

"Évaluation de deux méthodes de prédiction de l'occupation du sol basées sur les rotations de cultures dans une perspective de prédiction d'inondations boueuses"

Présenté par Miguel Urdiain-Arraiza

Promoteur(s) : Prof. Charles Bielders (UCL/ELI/ELIE)

Lecteurs : Prof. Pierre Defourny (UCL/ ELI/ELIE)
PhD. Alexandre Maignard (CRA)
PhD. Valérie Obsomer (SPW/DGO3)

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

Remerciements

Je souhaite remercier sincèrement toutes les personnes qui ont aidé de près ou de loin à la réalisation de ce mémoire. Sans leur aide celui-ci n'aurait pu voir le jour.

En premier lieu, je remercie mon promoteur le professeur Charles Bielders qui m'a suivi tout le long de ma démarche. Ses conseils et relectures me furent précieux au cours de cette recherche. Il a su m'accorder une certaine liberté tout en étant disponible et exigeant.

Merci à mes lecteurs, le professeur Charles Bielders, le professeur Pierre Defourny, Monsieur Alexandre Maignard et Madame Valerie Obomser qui ont pris leur temps pour lire ce mémoire et me donner leurs commentaires le concernant.

Merci à Baptiste de Walque et Pierre Demarcin de la cellule GISER qui m'ont transmis les couches de données des parcellaires agricoles anonymes.

Merci à Bernard Collignon, gestionnaire du Centre Alphonse de Marbaix, et à Monsieur Nicolas Simon, cartographe à la DGO3 qui m'ont apporté des informations précieuses sur les parcellaires agricoles anonymes.

Merci à Damien Jacques, doctorant au laboratoire de géomatique de l'UCL, qui m'a fourni des exemples de scripts pour matrices de transition.

Merci au SMCS qui m'a orienté sur les méthodes de statistiques à utiliser.

Merci à Nicolas Bellemans et Francois Waldner, chercheurs et doctorants au laboratoire de géomatiques de l'UCL, qui ont répondu à mes questions.

Enfin, merci à mes proches qui ont été présents pour moi lors de la réalisation de ce mémoire. Je remercie en particulier Camille Lefevre pour la relecture de celui-ci.

Tables des matières

TABLES DES MATIERES	I
LISTE DES FIGURES	III
LISTE DES TABLEAUX	IV
INTRODUCTION	1
1. ÉTAT DE L'ART	3
1.1 INONDATIONS BOUEUSES.....	3
1.1.1 Occurrence des inondations	3
1.1.2 Facteurs d'origine des inondations.....	3
1.1.3 Approches utilisables pour les risques d'inondations boueuses	4
1.1.4 Coûts des inondations boueuses et mesures de contrôle	5
1.2 CLASSIFICATION DE CULTURES.....	7
1.2.1 Télédétection	7
1.2.2 Information in situ - « LPIS ».....	9
1.3 SUCCESSIONS ET ROTATIONS DE CULTURES	11
1.3.1 Définition et intérêts.....	11
1.3.2 Rotations typiques en Wallonie.....	11
1.4 PRÉDICTION DE CULTURES.....	14
1.4.1 Échelle régionale.....	15
1.4.2 Échelle de l'exploitation.....	16
1.4.3 Échelle de la parcelle	17
1.4.4 Réseaux de neurones	18
1.5 ÉVALUATION DE LA PRÉDICTION ET DE LA CLASSIFICATION	19
2. MATÉRIELS ET MÉTHODES	21
2.1 DONNÉES ET OUTILS UTILISÉS.....	21
2.2 ZONE D'ÉTUDE	22
2.3 ANALYSE ET TRAITEMENT DES DONNÉES.....	25
2.3.1 Parcelles en bordure de zone coupées.....	25
2.3.2 Superposition de parcelles au sein des PAA.....	26
2.3.3 Évolution de la surface des PAA et du nombre de parcelles pour plusieurs zones géographiques.....	26
2.3.4 Évolution des types de cultures	27
2.3.5 Harmonisation et sélection des cultures.....	27
2.3.6 Comparaison des types de cultures pour plusieurs zones géographiques.....	27
2.3.7 Assemblage des différentes couches des PAA	28
2.3.8 Rotation de cultures	33
2.4 PRÉDICTION PAR MATRICES DE TRANSITION.....	34
2.4.1 Qu'est-ce qu'une matrice de transition ?	34
2.4.2 Résumé de la méthodologie	36
2.4.3 Calibration	37
2.4.4 Validation	39
2.4.5 Application aux groupes de cultures.....	42
2.5 PRÉDICTION PAR RÉSEAU DE NEURONES	43
2.5.1 Mise en place du système de réseau de neurones.....	43
2.5.2 Phase de calibration	45
2.5.3 Phase de validation	47

3. RÉSULTATS ET DISCUSSIONS.....	49
3.1 ANALYSE ET TRAITEMENT DES DONNÉES.....	49
3.1.1 <i>Superposition de parcelles parmi les PAA</i>	49
3.1.2 <i>Évolution de la surface des PAA et du nombre de parcelles pour plusieurs zones géographiques</i>	50
3.1.3 <i>Évolution des types de cultures</i>	51
3.1.4 <i>Harmonisation et sélection des cultures</i>	53
3.1.5 <i>Comparaison des types de cultures en fonction des zones géographiques</i>	56
3.1.6 <i>Assemblage des différentes couches des PAA</i>	57
3.1.7 <i>Rotation de cultures</i>	59
3.2 PRÉDICTION PAR MATRICES DE TRANSITION	61
3.2.1 <i>Calibration</i>	61
3.2.2 <i>Validation</i>	64
3.2.3 <i>Application aux groupes de cultures</i>	75
3.3 PRÉDICTION PAR RÉSEAU DE NEURONES.....	80
3.3.1 <i>Mise en place du réseau de neurones</i>	80
3.3.2 <i>Phase de calibration</i>	81
3.3.3 <i>Phase de validation</i>	83
3.3.4 <i>Application aux groupes de cultures</i>	84
CONCLUSION	86
BIBLIOGRAPHIE	88

Liste des figures

Fig. 1. Zone d'étude et agronomie. a) Régions agricoles de Belgique, b) Zones agro-géographiques de Wallonie (Sources des données : SPW)	22
Fig. 2. Occupation du sol et SIGISER (Sources des données : SPW).	23
Fig. 3. Les 3 exemples d'erreurs de superpositions.....	26
Fig. 4. Schéma explicatif de la fonction "Union" d'ArcGIS	28
Fig. 5. Schéma explicatif de la fonction "Intersect" d'ArcGIS.....	29
Fig. 6. Erreurs dues à la fonction « Intersect »	30
Fig. 7. Méthode des centroïdes avant intersection.....	31
Fig. 8. Méthode des centroïdes : différence avec ou sans l'option « Inside »	32
Fig. 9. Chaines de Markov : exemple fictif de graphe probabiliste et matrice de transition	34
Fig. 10. Résumé de la méthodologie : Calibration et Validation	37
Fig. 11. Exemple de Calibration : input à 2 années antérieures (2010_2011)	39
Fig. 12. Exemple de Calibration : input à 2 années antérieures (2014_2015)	40
Fig. 13. Validation appliquée aux points d'intersection	40
Fig. 14. Validation appliquée aux polygones intersectés	41
Fig. 15. Schéma général de la méthode des réseaux de neurones.	44
Fig. 16. Schéma de la phase de calibration réelle.	46
Fig. 17. Schéma de la phase de validation réelle.....	47
Fig. 18. Phase de validation : vérification de l'erreur.....	48
Fig. 19. Evolution des cultures principales de 2008 à 2016 (avec froment)	52
Fig. 20. Evolution des cultures principales de 2008 à 2016 (sauf froment)	52
Fig. 21. Calibration : évolution du PBP en fonction du nombre d'années antérieures.....	63
Fig. 22. PBP en utilisant des matrices seules, plus ou moins récentes, pour prédire 2016.	66
Fig. 23. Evolution du PBP et de l'Exactitude en fonction du nombre d'années antérieures.....	66
Fig. 24. Evolution du PBP : comparaison Calibration, Validation et matrices fusionnées.....	67
Fig. 25. Evolution de l'Exactitude : comparaison matrices seules et fusionnées	68
Fig. 26. Comparaison PBP et % de surface bien prédite de 2016.....	74
Fig. 27. Comparaison résultats 23 cultures et 4 groupes de cultures après prédiction.....	76
Fig. 28. Regroupement avant prédiction : résultats.....	78
Fig. 30. Groupes de cultures : comparaison PBP et % de surface bien prédite de 2016	79
Fig. 31. Réseau de neurone à plusieurs années d'input mises sur un pied d'égalité.....	80

Liste des tableaux

Tableau 1. Cultures typiques des régions agricoles wallonnes	12
Tableau 2. Rotations typiques de cultures en Région wallonne	13
Tableau 3. Coordonnées de la zone d'étude (Lambert 72)	22
Tableau 4. Résumé de l'occupation du sol (Sources des données : SPW ¹)	23
Tableau 5. Données d'étude des rotations de cultures	33
Tableau 6. Exemple de table de prédiction.....	35
Tableau 7. Matrice de transition fictive : exemple de combinaison de cultures	35
Tableau 8. Calibration : Inputs et outputs et création de matrices de transitions	38
Tableau 9. Validation : matrices de transition et inputs associés en vue de prédire les cultures de 2016.....	39
Tableau 10. Exemple de binarisation d'une parcelle avec le code 311.....	44
Tableau 11. Superposition de parcelles parmi les PAA de 2008 à 2016 sur notre zone d'étude.	49
Tableau 12. Résumé des PAA de 2008 à 2016 : Wallonie, Hesbaye, Zone d'étude.	50
Tableau 13. Harmonisation des codes cultures	54
Tableau 14. Cultures gardées pour prédiction	55
Tableau 15. Comparaison des types de cultures en 2016 en Wallonie, Hesbaye et Zone d'étude.	56
Tableau 16. Evolution du nombre de points d'intersection selon l'assemblage des PAA	58
Tableau 17. Nombre de classe de cultures par année	58
Tableau 18. Rotations les plus importantes en fonction du nombre d'années	59
Tableau 19. Exemple de matrice de transition avec 23 cultures : 2010_2011	62
Tableau 20. Evolution du nombre de combinaisons en fonction du nombre d'années antérieures utilisées.....	64
Tableau 21. Evolution du nombre de prédictions en fonction du nombre d'années antérieures utilisées.....	65
Tableau 22. Matrice de confusion sur l'Exactitude de la prédiction par matrices fusionnées de 2 années.	70
Tableau 23. . Matrice de confusion sur l'Exactitude de la prédiction par matrices fusionnées de 5 années	71
Tableau 24. Score F1 pour chaque culture pour 2 (haut) et 5 années (bas).	73
Tableau 25. Regroupement des cultures sous les groupes : Hiver, Printemps, Prairie, Autre	75
Tableau 26. Groupes de cultures : Matrice de transition 1 an.....	76
Tableau 27. Groupes de culture : Matrice de transition 2 ans fusionnée.....	77
Tableau 28. Regroupement avant prédiction : Matrice de confusion de l'exactitude de prédiction par matrices de 3 ans fusionnée.....	78
Tableau 29. Phase de calibration : impact individuel des types d'inputs.....	81
Tableau 30. Phase de calibration : Combinaison de type d'inputs.	82
Tableau 31. Phase de validation.....	83
Tableau 32. Groupes de cultures : calibration.	84
Tableau 33. Groupes de cultures : validation.....	85

Introduction

Les catastrophes liées à la boue existent depuis la nuit des temps. Les coulées boueuses, tout comme les avalanches et les éruptions volcaniques, font partie des phénomènes naturels de notre monde (Revet, 2017). Cependant, les inondations boueuses sont des phénomènes plus récents. Identifiées dans les années 80 et moins imposantes que les coulées boueuses (Auzet, 1987), celles-ci sont le résultat du transport par ruissellement de particules de terre provenant en général des champs lors de fortes pluies. Sans être aussi meurtrières que les coulées boueuses, les inondations boueuses entraînent des dégâts importants sur les infrastructures publiques (routes, réseau d'égouts), privées (habitations, caves et jardins) et sur l'environnement.

Ces inondations, observées entre autres dans la ceinture loessique européenne, engendrent des coûts importants à la société. La Belgique centrale est fort touchée avec des coûts chiffrant de 16 à 172 millions d'euros chaque année (Evrard et al., 2007). Ces inondations devenant de plus en plus fréquentes, une augmentation des moyens est mise en œuvre afin de les comprendre pour pouvoir ensuite les prévenir, les contenir et dans l'idéal les anticiper.

Un des principaux facteurs de ces inondations boueuses est l'érosion du sol au niveau des cultures agricoles (Boardman and Stammers, 1984). Cette érosion dépend en effet du type de culture présente. Il est donc capital de connaître la culture sur une parcelle afin de pouvoir identifier les zones à risque d'inondation boueuse.

Récemment familiarisé avec les concepts d'érosion des sols, de photogrammétrie et de géomatique, j'ai saisi cette opportunité pour mettre mes connaissances en application dans le but de développer une technique qui permette de prédire la prochaine culture présente sur une parcelle à l'avance. Cette prédiction s'inscrit dans un objectif plus large de « système d'alerte » précoce qui modéliserait les risques d'inondations boueuses et qui permettrait aux communes d'identifier les sites à risque afin d'anticiper les problèmes.

Une méthode de prédiction de cultures peut se baser sur des techniques de télédétection et/ou sur l'utilisation de séries temporelles des cultures précédentes. En effet, il est possible d'utiliser l'information des cultures présentes sur une parcelle l'année précédente afin d'avoir une idée de la prochaine culture. C'est possible grâce au concept de « rotation de culture » qui met en avant une logique dans la succession des cultures sur une parcelle.

De nombreux modèles permettent de simuler ces rotations de cultures. Il en existe qui se basent uniquement sur les connaissances théoriques mais la plupart sont des modèles qui reposent sur une approche statistique. Ce sont sur ces méthodes probabilistes que s'oriente particulièrement ce mémoire. La méthode que nous avons mise en avant est l'utilisation de chaînes de Markov et de matrices de transition. La seconde méthode est celle des réseaux de neurones, de plus en plus populaire dans le monde scientifique.

Finalement, l'objectif de ce mémoire est clair. Tout d'abord, analyser les différentes cultures présentes ainsi que leurs rotations. Ceci sera effectué pour une zone d'étude de 12.5 km² située en Hesbaye. Le second objectif est de démontrer l'intérêt et l'efficacité des deux méthodes à prédire une culture sur une parcelle à partir des cultures précédentes. À cet effet, nous évaluerons l'influence de différents facteurs, en particulier le nombre d'années antérieures prises en compte pour la prédiction, ou encore le nombre de classes de cultures que l'on cherche à prédire.

Pour obtenir des réponses complètes et pertinentes à ces objectifs, ce travail s'articule en 3 parties elles-mêmes divisées selon les sujets traités. Premièrement, un état de l'art décrit de manière plus détaillée le contexte et les méthodes qui peuvent entourer de près ou de loin notre problématique. Cet état de l'art définit tout d'abord le concept d'inondations boueuses en détaillant l'occurrence de celles-ci, les facteurs qui les influencent, les approches de modélisation ainsi que les coûts et les mesures de contrôle existants. Il s'intéresse ensuite aux méthodes de classifications de cultures, aux rotations de cultures et aux méthodes de prédictions des cultures et de leurs rotations. Ces méthodes de prédictions sont présentées selon l'échelle spatiale à laquelle elles s'intéressent. Enfin, les différentes techniques d'évaluation des méthodes de classification et de prédiction sont rapidement citées.

Deuxièmement, les matériels et méthodes utilisés décrivent le processus par lequel les résultats ont été obtenus. Cette partie présente tout d'abord les données et outils utilisés ainsi que la zone d'étude. Celle-ci explique ensuite la méthode suivie afin d'analyser et traiter les données. Enfin, elle traite en détail les méthodes de prédiction utilisées. Celles-ci comprennent la prédiction par matrices de transition et la prédiction par réseaux de neurones.

Troisièmement, les résultats seront présentés et discutés. Pour clôturer ce mémoire, la conclusion reprendra les points importants que nous avons pu mettre en exergue. Elle définira les méthodes qui ont le mieux fonctionné et donnera des pistes de réflexions afin d'améliorer ces méthodes.

Je vous souhaite une agréable lecture.

1. État de l'art

1.1 Inondations boueuses

Les inondations boueuses sont des inondations causées directement par ruissellement, transportant de grandes quantités de sol comme sédiments en suspension ou en charge de fond provenant fréquemment de terres agricoles. Ces inondations causent des dégâts aux propriétés et aux infrastructures (Boardman et al., 2006).

1.1.1 Occurrence des inondations

Les inondations boueuses sont particulièrement fréquentes en Belgique centrale et plus précisément au niveau de la ceinture loessique. Celle-ci comprend la région limoneuse et sablo limoneuse. Selon Verstraeten and Poesen (1999), 43% des municipalités doivent y gérer des inondations boueuses de temps en temps et 35% ont des problèmes d'inondations à petite échelle dû au débordement de cours d'eau. Pour ces communes, la fréquence de ces événements est importante. Dans 20% des cas, la période de retour est inférieure à 1 an et dans 90% des cas elle est inférieure à 5 ans. Les municipalités touchées par les inondations boueuses reconnaissent à 89% que le problème provient du ruissellement des terres cultivées. Plus récemment, Evrard et al.(2007) déclarent que 79% des municipalités de Belgique centrale ont été touchées au moins une fois par des inondations boueuses ces 10 dernières années. Et parmi celles-ci, 22% ont été touchées plus de 10 fois durant cette période. Le passage de 43% dans les années 90 à 79% en 2005 montre une énorme augmentation du nombre d'inondations boueuses ces dernières années.

1.1.2 Facteurs d'origine des inondations

Le premier facteur indispensable pour des inondations boueuses est la présence de pluie. Selon Evrard et al. (2007), les interventions des pompiers concernant des inondations boueuses surviennent à 85% du temps entre mai et septembre suite à événements pluvieux convectifs locaux (orages). Les événements pluvieux avec plus de 45mm de pluie conduisent dans 100% des cas à des inondations.

Cependant, cette quantité de pluie nécessaire aux inondations dépend du mois. C'est là qu'intervient le deuxième facteur : les cultures. Selon Evrard et al. (2007), en mai et juin, la surface du sol est peu recouverte dû aux cultures de printemps, principalement de maïs, pomme de terre et betterave. C'est à ce moment-là que l'érosivité est à son maximum et les inondations sont déclenchées à partir de 25mm de pluie. Finalement, l'augmentation du nombre d'inondations ces dernières années peut être en partie expliqué selon Evrard et al. (2007) par l'expansion des cultures sarclées, vergers et légumes

au détriment des cultures céréalières. En plus des cultures, le remembrement et l'expansion de zones urbaines favorisent un ruissellement rapide et une mauvaise infiltration.

Enfin, les derniers facteurs sont liés au sol. Les sols limoneux et sablo limoneux de la ceinture de loess sont particulièrement sensibles à l'érosion. En effet, ils possèdent un facteur d'érodibilité important. De plus si la pente est importante, le risque d'érosion augmente. Selon Evrard et al. (2007), celle-ci couplée aux réseaux de routes et de fossés facilitent le transfert d'écoulements de terres cultivées jusqu'aux zones habitées dans 64% des cas. Le reste des écoulements sont transférés par des phénomènes d'érosion (rigoles et ravines temporaires). Ces écoulements de boues sont générés à 90% en amont au niveau de versants (1-30 ha, par érosion en nappe) ou au niveau de petits bassins (10-300 ha, par érosion concentrée et 1 thalweg). Selon Verstraeten and Poesen (1999), une différence d'altitude importante dans une municipalité est un facteur qui augmente les inondations boueuses. De même, dans les municipalités touchées par des inondations boueuses, les fermiers cultivent plus souvent sur des pentes importantes (supérieures à 4%).

1.1.3 Approches utilisables pour les risques d'inondations boueuses

Il existe de nombreux modèles capables de prédire un risque de d'inondations boueuses. La plupart sont des modèles d'érosion. Parmi ceux-ci, il y a deux types de modèles : ceux qui modélisent uniquement le détachement et le transport de particules de sol et ceux qui modélisent en plus la redéposition des sédiments. Deux modèles sont présentés ci-dessous : le RUSLE et les indices de connectivités de Borselli. Le premier modélise exclusivement le détachement et transport de particules de sol et le deuxième arrive à coupler les deux processus.

Modèle USLE - RUSLE

Le modèle d'érosion le plus connu dans le monde est l'équation universelle de perte en sol (USLE; Wischmeier and C, 1978). Il fut développé aux Etats-Unis mais est applicable en Belgique si des adaptations sont faites sur certains paramètres. La dernière version de l'USLE est la Revised USLE (RUSLE; Renard et al., 1991). Celle-ci calcule la perte en sol moyenne annuelle A en tonne par hectare et par an. Ce calcul est réalisé au niveau parcellaire. Il est le produit de 5 facteurs :

$$A = R.K.LS.C.P \quad (1)$$

Où R est l'érosivité de la pluie [$\text{MJ.mm.ha}^{-1}.\text{h}^{-1}.\text{an}^{-1}$], K est l'érodibilité du sol [$\text{T.ha.h.MJ}^{-1}.\text{mm}^{-1}$], LS est le facteur topographique [-], C est le facteur représentatif de la couverture du sol [-] et P est le facteur d'aménagement antiérosif [-].

Le facteur C représente donc l'impact de la couverture qui permet de jouer sur les pertes en sol. Il est fonction du Soil Loss Ratio (SLR) de la culture et de l'érosivité de la pluie (Renard et al., 1997). La plupart des facteurs qui constituent le SLR varient en fonction du type de culture utilisé sur le sol actuellement et antérieurement. Il serait donc intéressant de connaître à l'avance la culture présente sur les parcelles du bassin versant afin de pouvoir prévoir le risque d'érosion au moyen de ce modèle.

Modèle Indice de Connectivité – Borselli

Le modèle basé sur un indice de connectivité des sédiments (Borselli et al., 2008) calcule un indice adimensionnel en chaque pixel d'un MNT. Cet indice représente une information sur la connectivité des sédiments en ce pixel (Dewalque, 2016). La connectivité est estimée par rapport à un 'puits' qui correspond à l'exutoire du bassin versant. Ce puits peut être, dans le cas d'inondations boueuses, apparenté au récepteur de risque, c'est-à-dire aux infrastructures touchées en général. L'indice de Borselli est le logarithme du rapport entre la distance en amont du pixel et la distance en aval du pixel :

$$IC_k = \log_{10} \frac{\bar{W}_k \bar{S}_k \sqrt{A_k}}{\left(\sum_i \frac{d_i}{\bar{W}_i \bar{S}_i} \right)_k} \quad (2)$$

Où IC_k [-] est l'Indice de Connectivité de Borselli du pixel k, $\bar{W}_k$ [-] est la moyenne des facteurs de pondération sur toute l'aire contributive en amont du pixel k, $\bar{S}_k$ [m.m⁻¹] est la moyenne des facteurs de pente sur toute l'aire contributive en amont du pixel k et A_k [m²] est l'aire contributive en amont du pixel k. Où d_i [m] est la longueur de la cellule i faisant partie du chemin d'écoulement depuis le pixel k jusqu'au puit, $\bar{W}_i$ [-] est le facteur de pondération de la cellule i, qui reflète l'occupation du sol (ici représenté par le facteur C) et $\bar{S}_i$ [m.m⁻¹] est le facteur de pente.

Parmi tous les facteurs qui influent sur l'indice de connectivité, l'occupation du sol intervient deux fois. Une fois au numérateur au niveau de la production de sédiments et via le paramètre $\bar{W}_k$. Et une fois au dénominateur au niveau de la redéposition des sédiments et via le paramètre $\bar{W}_i$. Il serait donc à nouveau intéressant de connaître à l'avance la culture présente sur les parcelles du bassin versant afin de pouvoir prévoir le risque d'érosion.

1.1.4 Coûts des inondations boueuses et mesures de contrôle

Selon Evrard et al. (2007), en région centrale belge, les coûts totaux sociétaux des inondations boueuses chaque année s'élèvent de 16 à 172 millions d'euros selon l'intensité des tempêtes. Ces coûts très importants justifient la mise en place de mesures afin de contrôler l'érosion et limiter un écoulement vers les zone d'infrastructures.

Ces mesures peuvent être soit préventives, soit curatives. Les mesures préventives évitent le ruissellement et le détachement du sol. On compte parmi elles les bonnes pratiques agricoles (rotations de cultures), techniques culturales simplifiées, couverture du sol en hiver ou le cloisonnement inter-buttes. Les mesures curatives permettent de retenir l'eau et les sédiments. Parmi les pratiques curatives, il y a la mise en place de bande et chenaux enherbés, de murets, de digues ou encore de fascines (Biielders, 2016).

Les mesures préventives sont compliquées à mettre en place car elles demandent au fermier de changer sa façon de travailler et éventuellement ses cultures. De plus, des mesures préventives sont déjà imposées via la conditionnalité de la PAC. Celle-ci oblige les parcelles possédant des pentes de plus de 10% à ne pas cultiver de cultures sarclées à moins d'installer une bande enherbée de 6m le long de la parcelle.

Les mesures curatives peuvent être mises en place dès qu'un risque de d'inondation boueuse est remarqué sur une parcelle. Il faut souvent l'aval du fermier mais la commune peut organiser les travaux si le risque est avéré. Par exemple, une parcelle peut devenir à risque d'une année à l'autre à cause d'un changement de culture. Si ce changement de culture est remarqué avant les inondations qui commencent en mai, il est possible d'agir afin de limiter les dégâts. Le but est donc de savoir le plus tôt possible les cultures présentes sur les parcelles.

1.2 Classification de cultures

Afin de connaître les cultures présentes dans une zone d'un ensemble de parcelles et pouvoir réaliser un modèle qui permet de prédire le risque d'érosion, il faut que chaque parcelle soit classifiée selon la culture qu'elle représente. Cette classification peut se faire principalement par 2 méthodes : la télédétection et l'information in situ.

Dans notre cas, il faut différencier la cartographie de l'occupation du sol (land cover mapping) des terres cultivées (cropland mapping) et des types de cultures (crop type mapping). À noter que les deux derniers termes sont utilisés différemment en fonction des pays et des administrations.

La cartographie de l'occupation du sol prend en compte l'entièreté de la surface terrestre et la divise en classes assez générales comme les plans d'eau, les zones urbaines, les forêts feuillues, les forêts conifères, les prairies et les terres cultivées. Selon (Gregorio and FAO, 2005), les terres cultivées sont définies par le JECAM (Joint Experiment of Crop Assessment Monitoring) comme : « un terrain de minimum 0.25 ha (largeur min 30 m) qui est semé/planté et récoltable au moins une fois tous les 12 mois après la date de semis/plantation ». Dans ce cas-ci, les prairies ne sont pas considérées comme terres cultivées. Lorsque nous discutons du type de culture, nous parlons généralement des différentes espèces cultivées (par ex : maïs, blé, orge, betterave, pomme de terre) qui constituent la carte des terres cultivées. Dans certains cas, les prairies peuvent être prises en compte, qu'elles soient temporaires ou permanentes.

1.2.1 Télédétection

Moyens techniques de télédétection

Les moyens techniques de cartographie via télédétection sont nombreux. Il est possible d'utiliser soit des images satellites, soit aériennes. Parmi les images satellites, il faut distinguer les données de radars et de capteurs optiques. Les premiers sont des instruments actifs qui émettent leur propre rayonnement vers la cible qui est réfléchi et analysé par le capteur. Les longueurs d'ondes sont plus élevées, de l'ordre du micro-ondes (1-300 mm) (Campbell and Wynne, 2011). Les seconds sont des instruments passifs qui captent l'énergie du soleil réfléchi ou retransmise par la cible dans des longueurs d'ondes de l'ordre du visible (400-700 nm) et du thermique (3-15 μ m).

Les capteurs radar les plus communs sont nommés SAR (Synthetic Aperture Radar). Les radars peuvent avoir différentes bandes spectrales selon leurs longueurs d'ondes : X, C, S, L, P, etc. (Ressources naturelles Canada, 2008a). La résolution spectrale des données optiques peut être classée par ordre croissant, de panchromatique (PAN, 1 bande spectrale) à multi-spectral (MULTI, 3 à 15 bandes), jusqu'à hyperspectral (des centaines de bandes) (Campbell and Wynne, 2011).

La résolution spatiale d'une image satellite correspond à la taille de la zone couverte par un pixel de l'image (EOEdu, 2016). Grâce à la technologie, on arrive à des résolutions spatiales de plus en plus élevées ces dernières années. Un bon exemple est Sentinel-2 lancé récemment qui possède des données MULTI avec 13 bandes et une résolution spatiale allant de 10 à 60 m selon les bandes (ESA, 2017).

La résolution temporelle d'un satellite est le temps que prend un satellite pour effectuer un cycle orbital complet. C'est le temps écoulé entre la prise d'une image à un endroit et la prochaine image prise à cet endroit. (Ressources naturelles Canada, 2008b). Sentinel-2 possède une résolution de 5 jours.

Méthodes de classifications

Les informations contenues dans les bandes spectrales permettent de classifier les pixels en fonction de leur occupation du sol, du type de culture, etc. En effet, lorsque des pixels possèdent les mêmes caractéristiques dans les bandes spectrales analysées, on peut les regrouper en classes spectrales. l'étape suivante est de faire correspondre ces classes spectrales avec une signification réelle et de les associer à des classes d'information (zone artificielle, forêt, plan d'eau, types de cultures, etc.)(Ressources naturelles Canada, 2008c). Il y a deux grands types d'approches de classification : classification supervisée et non supervisée. Les classifications non supervisées se font, comme leur nom l'indique, sans aucune supervision humaine lorsqu'il s'agit de prédéterminer des classes d'informations pour certaines classes spectrales.

Les classifications supervisées sont plus courantes et sont influencées par l'utilisateur dès le début. Celui-ci identifie sur son image des zones homogènes dont il peut certifier la correspondance avec certaines classes d'information. L'association classe spectrale et classe d'information est dès lors déjà prédéterminée. L'ordinateur peut ensuite analyser le reste de l'image et reconnaître certaines régions ayant les mêmes propriétés que les classes prédéfinies.

Différentes manières de traiter l'image

Comme vu à propos des résolutions temporelles, il y a toute une série d'images satellites récoltées durant l'année qui permet de suivre l'évolution des cultures. À partir de ces images, il est possible de n'en sélectionner qu'une seule et de classer les cultures selon celle-ci ou d'utiliser plusieurs images pour classer les cultures (Langley et al., 2001). Généralement, la classification se fait à la fin de la saison (septembre-octobre) lorsque toutes les images liées à la culture ont été récoltées afin d'utiliser les meilleures images possibles. Cette classification est bien sûr la plus précise que l'on puisse espérer car elle est faite avec toutes les cartes en main. L'identification des cultures plus tôt dans la saison est moins exacte dû à la couverture nuageuse et à la disponibilité des images moins importante (Inglada

et al., 2016). Cependant, l'intérêt de connaître le type de culture plus tôt dans la saison est primordial dans notre cas. Notre modèle de risque d'érosion a pour but final de pouvoir prendre des mesures avant les événements à risque se déroulant à partir de mai.

Il est donc possible de connaître le type de culture dès avril ou mai grâce au early crop type mapping. En effet, selon l'étude de Inglada et al. (2016) qui utilise des images optiques Landsat8 couplées à des images SAR Sentinel-1 les résultats démontrent un coefficient kappa¹ supérieur à 0.75 début mai. Cela correspond à une concordance excellente selon Fleiss (1981). L'exactitude de classification est de 78%. Leur classification prend en compte 5 cultures en plus des prairies et du sol nu (correspondant à une culture d'été non émergée lors de la vérification in situ). Cependant, ce qui nous intéresse est d'avoir une idée encore plus tôt des cultures présentes dans la saison afin de prendre des mesures, c'est-à-dire vers fin janvier. À cette date, le coefficient kappa est d'environ 0.60 selon Inglada et al. (2016) et l'exactitude de 68% afin de classer la culture de mai. Selon Osman et al. (2015), en utilisant une classification en temps réel selon le nombre d'images disponibles mais sans SAR, on arrive à seulement 40% d'exactitude fin janvier.

1.2.2 Information in situ - « LPIS »

En ce qui concerne les méthodes dites « in situ », il n'en existe pas autant. Un système utilisé principalement en Europe depuis 2003 est le LPIS : le Land Parcel Identification System. C'est une couche vectorielle d'information géographique annuelle précisant la localisation des limites de parcelles et attribuant à chacune d'entre elles un identifiant en plus d'autres informations pertinentes telles que leur superficie et la culture principale emblavée (Leteinturier et al., 2007). Le LPIS est un outil faisant partie du IACS/SIGEC (Integrated Administration and Control System/ Système Intégré de Gestion Et Contrôle). Le SIGEC fut créé en 1992 pour améliorer l'efficacité des paiements directs des agriculteurs sous la PAC (Politique Agricole Commune). Son objectif était de s'assurer qu'il y ait un paiement correct aux agriculteurs et qu'il y ait traçabilité de paiement (FAO, 2006).

Le LPIS peut être de différentes formes selon les pays. La plupart du temps, on utilise des orthophotoplans² et les agriculteurs remplissent leur déclaration de superficie agricole sur ceux-ci. En Belgique, les agriculteurs doivent remplir leur déclaration de superficie pour fin avril de l'année en cours. Ces déclarations de superficie agricoles concernent la culture qui sera récoltée dans l'année car il est possible qu'il y ait un semis de printemps et un semis d'hiver la même année. Seule la récolte de la culture de printemps sera mentionnée. Il y a plus de 90 types de cultures qui peuvent être encodées.

¹ Expliqué au point 1.5 et dans la méthodologie

² Photographies aériennes sur lesquelles sont effectuées un redressement différentiel afin d'obtenir un plan

Jusqu'à récemment, la déclaration de superficie se faisait sous format papier mais le format électronique est maintenant obligatoire. La couche vectorielle n'est disponible en ligne sur le Géoportail de Wallonie¹ qu'en janvier de l'année suivante, soit plusieurs mois après la fin de la saison. Le LPIS est considéré comme très fiable car chaque année 7% des parcelles qui le constituent sont contrôlées (Leteinturier et al., 2007). Aux USA, l'équivalent du LPIS est le USDA Cropland Data Layer qui renseigne les types de cultures avec une résolution de 30 m (Boryan et al., 2011).

¹ geoportail.wallonie.be

1.3 Successions et rotations de cultures

1.3.1 Définition et intérêts

Une rotation de cultures est une séquence de cultures pratiquées successivement sur un certain champ (Wibberley, 1996). Généralement, chaque année une culture suit celle de l'année passée en suivant un ordre défini par le fermier. L'intérêt des rotations de cultures peut être multiple selon Castellazzi et al. (2008). Tout d'abord économique, en maximisant le profit en fonction des ressources du sol et l'optimisation des semis et récoltes. Ensuite, un intérêt agricole si le rendement de production est l'objectif. Et enfin, environnemental si le but est de limiter l'utilisation des pesticides ou bien encore de limiter l'érosion.

Ces 3 points sont améliorés par des exemples concrets selon Leteinturier et al. (2007):

- La rotation diversifiée contribue à rompre le cycle vital des organismes nuisibles aux cultures et restreint les agents pathogènes et adventices (Boller et al., 1997; Gerowitt, 2003).
- La structure du sol est améliorée grâce aux racines variables des différentes cultures (Hamza and Anderson, 2005), ces racines permettent aussi une meilleure exploration du sol et une meilleure nutrition des plantes ainsi qu'une augmentation de la séquestration du carbone de l'atmosphère (Dendoncker et al., 2004).
- Le sol est enrichi en éléments nutritifs provenant des diverses cultures et améliore la qualité de la matière organique de celui-ci. Les légumineuses par exemple permettent l'ajout d'azote symbiotique (Leteinturier et al., 2007).
- L'érosion est limitée grâce à l'utilisation de céréales et de cultures sarclées alternées à la place de monocultures en rang. Les bienfaits sont encore plus importants si il y a utilisation de couverture herbacée (Jankauskas and Jankauskiene, 2003).
- La biodiversité est favorisée (Albrecht, 2003) par des paysages dits en mosaïques (Paoletti, 1999) plutôt qu'uniformes dues au monocultures.

1.3.2 Rotations typiques en Wallonie

Les rotations de cultures à l'échelle Wallonne ont été particulièrement étudiées par Leteinturier et al. (2007) et Bielders et al. (2011). Leteinturier et al. (2007) réalisent un diagnostic agronomique et environnemental des successions culturales au niveau des régions agricoles wallonnes. Ils analysent l'effet de la qualité des cultures précédentes sur les suivantes, le temps de retour moyen des cultures, le nombre de cultures dans une succession et la durée des périodes d'intercultures. Tout est analysé à l'échelle de la parcelle et les résultats sont agrégés à l'échelle régionale. Pour cette étude, Leteinturier

et al. (2007) se sont basés sur les LPIS de 7 années allant de 1997 à 2003. Dans leurs travaux, les prairies permanentes ne sont pas prises en compte car elles ne font pas partie du cycle de rotation. Les prairies temporaires non plus car elles sont souvent identifiées comme telles alors qu'elles sont permanentes. Enfin, certaines cultures marginales et appellations vagues ne sont pas considérées non plus ni les petits polygones d'intersection (0.3ha max).

Les cultures les plus présentes par région agricole wallonne sont définies au tableau 1 (Bielders et al., 2011). Cela montre déjà une certaine diversité dans les cultures parmi les régions agricoles wallonnes.

Tableau 1. Cultures typiques des régions agricoles wallonnes

Région agricole	Culture
Région (sablo) limoneuse	Betterave, Pomme de terre, Maïs, Lin, Froment, Escourgeon
Condroz	Betterave, Maïs, Froment, Escourgeon, Colza, Epeautre
Famenne	Maïs, Froment, Escourgeon, Colza, Epeautre
Région Jurassique	Maïs, Froment, Epeautre

Typiquement, une rotation de culture fait en sorte d'avoir une certaine diversité de cultures selon les années. Cependant, cette diversité n'est pas totale. Ainsi, Leteinturier et al. (2007) montrent que sur une période de 7 ans, il y aura en moyenne 3 à 4 cultures différentes observées sur une parcelle. Cela signifie donc que certaines cultures réapparaissent sur les 7 années et ne sont pas uniques.

Comme vu pour les types de cultures, ces rotations de cultures sont aussi très différentes selon la région agricole. Ci-dessous sont reprises les rotations les plus fréquentes en Wallonie en combinant les résultats de Bielders et al. (2011) et Leteinturier et al. (2007).

Tableau 2. Rotations typiques de cultures en Région wallonne

Type de rotation	Cultures
Monocultures	Maïs – Maïs – etc.
	Froment – Froment – etc.
	Escourgeon – Escourgeon – etc.
Succession de 2 cultures	Betterave – Froment
	Froment – Betterave
	Escourgeon – Betterave
	Maïs – Froment
	Froment – Maïs
	Pomme de terre - Froment
	Froment – Escourgeon
Succession de 3 cultures	Froment – Escourgeon – Betterave
	Froment – Escourgeon – Pomme de terre
	Froment – Escourgeon – Maïs
	Betterave – Froment – Pomme de terre
	Betterave – Froment – Maïs
	Pomme de terre – Escourgeon - Betterave

Ces successions de cultures les plus importantes montrent certains schémas répétitifs. Aucune de ces cultures dites de printemps (maïs, pomme de terre, betterave, lin) ne se suivent successivement hormis la monoculture de maïs. Celles-ci montrent que les rotations de cultures restent encore fort utilisées et qu'une culture de printemps est suivie par une culture d'hiver. Une culture d'hiver peut quant à elle être suivie par une culture de printemps ou d'hiver. Dans les successions à 3 années est par exemple représentée la rotation « Froment – Escourgeon – Betterave » typique de la région limoneuse et de la Hesbaye (Seutin et al., 2011). Selon Osman et al. (2015) ces régularités dans les pratiques agricoles permettent de prédire avec une certaine exactitude le type de culture présent sur un champ si l'on connaît les cultures précédentes.

1.4 Prédiction de cultures

Comme vu précédemment, nous avons besoin de prédire les cultures le plus tôt possible. Selon Osman et al. (2015), la prédiction des terres cultivées peut se faire soit via la déclaration administrative des fermiers, soit via la télédétection. La télédétection dépend de la surface du sol recouverte par les cultures. Jusqu'à mi-février, les seules cultures présentes sont les cultures d'hiver et intercalaires et ne sont pas très développées, ce qui limite leur identification. Sachant que généralement les cultures de printemps sont semées vers mars-avril, il y a peu de données disponibles avant cette période. Il est possible de contourner ce manque d'information en utilisant la déclaration administrative des fermiers, via les LPIS nationaux. Cette approche modélise les rotations de cultures afin de prédire, dès l'obtention des LPIS, la culture qui a le plus de chance d'être présente sur le champ en utilisant les séquences de cultures des 2-5 années précédentes (Osman et al., 2015). Afin d'avoir une idée des possibilités autour de ce sujet, il est important de voir ce qui existe comme avancées à ce propos.

Au niveau des types de méthodes, on a soit :

- Des approches à partir de connaissances théoriques : modèle sur la science du vivant, économie, ou expertise agronomique. Ces méthodes sont le plus souvent déterministes.
- Des approches basées sur l'apprentissage automatique des données via le machine learning : rotations de cultures, utilisation de paramètres influençant la culture utilisée. Ces méthodes sont les plus souvent stochastiques.

Au niveau de l'échelle de prédiction celle-ci peut être :

- Régionale : représentation des différentes rotations de cultures dans une zone large soumise à une politique régionale. Par exemple l'intérieur d'un bassin versant.
- Exploitation : représentation des cultures qu'une ferme gère, quelles rotations doivent être utilisées. Les conseils de rotation sont donnés au fermier, à lui d'ensuite appliquer ceux-ci à l'échelle de ses différentes parcelles.
- Parcelle : prédiction d'une certaine culture présente sur la parcelle l'année suivante. Pas de conseil mais une prédiction.

Enfin, la possibilité d'utiliser les réseaux de neurones dans le cadre des prédictions de cultures a été étudiée. Quelques travaux liés aux réseaux de neurones et aux cultures agricoles ont été analysés.

1.4.1 Échelle régionale

Approches théoriques

Tout d'abord, la connaissance de l'occupation du sol peut être utile lors de simulations hydrologiques. Wechsung et al. (2000) se sont intéressés à la variation de débit dans les rivières en fonction de l'occupation du sol. En utilisant le modèle SWIM, ils ont associé la fertilité des sols à leur type (sableux, limono-sableux, limoneux) et donc relié ces types de sols à certains types de cultures plus ou moins sensibles à la fertilité. Selon le type de sol, ils ont donc déterminé une rotation de culture type. Ils ont ensuite créé un générateur stochastique de culture afin d'attribuer ces rotations de cultures aux sols cultivés. Les rotations de cultures sont ensuite prises en compte dans le modèle hydrologique.

Dans la même idée, Klöcking et al. (2003) s'intéressent aux scénarios d'occupation du sol et leurs impacts sur la dynamique de l'eau ainsi que sa qualité. Cela reste cependant des scénarios de cultures probables à l'échelle d'un bassin versant régional et pas une prédiction à l'échelle de la parcelle. Aussi, ils s'intéressent plus à la prédiction verticale (temporelle) et utilisent des connaissances expertes et règles agronomiques afin d'y parvenir.

Et Salmon-Monviola et al. (2012), toujours dans le but d'évaluer l'impact des pratiques agricoles sur l'environnement, essayent de reconstruire les dynamiques spatiales des systèmes de cultures passées à partir des données disponibles et de construire les scénarios des systèmes de cultures futures. Ici, la modélisation s'est faite en séparant les cultures d'été (méthode stochastique via chaînes de Markov) et les cultures d'hiver (via connaissance experte déterministe).

ROTOR est un modèle qui prédit les meilleures rotations de cultures pour des systèmes spécifiques à l'agriculture biologique (Bachinger and Zander, 2007). Dans leur cas, il est important d'avoir une rotation de culture optimale car il n'y a pas de fertilisants minéraux et pas de pesticides. Ils prennent donc en compte les activités liées à la production de la culture (labour, récolte, semis) afin de les optimiser selon le rendement, les performances économiques, la balance en nitrate et le développement des adventices. Les données proviennent d'expériences sur le terrain, d'essais en ferme, d'enquêtes, de connaissances expertes. On reste cependant à l'échelle de systèmes agricoles et non à l'échelle parcellaire.

Enfin, un modèle qui utilise les bases de données de cultures d'une exploitation et d'une région sur plusieurs années fournies par le SIGEC est CropRota (Schönhart et al., 2011). Il utilise les données des 7 dernières années couplées à des critères agronomiques afin de générer des rotations typiques pour les exploitations et les régions. Cela reste une méthode appropriée d'estimation des rotations de cultures à l'échelle régionale donnée mais pas une prédiction de culture sur une parcelle donnée.

Une part importante de la prédiction qui n'est pas prise en compte avec la connaissance théorique est le changement des conditions qui peuvent affecter le choix des types de cultures. A moyen-court terme, une décision politique ou un changement de contexte économique peut modifier les types de cultures choisies. A moyen-long terme, des problèmes phytosanitaires ou le changement climatique peuvent aussi modifier le type de cultures sur les champs. Par exemple, Supit et al. (2012) ont pris en compte les effets du changement climatique sur les rendements des cultures européennes. Ils montrent que la prise en compte des changements de rendements peut jouer sur le choix du type de culture.

Approches d'apprentissage automatique

Parmi les études réalisées sur base de données existantes spatio-temporelles, CarrotAge de Le Ber et al. (2006) est un logiciel basé sur un modèle de Markov caché (HMM). Il a été développé pour étudier les scénarios de cultures d'un territoire et utilise une base de données « TerUti » qui rend compte du type d'occupation du sol par zones de 9ha. CarrotAge ne renseigne cependant pas sur les types de cultures à l'échelle de la parcelle.

Enfin, une étude de (Xiao et al., 2014) a utilisé les données TerUti et une méthodologie de data mining stochastique afin de représenter les scénarios de cultures à une échelle régionale large. Ces chercheurs ont utilisé des modèles de Markov cachés ainsi que de l'analyse cluster non supervisée. Cependant, bien que les données TerUti qui couvrent 1992 à 2003 aient une résolution de 300m, elles ne permettent pas une prédiction précise à l'échelle de la parcelle dans ce modèle.

1.4.2 Échelle de l'exploitation

Approches théoriques

Parmi les premiers modèles qui se sont intéressés à l'élaboration de rotations de cultures, on peut tout d'abord citer l'étude de (El-Nazer and McCarl, 1986). Ces auteurs s'intéressent au choix des rotations de cultures non pas en considérant des rotations prédéterminées mais en permettant au modèle de déterminer librement la rotation en maximisant le profit et en tenant compte de l'aversion au risque. La méthode utilise la programmation linéaire et l'analyse par régression de l'historique des cultures précédentes.

Un autre modèle est ROTAT (Dogliotti et al., 2003). Ce logiciel basé sur la programmation linéaire a été développé pour générer des rotations de cultures en se basant sur des critères agronomiques explicites. Le logiciel génère toutes les rotations possibles sur base d'une liste prédéfinie. L'utilisateur, grâce à ses connaissances expertes, utilise des filtres pour limiter certaines combinaisons de cultures. Dans le cas de Dogliotti et al. (2003), 840 rotations de cultures ont été modélisées et analysées du

point de vue du rendement économique, de l'érosion du sol, du risque de maladies et adventices, du besoin en azote et du lessivage de celui-ci. Cependant, ce modèle travaille à l'échelle de l'exploitation et non des parcelles.

Ensuite, Haneveld and Stegeman (2005) ont proposé une méthode pour donner les successions de cultures exigées dans le but d'une planification de la production agricole. Pour ce faire, ils déterminent les successions de cultures inadmissibles. Ils ont créé un algorithme mathématique qui se base sur la programmation linéaire sur plusieurs années afin de résoudre le problème comme un « max flow ».

Enfin, Detlefsen (2004) tente de modéliser les rotations de culture en utilisant la méthode du « network flow modeling » qui se rapproche d'une méthode de programmation linéaire avec une structure spéciale. Il s'intéresse à définir une rotation de culture selon les cultures dans la rotation et le nombre d'hectares pour chaque culture.

Approche d'apprentissage automatique

Aurbacher and Dabbert (2011) présentent une approche dans le but de combler l'écart entre la connaissance d'une répartition de culture sur une région et la connaissance d'une succession de cultures sur une parcelle précise. Ils se basent sur deux nouveautés : 1) l'utilisation d'un minimum cross entropy afin d'avoir une seule solution lors de l'utilisation d'un modèle de programmation linéaire au niveau de la parcelle 2) l'utilisation de chaînes de Markov calibrées qui permettent, contrairement aux simulations stochastiques inconditionnelles, d'avoir des séquences de cultures plus proches des résultats des modèles de programmation linéaire.

1.4.3 Échelle de la parcelle

Approches d'apprentissage automatique

Castellazzi et al. (2008) proposent une représentation mathématique des rotations et une structure pour modéliser les rotations de cultures. Les rotations peuvent être classées selon leur variabilité interne et leur tendance cyclique. On peut quantifier les rotations comme des matrices de transitions représentant des processus stochastiques avec la propriété markovienne que l'attribution d'une année dépend de l'année précédente (Cox and Miller, 1977). Ces matrices pourraient aussi permettre de modéliser la transition entre rotations à cause de variables externes (climat, changement du marché).

Dernièrement, Osman et al. (2015) ont poussé le concept des modèles de Markov plus loin. Ils ont développé un modèle de Markov qui prend en compte les connaissances expertes. Ils ont donc étendu le concept de Markov Network à la logique pour créer un Markov Logic Network (MLN). Au lieu de s'intéresser seulement aux probabilités qu'une certaine culture soit présente en fonction de la culture précédente, ils ajoutent d'autres types de connaissances plutôt logique. Par exemple, une nouvelle loi

sur les nitrates, des changements dans les prix ou une loi sur les biocarburants pourraient changer les rotations de cultures selon Osman et al. (2015). Non seulement les probabilités peuvent changer en fonction de ces règles logiques que l'on impose, mais dans certain cas il est possible de tout simplement les remplacer par de la logique de premier ordre (if-then). Tout cela est possible en associant un poids à chaque formule qui correspond à une séquence de culture. Les poids associés doivent bien sûr être appris pour chacune des séquences possibles de cultures.

1.4.4 Réseaux de neurones

Dans le domaine des réseaux de neurones, peu de travaux ont été réalisés sur la prédiction de culture. On peut néanmoins citer le travail de Dahikar and Rode (2014) qui s'intéresse à la prédiction d'une culture à un endroit donné afin d'avoir le meilleur rendement possible. Ils utilisent un Artificial Neural Network (ANN) basé sur le feedforward back propagation. Ils ont utilisé des paramètres mesurables comme le type de sol, le pH, l'azote, le phosphate, la température, l'humidité et la pluviométrie pour déterminer le rendement des cultures. Ainsi, la meilleure culture est déterminée selon les paramètres mesurés. Cette méthode ne prend cependant pas en compte les rotations de cultures parmi ces paramètres. D'autres travaux ont été réalisés à l'aide d'ANN sur le sujet du rendement des cultures. On peut citer le travail de Ranjeet and Armstrong (2014) qui s'intéresse au rendement des cultures au Népal. Une comparaison de modèle de régression et de réseau de neurone a été réalisée sur les rendements de riz en Thaïlande par Na-udom and Rungrattanaubol (2015). Une comparaison entre les modèles temporels et spatiaux de prédiction de rendements de culture a aussi été réalisée par Guo and Xue (2014). Enfin, Paswan and Begum (2013) ont réalisé une étude qui regroupe les possibilités des modèles de régression et de neural network dans le domaine de la prédiction de rendement. Le but serait, dans ce mémoire, de tenter d'utiliser les réseaux de neurones sur des données temporelles de cultures et d'utiliser les rotations parmi les réseaux de neurones.

1.5 Évaluation de la prédiction et de la classification

Suite à la prédiction et la classification, il est possible d'utiliser différents outils afin d'évaluer leur exactitude ainsi que leur incertitude. Nous utiliserons le mot exactitude pour mentionner si la prédiction/classification est en accord avec la réalité. Nous utiliserons le mot incertitude pour mentionner, lors d'une classification, la probabilité qu'à un pixel d'appartenir à différentes classes.

Généralement, la meilleure des façons d'évaluer la performance d'une classification ou prédiction est de les comparer avec des observations au sol. Il est aussi simplement possible de prédire une année pour laquelle on connaît déjà le parcellaire agricole pour ensuite comparer les prédictions avec notre parcellaire.

Une méthode très souvent utilisée pour visualiser les incertitudes en classification d'images satellites est le Maximum Probability qui apparaît dans les vecteurs de probabilités qui décident de la classe à assigner au pixel. Par exemple, Van der Wel et al. (1998) utilisent une carte avec une échelle de gris des maximum probability. Toujours dans le domaine des incertitudes de classification par pixel, on peut citer les méthodes de Normalized Uncertainty Criterion (NUC) basées sur l'entropie (Dehghan and Ghassemian, 2006), de root mean squared error (RMSE) (Lobell and Asner, 2004) ou encore plus récemment de l'Equivalent Reference Probability (ERP) développé par Bogaert et al. (2016).

En ce qui concerne l'exactitude, celle-ci n'était pas considérée si importante jusque fin des années 80. Selon Congalton (1991), on évaluait l'exactitude sur base d'un seul chiffre qui donnait l'exactitude globale de la classification. On se retrouvait alors avec une exactitude non spécifique qui ignorait les exactitudes en fonction de la localisation. De plus, une autre mauvaise méthode était d'utiliser les mêmes données pour entraîner le modèle et pour évaluer l'exactitude, ce qui surestimait l'exactitude. Une méthode classique qui est dès lors utilisée est la création d'une matrice d'erreurs ou matrice de confusion Congalton (1991). C'est une sorte de table de contingence en deux dimensions (actuel et prédit) qui peut être utilisée pour des données de tous types d'unités : pixel, agrégat de pixel, polygones, etc.

À partir de la matrice de confusion, il a été possible de calculer différentes valeurs statistiques qui permettent d'évaluer de manière analytique et descriptive les résultats. Parmi les méthodes descriptives il existe trois valeurs triviales : l'exactitude du producteur, de l'utilisateur et l'exactitude globale. Parmi les méthodes analytiques, on utilise surtout les techniques discrètes multivariées (Congalton et al., 1983). Une première technique est de normaliser les matrices de confusion ce qui permet d'être plus complet et de mettre au même niveau l'exactitude du producteur et de l'utilisateur. Une deuxième technique est l'analyse Kappa qui détermine si l'exactitude entre deux matrices de

confusion est significativement différente. Il analyse via une statistique nommé KHAT si l'accord entre les prédictions et la réalité n'est pas dû à la chance. Seulement, la pertinence de l'analyse Kappa a depuis été réfutée par Pontius and Millones (2011) qui pousse plutôt à l'analyse de deux paramètres : quantity disagreement et allocation disagreement. Enfin, il faut citer aussi la méthode du F-score ou F-measure conseillée par Labatut and Cherifi (2012).

2. Matériels et méthodes

2.1 Données et outils utilisés

Les données principales utilisées sont le LPIS Wallon fourni par le SIGEC dans le cadre de la PAC et mis en œuvre par l'Organisme Payeur de Wallonie (OPW). Ce LPIS connu sous la dénomination de « parcellaire agricole anonyme » (PAA) est disponible sur le Géoportail de Wallonie a priori dès fin janvier de l'année qui suit le semis de culture.

Les PAA de 2008 à 2016 ont été demandés au Service Public de Wallonie (SPW). Une couche de données du PAA est composée de polygones qui correspondent chacun à une parcelle. Chaque parcelle Wallonne est donc représentée et possède différentes informations liées à elle, dont :

- L'année de la campagne de recensement du PAA (2008, 2009, etc.)
- Un code culture formé de 1 à 4 chiffres qui correspond à un type de culture (6, 311, 71, 9811, etc.). Le nom de la culture associée est présent sous forme de texte (Prairie, Froment d'hiver, Betterave, Chicorée à inuline, etc.)
- La surface de la parcelle en m² ou en ha
- Le périmètre de la parcelle en m

D'autres données sont fournies selon l'année du PAA mais ne sont pas utiles pour ce mémoire.

Au niveau des outils, l'entièreté des moyens utilisés ont été informatiques. Le programme ArcGIS a été utilisé pour le traitement des données SIG, la visualisation de celles-ci ainsi que la création de cartes. Le programme R (R Development Core Team, 2008) avec l'interface Rstudio a été utilisé pour réaliser les différentes analyses statistiques.

2.2 Zone d'étude

Les critères de sélection du site étaient nombreux. Il fallait tout d'abord une zone homogène tant au niveau agricole qu'au niveau agro-géographique se trouvant en Wallonie. La proximité avec Louvain-la-Neuve était un avantage en cas de nécessité de validation de terrain. Il fallait ensuite un nombre de parcelles suffisant pour pouvoir en dériver des probabilités d'occurrence robuste. Enfin, il fallait une zone sujette à l'érosion et aux inondations boueuses.

Sur base de ces critères, une zone carrée de 12.5 km de côté a été sélectionnée, soit une surface de 156,25 km². Elle est située dans la région agricole limoneuse (a) et la zone agro-géographique de la Hesbaye (b) (Fig. 1) Les coordonnées de la zone sont renseignées au tableau 3.

Tableau 3. Coordonnées de la zone d'étude (Lambert 72)

	Min	Max
X	184500	197000
Y	145500	158000

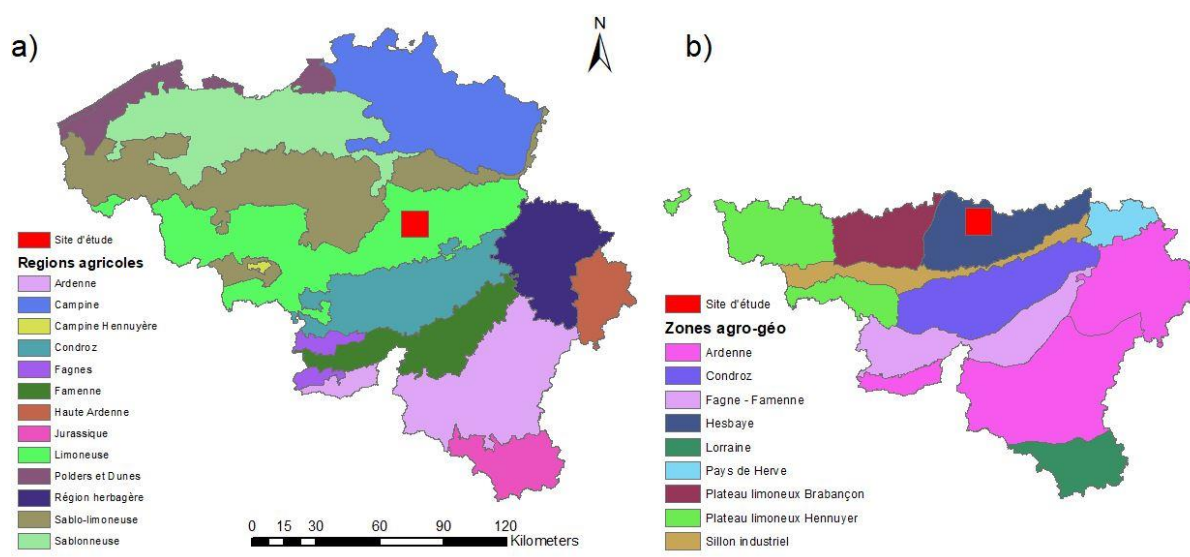


Fig. 1. Zone d'étude et agronomie. a) Régions agricoles de Belgique, b) Zones agro-géographiques de Wallonie (Sources des données : SPW¹)

La zone d'étude s'étend sur 8 communes du Brabant Wallon : Jodoigne, Orp Jauche, Hélécinne, Lincen, Ramillies, Hannut, Eghezée et Wasseiges (Fig. 2). Il y a 4835 parcelles dans notre zone d'étude. Enfin, les endroits à fortes inondations boueuses ont été déterminés grâce aux Sites d'Intérêts GISER

¹ <http://geoportail.wallonie.be>

(SIGISER). La couche de données contenant tous les SIGISER de Wallonie a été fournie par la cellule GISER du SPW. La zone compte 36 SIGISER d'une surface allant 0.02 km² à 4.6 km². La surface totale de ceux-ci est de 20.3 km², ce qui équivaut à environ 13% de la surface de la zone d'étude.

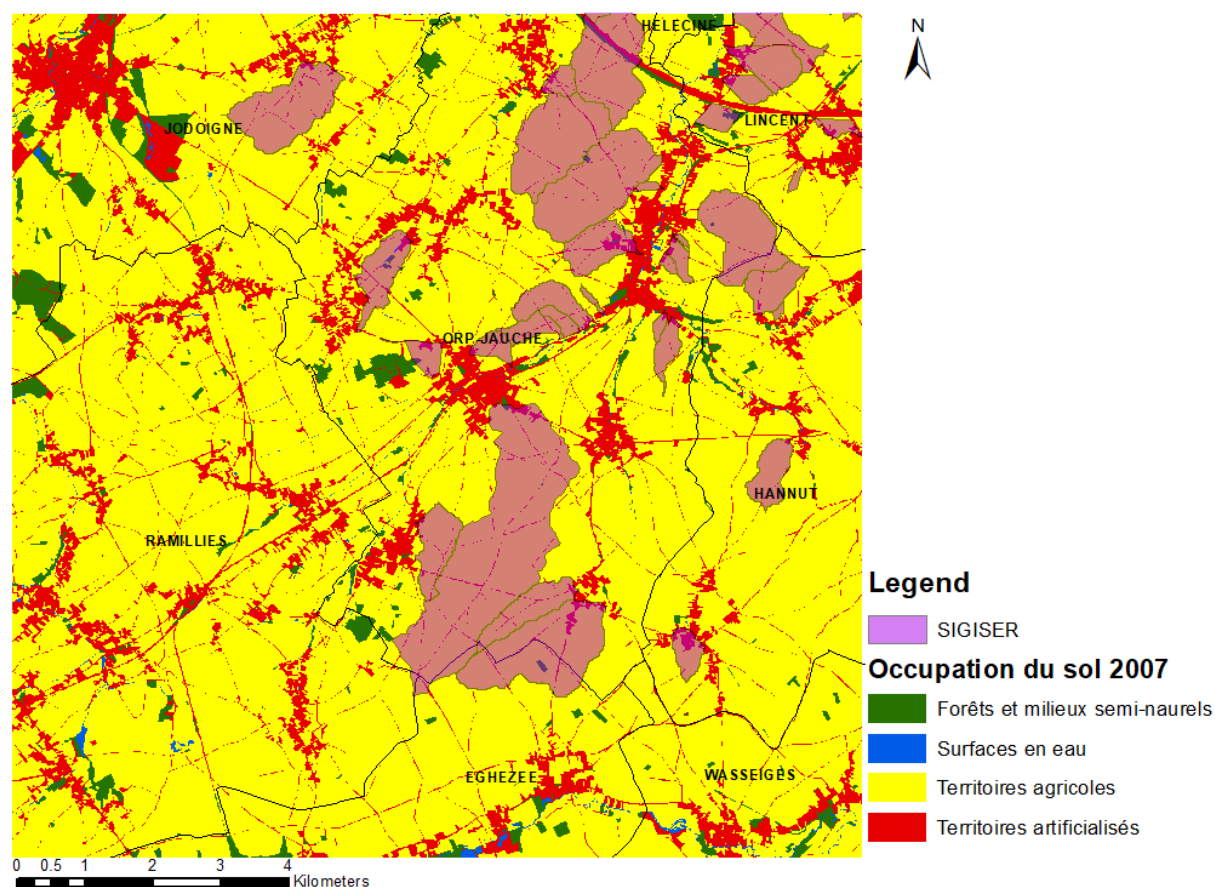


Fig. 2. Occupation du sol et SIGISER (Sources des données : SPW¹).

La Fig. 2 renseigne aussi sur l'occupation du sol de la zone d'étude détaillée dans le tableau 4. Le pourcentage de territoires agricoles est bien supérieur à la moyenne wallonne (53%) et confirme le côté agricole de la zone.

Tableau 4. Résumé de l'occupation du sol (Sources des données : SPW¹).

Occupation du sol	Aire totale (km ²)	Pourcentage de la surface totale (%)
Territoires agricoles	132.22	84.6
Territoires artificialisés	18.85	12.1
Forêts et milieux semi-naturels	4.52	2.9
Surfaces en eau	0.66	0.4

¹ <http://geoportail.wallonie.be>

Le relief est peu marqué et très progressif avec une pente moyenne de 3.97%. C'est une zone d'agriculture particulièrement peu vallonnée comparé à la pente moyenne en Wallonie qui est de 8.14%. Le plus haut point situé au sud de la zone culmine à 163 m et le point le plus bas au nord de la zone à 57 m. Au niveau hydrographique, la zone d'étude se situe principalement dans le bassin versant de la Dyle-Gette et plus précisément dans le sous-bassin de la Petite Gette. Le nord de la zone est traversé par le bassin versant de la Meuse et plus précisément par le sous bassin de la Mehaigne.

La géologie de la zone d'étude est très diverse. Les infos qui suivent proviennent de la carte géologique de Wallonie¹. La zone représente la frontière entre la Formation de Bruxelles/Sint-Huibrechts-Hern située à l'ouest et la formation de Hannut/Dormaal à l'est. En plus de ces 4 formations qui constituent la majorité, il faut compter les formations de Gulpen et de Ramillies qui apparaissent à certains endroits. Enfin, des alluvions modernes sont souvent présents à l'emplacement du réseau hydrographique.

Les types de sols trouvés dans notre zone sont quasi uniquement des sols limoneux à drainage naturel favorable. Ils possèdent une charge caillouteuse inférieure à 5%. Ces sols limoneux sont principalement des sols limoneux à horizon B textural (Aba) ou des sols sur limon (Abp) de temps en temps faiblement gleyifiés (ADp) (Sources SPW²)

Le climat dans la zone d'étude est apparenté à celui de Orp-Jauche. Celui-ci est classé comme Cfb c'est-à-dire un climat tempéré humide de type océanique. Les températures moyennes sont de 9.8°C et les précipitations annuelles de 814 mm Les mois de Juin et Juillet sont les deux mois avec les précipitations moyennes les plus importantes : 77 et 73 mm (Climate-data.org, 2012).

¹ <http://carto1.wallonie.be/geologie/viewer.htm>, SPW – DGARNE

² <http://geoportail.wallonie.be>

2.3 Analyse et traitement des données

L'analyse et le traitement des données s'est fait exclusivement sur les PAA de 2008 à 2016. Dans un premier temps, un traitement purement technique a été réalisé afin de vérifier la présence de parcelles coupées et superposées au sein des PAA. Deuxièmement, l'évolution des PAA selon les années a été analysée et comparée selon les zones géographiques (Wallonie, Hesbaye, zone d'étude) du point de vue de leur surface et du nombre de parcelles. Troisièmement, les types de cultures ont été analysés selon les années dans notre zone d'étude, et une harmonisation des codes cultures a été réalisée. Une comparaison de ces codes harmonisés selon leurs zones géographiques a également été faite. Enfin, les différentes méthodes d'assemblages des PAA sont expliquées. Le but est d'obtenir une seule couche de données qui rassemble toutes les successions de cultures selon les années pour une même parcelle. A partir de cette couche, les rotations de cultures typiques sont analysées.

2.3.1 Parcelles en bordure de zone coupées

Vu la forme carrée de la zone d'étude, un certain nombre de parcelles situées en bordure de la zone se sont vu tronquées. Ces parcelles ont été identifiées en utilisant la fonction « select by location ». L'élimination de ces parcelles sur chaque PAA entraînait le retrait de parcelles souvent différentes d'une année à l'autre. Lors de l'étape d'intersection des PAA, cela pourrait conduire à la perte d'un grand nombre de parcelles. Cette étape n'a donc finalement pas été réalisée afin de garder un échantillon complet.

2.3.2 Superposition de parcelles au sein des PAA

Pour chaque PAA individuel de 2008 à 2016, il existe, au sein de chacune des couches, des parcelles qui se superposent. Trois cas de figure peuvent se présenter. Tout d'abord, un recoupement entre certaines arrêtes de parcelles mal délimitées (Fig. 3a). Ensuite, certaines parcelles dupliquées (Fig. 3b). Enfin, le troisième cas correspond à de nouvelles parcelles entières qui se superposent à une partie d'autres parcelles (Fig. 3c).

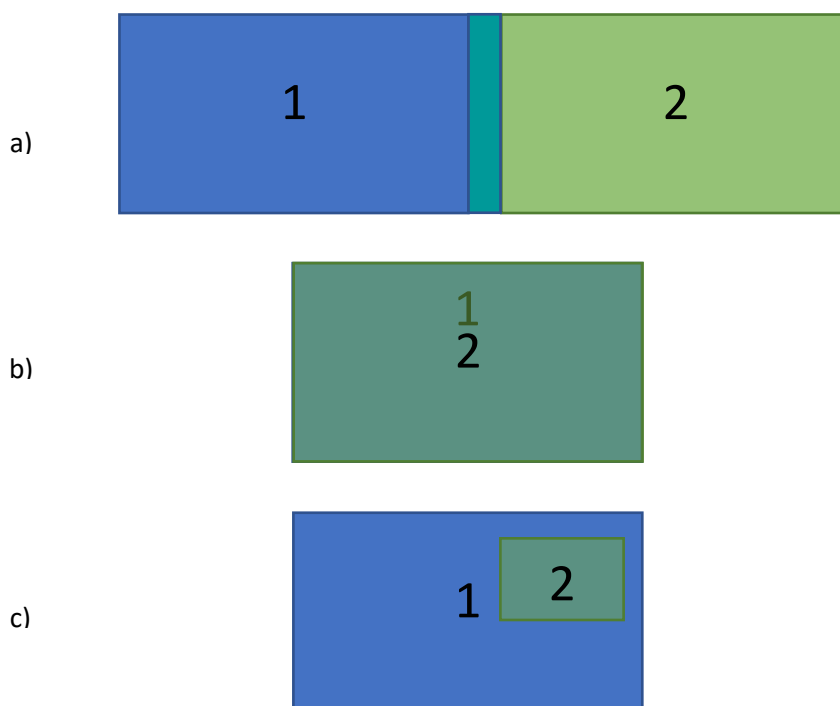


Fig. 3. Les 3 exemples d'erreurs de superpositions

Ces soucis de superposition peuvent influencer sur le nombre d'intersections des PAA de plusieurs années. La fonction « intersect » d'ArcGIS a été utilisée individuellement sur les PAA de chaque année afin d'évaluer ces erreurs. La correction a pu être faite pour les deux derniers cas en supprimant une des deux parcelles dans le cas de la duplication et en supprimant la plus petite parcelle dans le 3^e cas.

2.3.3 Évolution de la surface des PAA et du nombre de parcelles pour plusieurs zones géographiques

L'évolution de la surface des parcelles cultivées et du nombre de parcelles de 2008 à 2016 a été évaluée à partir des PAA au niveau de la Wallonie, de la Hesbaye et de notre Zone d'étude. Le but est de montrer que la zone d'étude représente mieux la Hesbaye que l'ensemble de la Wallonie. Mais aussi de vérifier la qualité des données des PAA dans leur ensemble.

Afin de mieux comprendre les informations contenues dans les PAA, la « Notice explicative de la déclaration de superficie et demandes d'aides 2015 » a été consultée¹. De plus, l'OPW² a été contacté afin d'avoir des informations complémentaires sur la création de la couche de données des parcellaires agricoles anonymes.

2.3.4 Évolution des types de cultures

L'évolution selon les années des types de cultures sur la zone d'étude a été analysée. Le but est de vérifier la situation stationnaire des PAA de 2008 à 2016 ce qui permet de prédire les années futures. En effet, des cultures qui varient énormément chaque année au niveau de leur représentation ne permettent pas d'observer de rotations claires dans les cultures.

2.3.5 Harmonisation et sélection des cultures

Les couches du PAA n'utilisent pas toujours précisément le même code culture d'une année à l'autre. En effet, l'administration a réalisé des petits changements chaque année afin que les agriculteurs mentionnent avec plus de précision leurs cultures. Par exemple, certaines années les prairies temporaires, permanentes et les tournières enherbées étaient distinctes avec leur propre code culture et d'autres années elles étaient reprises sous un même code culture. Ces différences de code pouvant entraîner des erreurs de prédiction ; un travail d'harmonisation a donc été effectué.

Après avoir harmonisé les codes cultures, il fallait sélectionner les cultures les plus importantes parmi plus de 40 cultures différentes. Pour ce faire, pour chaque année, le top 20 des cultures est réalisé. Cependant, les 20 cultures les plus importantes ne sont pas forcément les mêmes chaque année. Au final, ce sont 22 codes culture qui sont gardés. En plaçant le reste des cultures dans un code « Autre », le nombre de codes cultures final s'élève à 23.

Pour changer nos codes cultures, la fonction « select by attributes » d'ArcGIS a été utilisée afin de sélectionner les cultures qui doivent changer. Ensuite, la fonction « calculate geometry » a été utilisée afin de changer les codes cultures.

2.3.6 Comparaison des types de cultures pour plusieurs zones géographiques

Un calcul de la surface des parcelles de 2016 a été réalisé afin de connaître la proportion des 23 cultures dans notre zone, en Hesbaye et en Wallonie. Pour chaque zone géographique, les différences ont été mises en avant. Enfin, pour chaque culture, la taille moyenne des parcelles a été calculée.

¹ Information gracieusement fournie par M. G. Collignon, Ferme expérimentale De Marbaix, UCL.

² Mr. Nicolas Simon

2.3.7 Assemblage des différentes couches des PAA

L'assemblage des couches des PAA des différentes années est nécessaire afin d'obtenir une seule couche de données reprenant les infos des PAA de 2008 à 2015 pour chaque parcelle. C'est en effet les successions des cultures de ces années qui permettront d'en sortir les tendances afin de prédire l'année 2016. L'assemblage des différentes couches des PAA peut se faire de deux manières. L'utilisation de la fonction « Union » d'ArcGIS ou l'utilisation de la fonction « Intersect ».

La fonction « Union » prend en compte tous les détails possibles qui concernent l'évolution des parcelles au cours des ans. Elle prend donc aussi en compte l'apparition et la disparition de certaines parcelles (Fig.4). Par exemple, une parcelle de 2015 et une parcelle de 2014 avec un bout en moins formeront 2 parcelles avec la fonction union. La nouvelle parcelle (N°2 dans la Fig. 4) ne possèdera pas d'information sur 2014. Cependant, l'apparition ou disparition des parcelles est très compliquée à prédire et ne suit pas forcément de logique comme une rotation de culture. Cette méthode a donc été écartée.

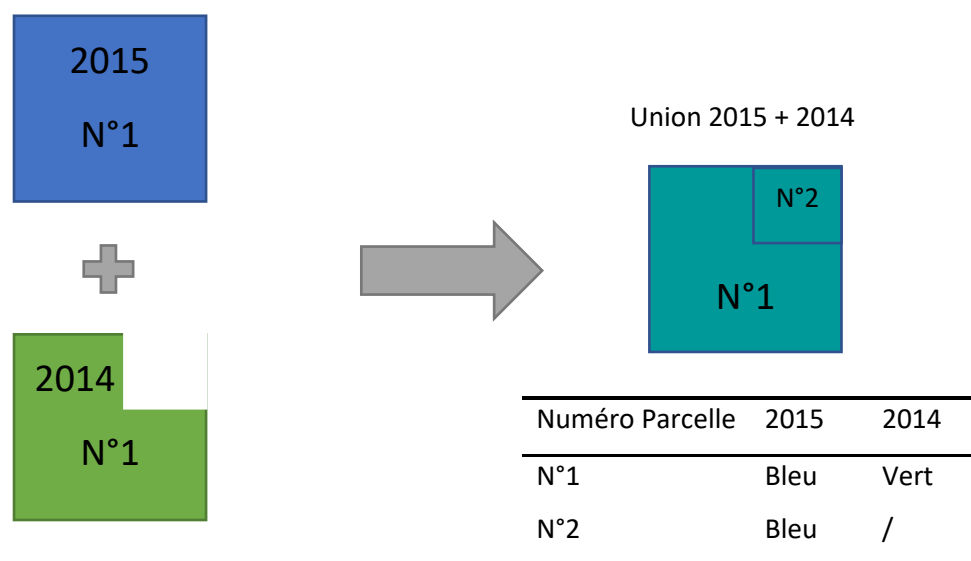


Fig. 4. Schéma explicatif de la fonction "Union" d'ArcGIS

La méthode « Intersect » ne garde pas les parcelles qui disparaissent ou apparaissent en fonction des années. Cette méthode ne conserve que l'intersection stricte des parcelles. Donc si un morceau ou une parcelle n'est pas présent une année, elle ne sera pas présente sur la couche finale intersectée et il n'y aura pas de nouvelle parcelle créée (Fig.5.).

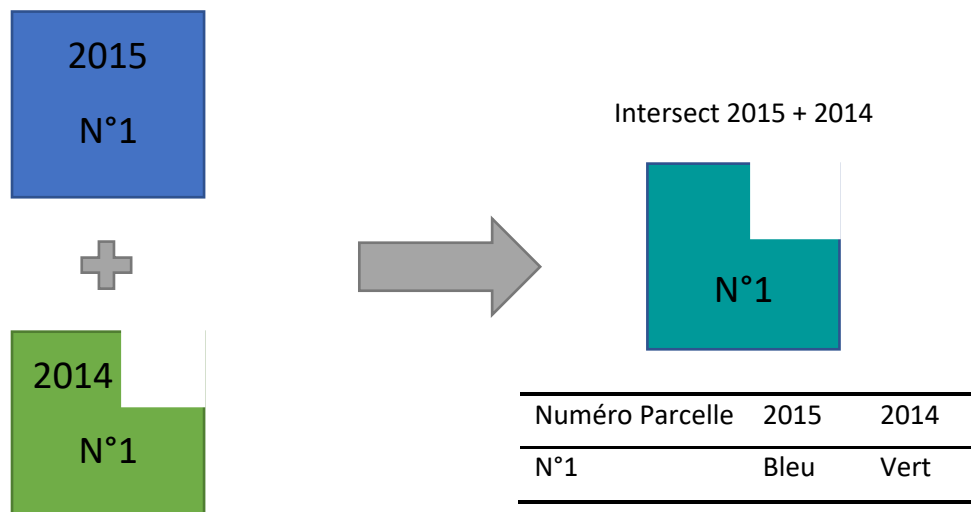


Fig. 5. Schéma explicatif de la fonction "Intersect" d'ArcGIS

Le souci rencontré avec la fonction « Intersect » est qu'elle peut créer de nouvelles parcelles aussi d'une autre manière. Soit suite à la division de grandes parcelles en plusieurs autres en fonction des années, soit lorsque deux parcelles normalement identiques entre 2 années possèdent des délimitations qui diffèrent même légèrement (voir Fig. 6). Le problème de ces parcelles supplémentaires créées est qu'elles « polluent » potentiellement l'information des rotations de cultures. En effet, elles ne suivent pas nécessairement une logique de succession de cultures telle que pensée par l'agriculteur.

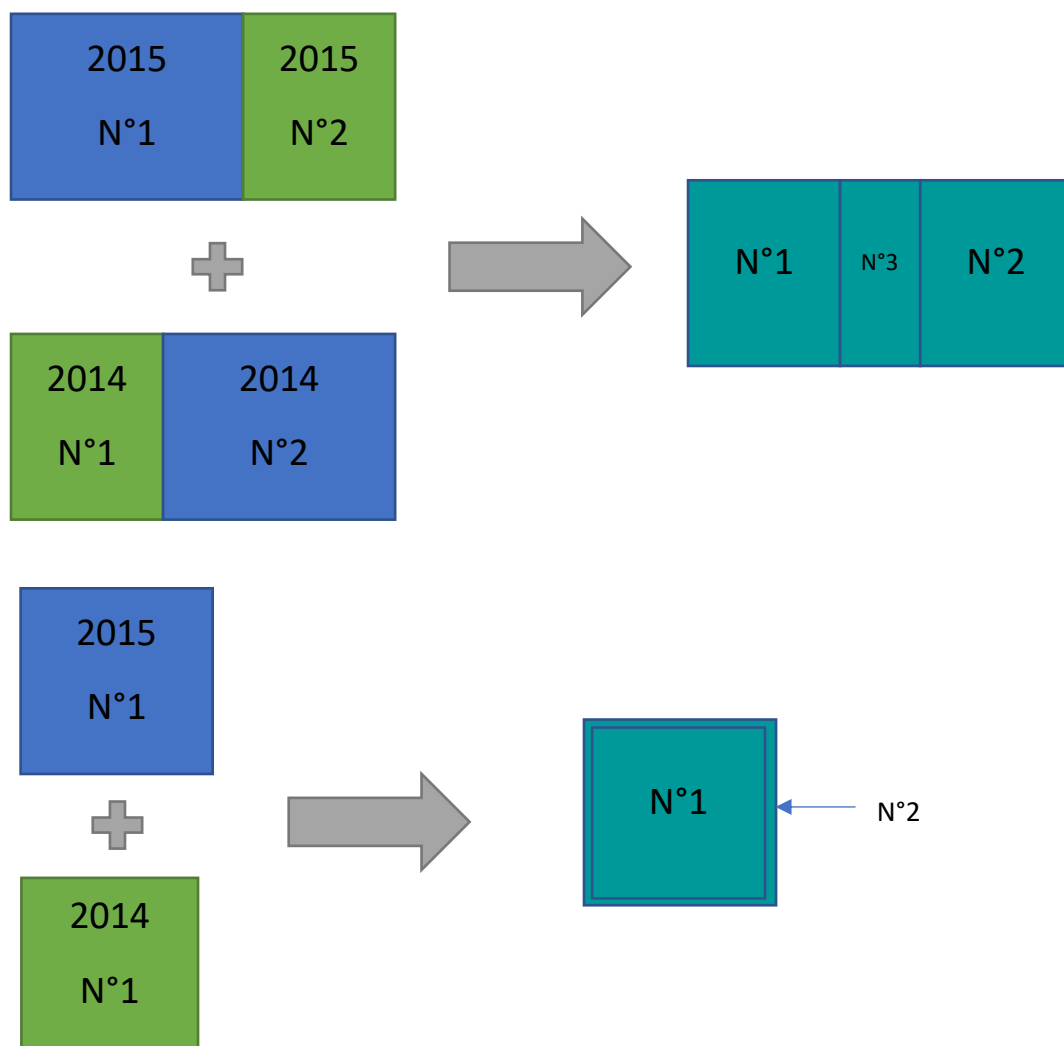


Fig. 6. Erreurs dues à la fonction « Intersect ».

La solution à ce souci fut d'utiliser une méthode appelée des « Centroides » (Fig. 7). Cette méthode utilise le centroïdes de chaque parcelle de 2015 grâce à la fonction « feature to point » d'ArcGIS. Ensuite il suffit de faire l'intersection des centroïdes de 2015 et des parcelles précédentes.

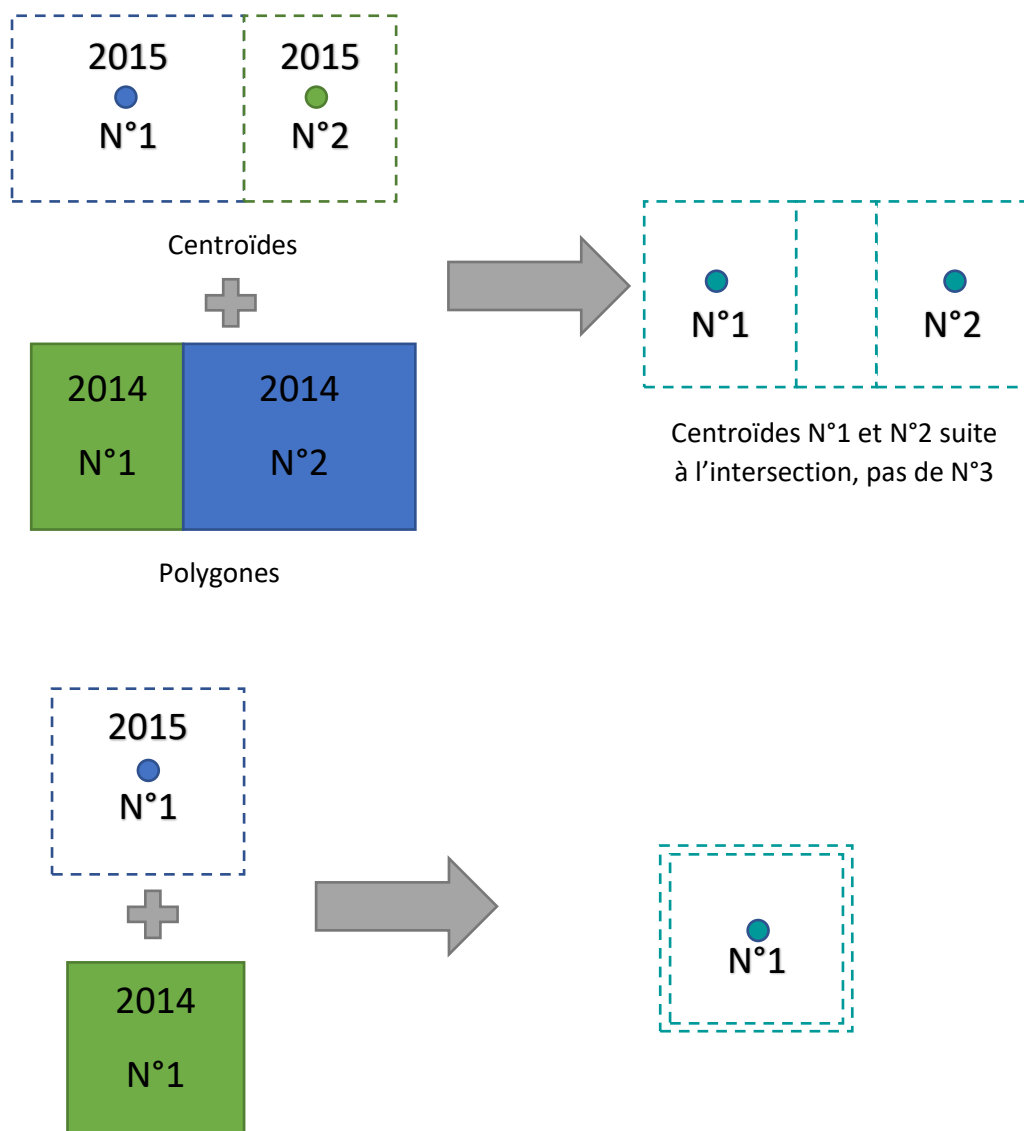
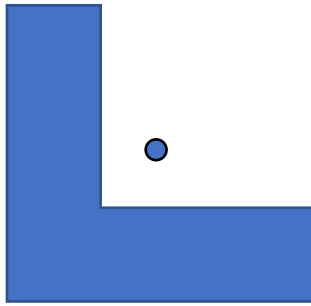


Fig. 7. Méthode des centroïdes avant intersection

Une limitation remarquée avec la méthode des centroïdes est que certains centroïdes ne se trouvent pas à l'intérieur des polygones à cause de la forme particulière des parcelles. Ces centroïdes peuvent soit se retrouver « dans le vide », soit se retrouver à l'intérieur d'un autre polygone adjacent. La première situation peut faire perdre des parcelles lors de l'intersection de plusieurs années, et la seconde situation pose le problème du choix lors de la détermination de la culture si les centroïdes possèdent une information différente. L'option « Inside » de la fonction a ainsi été utilisée afin que tous les centroïdes soient localisés à l'intérieur d'un polygone de 2015 (voir Fig. 8).

Sans l'option « inside »



Avec l'option « inside »

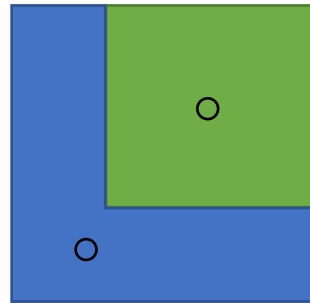
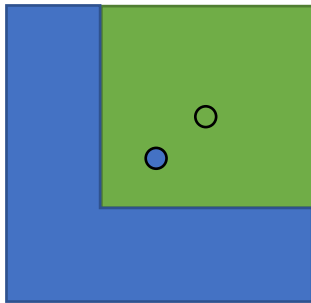
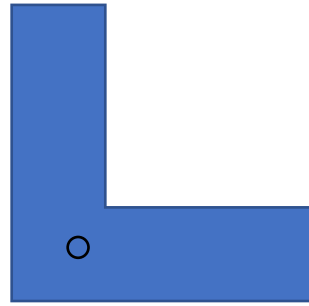


Fig. 8. Méthode des centroïdes : différence avec ou sans l'option « Inside »

À partir des centroïdes « Inside » de 2015, l'intersection est faite tout d'abord avec les polygones de 2014. En résultent des centroïdes « 2014_2015 » qui sont intersectés avec les polygones de 2013. Et ainsi de suite jusqu'au PAA de 2008. Le résultat final comprend seulement les centroïdes originaires de 2015 qui intersectent chaque année jusqu'en 2008.

2.3.8 Rotation de cultures

Pour 2 à 5 années successives, les rotations de cultures sont analysées. Pour chaque nombre d'années, les observations sont combinées afin de construire la base de données la plus riche possible (Tableau 5). Ainsi, pour construire la matrice de transition des rotations biennales, nous utiliserons les paires d'années 2015-2014, 2014-2013, ..., et 2009-2008. Cette approche suppose que la logique qui sous-tend les rotations n'a pas évolué sensiblement au cours de la période 2008-2015, c'est-à-dire que la proportion des différentes cultures au cours de la période est stationnaire.

Tableau 5. Données d'étude des rotations de cultures

2 années	3 années	4 années	5 années
2014_2015	2013_2014_2015	2012_2013_2014_2015	2011_2012_2013_2014_2015
2013_2014	2012_2013_2014	2011_2012_2013_2014	2010_2011_2012_2013_2014
2012_2013	2011_2012_2013	2010_2011_2012_2013	2009_2010_2011_2012_2013
2011_2012	2010_2011_2012	2009_2010_2011_2012	2008_2009_2010_2011_2012
2010_2011	2009_2010_2011	2008_2009_2010_2011	
2009_2010	2008_2009_2010		
2008_2009			
7 obs. comb. ¹	6 obs. comb.	5 obs. comb.	4 obs. comb.

1. Observations combinées par années

2.4 Prédiction par Matrices de transition

2.4.1 Qu'est-ce qu'une matrice de transition ?

C'est une matrice rectangulaire stochastique qui est liée aux chaînes de Markov. Les chaînes de Markov sont des processus qui donnent pour un état présent la probabilité de passage à un état futur. Ces processus ne sont pas dépendants des états antérieurs (pas de mémoire). Les chaînes de Markov peuvent être représentées par des graphes probabilistes.

La figure 9 représente le graphe probabiliste et la matrice de transition d'un exemple (non réel) avec 3 états possibles correspondant à des cultures : froment, betterave et prairie. Le présent correspond à l'année 2015 et le futur à l'année 2016 par exemple.

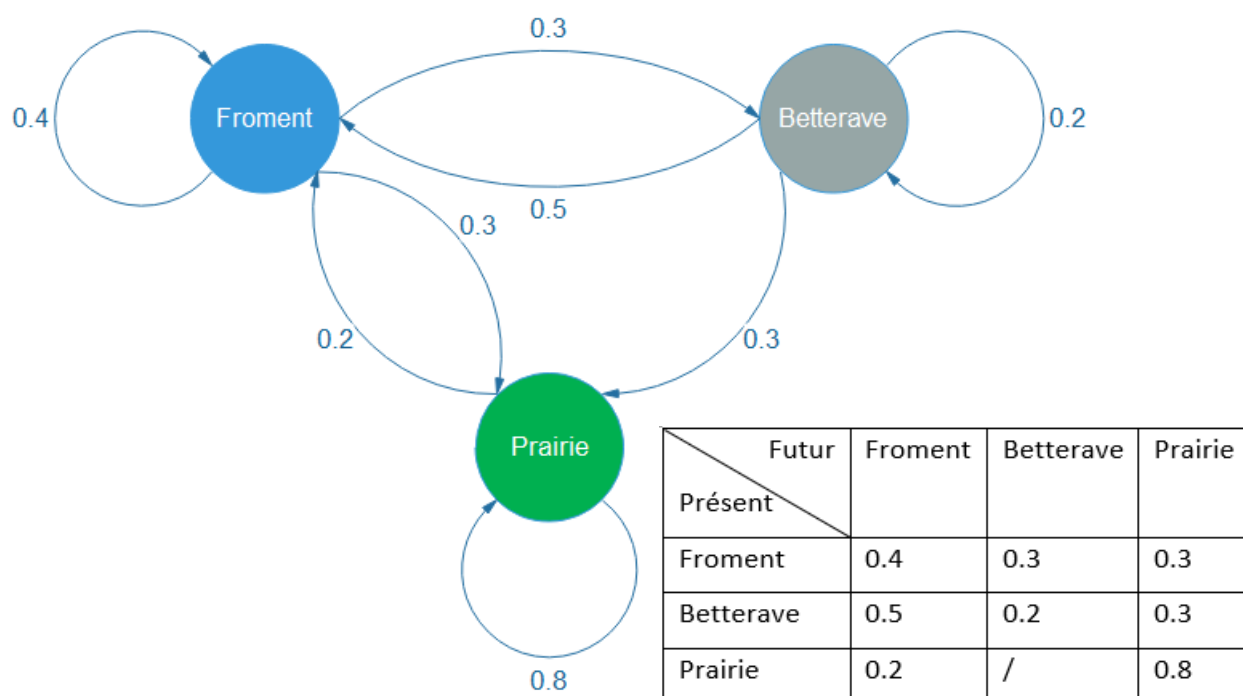


Fig. 9. Chaînes de Markov : exemple fictif de graphe probabiliste et matrice de transition

La matrice se lit de gauche à droite ; il suffit de choisir la ligne de la culture présente afin de connaître la probabilité de la future culture. Par exemple, la betterave a 50% de chance de devenir du froment l'an prochain. Dans certains cas, la culture n'a aucune probabilité de devenir une autre culture, comme c'est le cas ici avec les prairies qui ne peuvent pas devenir betterave. Enfin, la somme de toutes les probabilités d'une ligne correspond à l'unité. A partir de cette matrice de transition, il est possible de réaliser une table de prédiction la plus probable en gardant les probabilités de transition les plus élevées (Tableau 6).

Tableau 6. Exemple de table de prédiction

Présent (2015)	Futur (2016)
Froment	Froment
Betterave	Froment
Prairie	Prairie

Comme les chaines de Markov n'utilisent que l'état présent afin de prédire l'état futur, cela pose problème pour utiliser l'information relative à plusieurs années précédentes. La solution est de « coller » les informations des années précédentes afin qu'elles soient considérées comme une seule information représentant le « présent ». L'état présent représente alors des combinaisons de plusieurs cultures. Un exemple pour 3 cultures est repris au tableau 7 pour les combinaisons de 2 années. Il peut bien sûr y avoir des combinaisons d'encore plus d'années précédentes ; cependant le nombre de combinaisons possibles augmentera de façon exponentielle.

Tableau 7. Matrice de transition fictive : exemple de combinaison de cultures

Présent (2014-2015)	Futur (2016)		
	Froment	Betterave	Prairie
Froment – Froment	0.4	0.4	0.2
Froment – Betterave	0.7	0.1	0.2
Froment – Prairie	0.2	0.2	0.6
Betterave – Froment	0.4	0.4	0.2
Betterave – Betterave	0.6	0.2	0.2
Betterave – Prairie	0.5	0.2	0.3
Prairie – Froment	0.6	0.3	0.1
/	/	/	/
Prairie – Prairie	0.1	0.1	0.8

Le tableau se lit à nouveau de gauche à droite en allant de l'année la plus ancienne à la plus récente. Dans cet exemple, les valeurs précises n'ont pas été mentionnées. Cependant, il faut remarquer que

certaines combinaisons ne se produisent pas. Tout comme les prairies ne devenaient pas betterave lors de la matrice précédente, ici la combinaison Prairie-Betterave n'existe pas. C'est un souci qui n'est généralement pas lié à la méthode mais au jeu de données. Celui-ci sera rencontré lors de la partie validation de ce travail.

2.4.2 Résumé de la méthodologie

Afin de prédire la culture de l'année 2016 via les matrices de transition, il y a 2 phases à suivre : une phase de calibration et une phase de validation.

La phase de calibration utilise les données de 2008 à 2015 comme inputs (année « présente ») et output (année « future »). Par exemple, dans le cas d'une prédiction basée sur une seule année antérieure, la matrice de confusion et la table de prédiction sont construites pour chaque paire d'années possible à partir d'inputs et d'outputs. Cette table de prédiction est ensuite appliquée sur l'input utilisé pour créer la matrice de transition. Le nouvel output prédit est comparé à l'output réel. De cette façon, nous évaluons la qualité de la calibration.

Afin de prédire 2016, la phase de validation utilise les données de 2015 et antérieures comme inputs et les matrices de transition de la calibration. Le schéma à la Fig.10 résume l'ensemble de notre approche, les phases de calibration et validation seront détaillées par la suite.

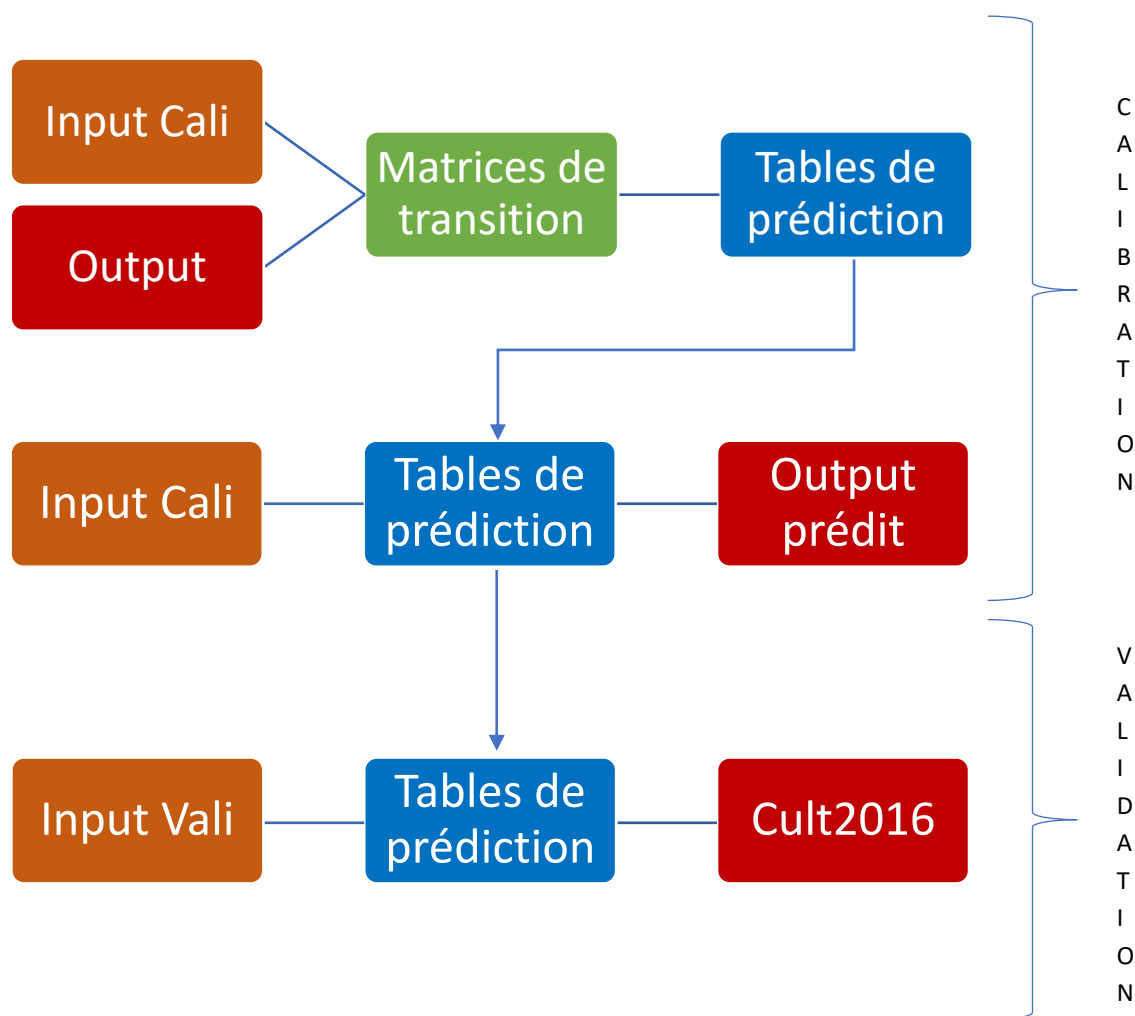


Fig. 10. Résumé de la méthodologie : Calibration et Validation

2.4.3 Calibration

Dans la calibration, les matrices de transition sont créées. Ces matrices sont créées grâce à la fonction « table » de R. Cette partie utilise des inputs Cali et des outputs observés afin de créer les matrices de transition. Pour les inputs, comme vu précédemment, il est possible de considérer des combinaisons d'années antérieures. Dans ce travail, le nombre de combinaisons maximum sera de 5 années. Pour les outputs, ceux-ci changent en fonction de l'année la plus récente des inputs utilisée (tableau 8). Par exemple, lorsque l'année la plus récente de l'input est 2008, l'output est 2009, et ainsi de suite. Il y a donc en tout 25 matrices de transition créées. Chacune de ces matrices de transition permettent de créer une table de prédiction.

Tableau 8. Calibration : Inputs et outputs et création de matrices de transitions

INPUTS					OUTPUTS
1 an	2 ans	3 ans	4 ans	5 ans	
2014	2013_2014	2012_2013_2014	2011_2012_2013_2014	2010_2011_2012_2013_2014	2015
2013	2012_2013	2011_2012_2013	2010_2011_2012_2013	2009_2010_2011_2012_2013	2014
2012	2011_2012	2010_2011_2012	2009_2010_2011_2012	2008_2009_2010_2011_2012	2013
2011	2010_2011	2009_2010_2011	2008_2009_2010_2011		2012
2010	2009_2010	2008_2009_2010			2011
2009	2008_2009				2010
2008					2009



Matrices de transition				
1 année	2 années	3 années	4 années	5 années
2014_2015	2013_2014_2015	2012_2013_2014_2015	2011_2012_2013_2014_2015	2010_2011_2012_2013_2014_2015
2013_2014	2012_2013_2014	2011_2012_2013_2014	2010_2011_2012_2013_2014	2009_2010_2011_2012_2013_2014
2012_2013	2011_2012_2013	2010_2011_2012_2013	2009_2010_2011_2012_2013	2008_2009_2010_2011_2012_2013
2011_2012	2010_2011_2012	2009_2010_2011_2012	2008_2009_2010_2011_2012	
2010_2011	2009_2010_2011	2008_2009_2010_2011		
2009_2010	2008_2009_2010			
2008_2009				

Ci-dessous est schématisé comme exemple la phase de calibration pour deux années d'inputs (2010_2011). La table de prédiction est utilisée sur l'input 2010_2011. Les cultures de 2012 prédites sont ensuite comparées aux valeurs réelles de 2012 pour tous les points d'intersection.

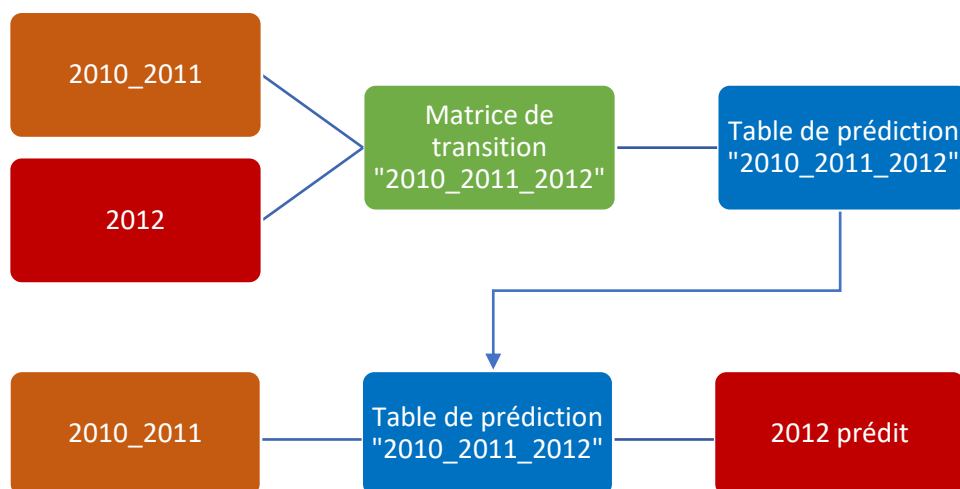


Fig. 11. Exemple de Calibration : input à 2 années antérieures (2010_2011)

2.4.4 Validation

Dans ce travail, la validation a pour but de prédire les cultures de 2016. L'input de validation renseigné dans le tableau 9 dépend du nombre d'années de la matrice de transition utilisée pour la prédiction. Chaque matrice créée lors de la calibration est testée sur l'input qui lui correspond. De plus, les matrices individuelles par année sont fusionnées pour donner une matrice fusionnée de 1, 2, 3, 4 et 5 ans. Ces matrices fusionnées seront utiles dans le cas où il n'est pas possible de prédire toutes les cultures (voir tableau 7). Il y aura donc par exemple 8 résultats pour les matrices qui se basent sur 1 année antérieure et 4 résultats pour les matrices à 5 années antérieures.

Tableau 9. Validation : matrices de transition et inputs associés en vue de prédire les cultures de 2016

1 année	2 années	3 années	4 années	5 années
Matrices de transition				
2014_2015	2013_2014_2015	2012_2013_2014_2015	2011_2012_2013_2014_2015	2010_2011_2012_2013_2014_2015
2013_2014	2012_2013_2014	2011_2012_2013_2014	2010_2011_2012_2013_2014	2009_2010_2011_2012_2013_2014
2012_2013	2011_2012_2013	2010_2011_2012_2013	2009_2010_2011_2012_2013	2008_2009_2010_2011_2012_2013
2011_2012	2010_2011_2012	2009_2010_2011_2012	2008_2009_2010_2011_2012	
2010_2011	2009_2010_2011	2008_2009_2010_2011		
2009_2010	2008_2009_2010			
2008_2009				
Fusion 1 an	Fusion 2 ans	Fusion 3 ans	Fusion 4 ans	Fusion 5 ans
Inputs				
2015	2014_2015	2013_2014_2015	2012_2013_2014_2015	2011_2012_2013_2014_2015

La figure 12 ci-dessous reprend la suite de l'exemple à la figure 11 pour la validation. Cet exemple met en avant le fait que n'importe quelle table de prédiction des années précédentes peut être utilisée afin de prédire 2016. Une comparaison sera d'ailleurs faite pour le démontrer.



Fig. 12. Exemple de Calibration : input à 2 années antérieures (2014_2015)

Cette validation peut se faire tout d'abord en utilisant les points d'intersection utilisés lors de la calibration comme input afin de pouvoir comparer les résultats de calibration et validation. Ces points d'intersection de 2008 à 2015 possèdent l'information de la prédiction de 2016. Ils sont intersectés avec les polygones de 2016 afin de pouvoir être comparés à la réalité (Fig. 13).

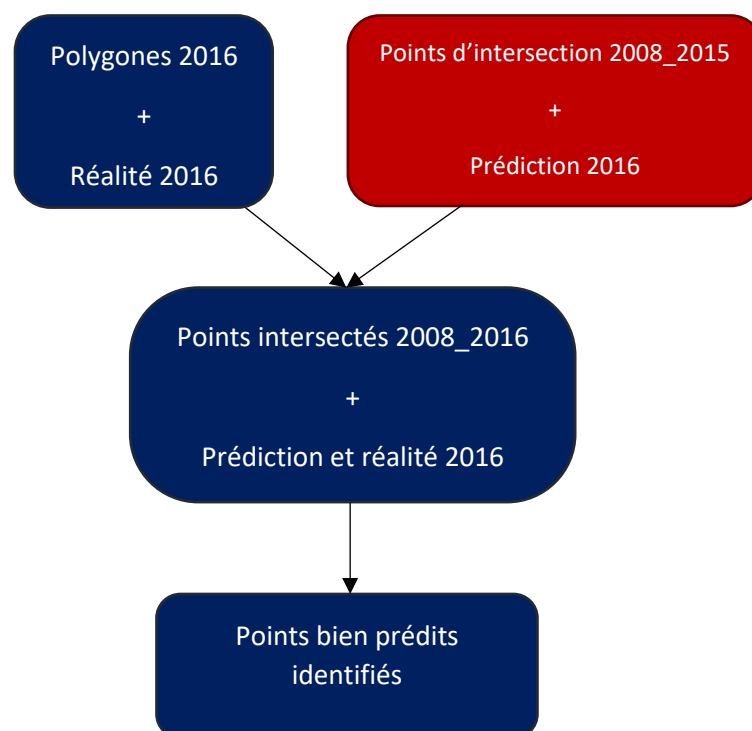


Fig. 13. Validation appliquée aux points d'intersection

L'autre approche de vérification des résultats est d'utiliser les polygones intersectés comme input afin de connaître la surface de cultures bien prédite (Fig.14). Les polygones intersectés correspondent aux PAA intersectés sans utiliser la méthode des centroïdes. Ensuite, l'intersection entre les prédictions et les polygones de 2016 possédant la réalité est réalisée.

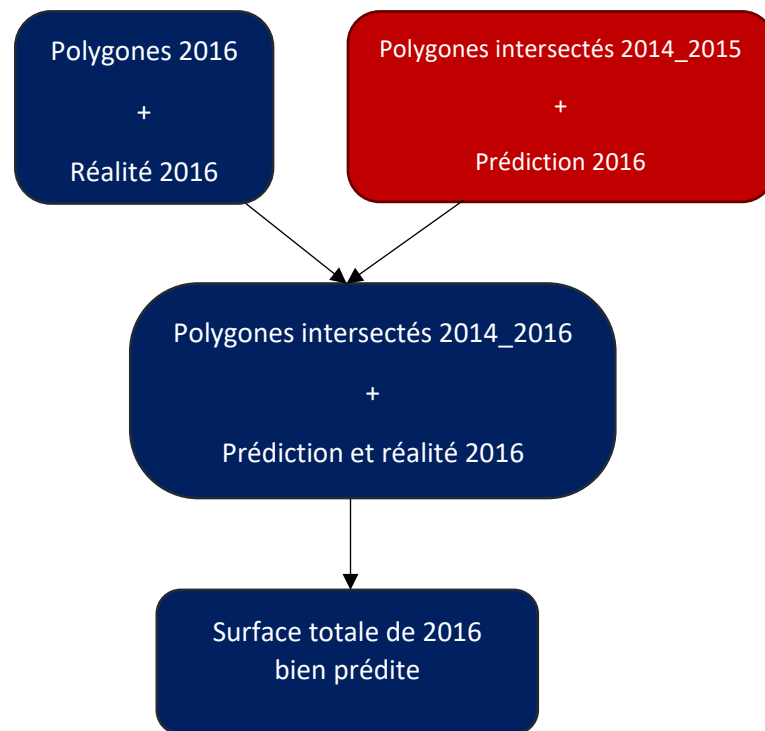


Fig. 14. Validation appliquée aux polygones intersectés

A partir de la validation sur les points d'intersection, une matrice de confusion a été réalisée. Celle-ci met en évidence différentes valeurs qui permettent d'analyser la prédiction. On y retrouve par exemple la sensibilité ou la précision. Ces valeurs permettent de calculer le score F1 ou encore le coefficient de Kappa.

2.4.5 Application aux groupes de cultures

La prédiction des cultures a aussi été tentée en réduisant le nombre de codes cultures. Au lieu de chercher à prédire 23 codes cultures, ceux-ci ont été réduits à 4 groupes de cultures. Ces 4 groupes de cultures sont les cultures d'hiver, de printemps, les prairies et les autres. Ces regroupements peuvent être réalisés de deux manières. Soit un regroupement après prédiction des 23 types de cultures par les matrices de transition. Soit un regroupement avant prédiction et avant la création des matrices de transition.

De manière générale, la méthodologie reste la même que pour la prédiction des 23 types de culture. Au niveau de la calibration, celle-ci a été réalisée afin de créer la matrice de transition mais n'a pas été évaluée. Suite à cela, la validation a été appliquée aux points d'intersection. Ces résultats ont permis la création de matrices de confusion afin de vérifier l'erreur de prédiction. Enfin, la validation a aussi été appliquée sur les polygones intersectés afin de connaître la surface bien prédite.

2.5 Prédiction par Réseau de neurones

Cette partie a pu être réalisée en utilisant le package 'neuralnet' dans R (Fritsch et al., 2016).

2.5.1 Mise en place du système de réseau de neurones

Tout comme pour les matrices de prédiction, l'approche par réseau de neurone fait intervenir une phase de calibration et une phase de validation. La phase de calibration utilise les données avant 2015 comme input et les données de 2015 comme output. La phase de validation utilise les données avant 2016 comme input et utilise les réseaux de neurones calculés lors de la calibration afin de prédire les cultures pour 2016. Le schéma à la Fig.15 résume l'ensemble de notre approche et met en avant un exemple qui utilise seulement 1 année précédente comme input.

Comme montré à la Fig.15, les inputs qui peuvent être utilisés pour une a plusieurs années précédentes sont : les codes cultures de chaque parcelle, les aires de chaque parcelle et les périmètres de chaque parcelle. Les outputs sont toujours les codes cultures de l'année suivante. L'utilisation de l'aire et du périmètre des parcelles des années précédentes pourrait permettre d'affiner la prédiction de la prochaine culture. En effet, certaines cultures sont typiquement associées à une aire particulièrement importante comme la pomme de terre. En ce qui concerne le périmètre, il peut permettre de distinguer deux parcelles sur base de leurs formes. Par exemple deux parcelles, l'une carré et l'autre rectangulaire très allongée. Si celles-ci possèdent la même aire, la rectangulaire possède un plus grand périmètre. Cette caractéristique est intéressante par exemple dans le cas des cultures « prairies et fourrages » qui incluent beaucoup de bandes enherbées.

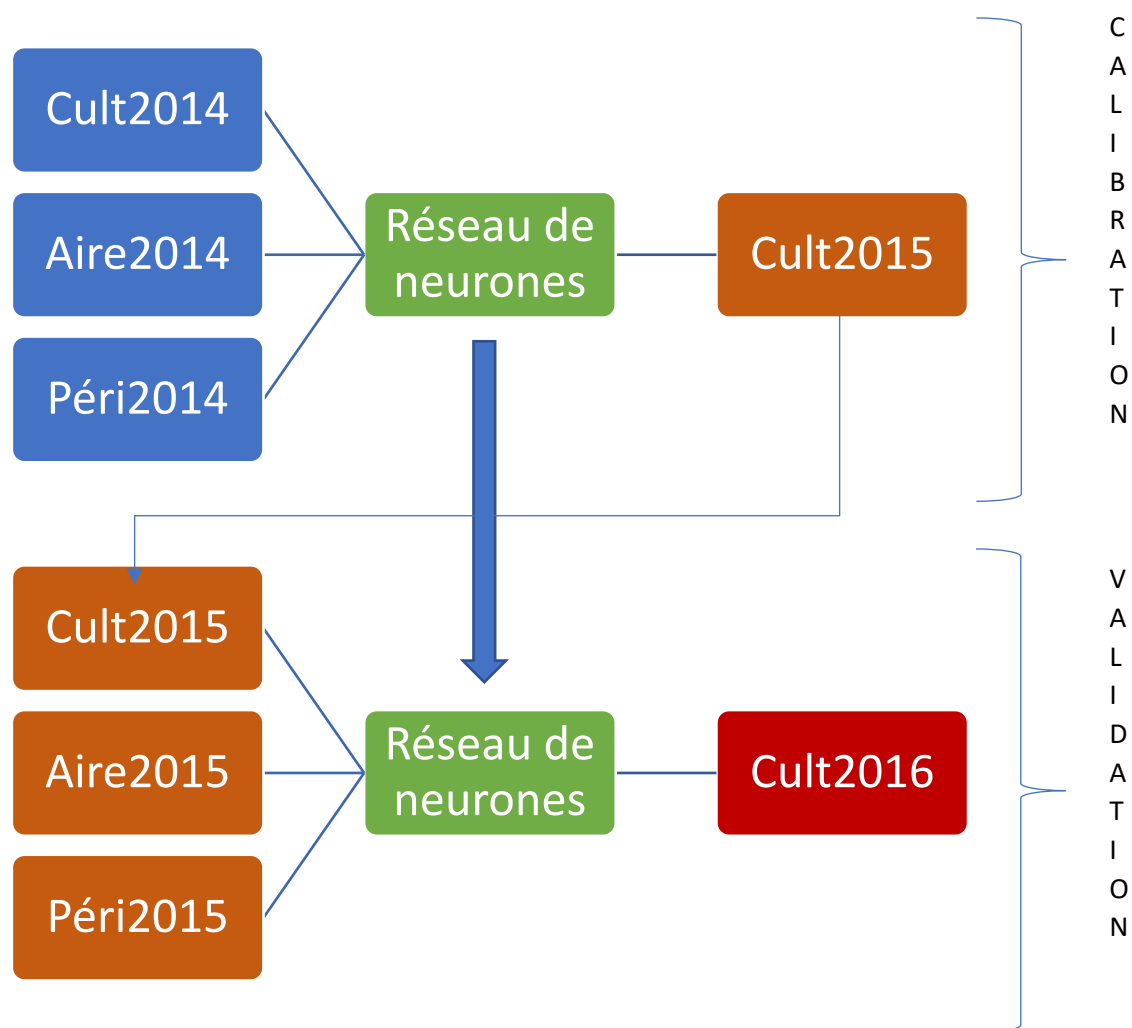


Fig. 15. Schéma général de la méthode des réseaux de neurones.

Cependant, tous les types de données ne sont pas utilisables dans le cadre des réseaux de neurones. Les données de codes cultures sont des variables catégorielles. Elles représentent un nombre fini de catégories sans ordre hiérarchique logique. Les variables codes cultures ont donc dû être transformées en variables binaires (tableau 10). Pour chacun des codes cultures, un input est créé. Pour une année donnée, chaque culture est alors reprise dans une colonne distincte. Ces inputs possèdent deux possibilités : 0 ou 1. Par exemple pour une parcelle n°1 qui correspond au code 311 :

Tableau 10. Exemple de binarisation d'une parcelle avec le code 311.

	201	202	311	...	931	951	Autre
Parcelle n°1	0	0	1	0	0	0	0

Les données représentant l'aire et le périmètre sont des valeurs continues. Elles sont donc parfaites pour les réseaux de neurones. Cependant, afin que ces données soient dans la même fourchette de valeurs que les codes cultures et afin que le réseau de neurone fonctionne de manière plus efficace,

celles-ci ont été normalisées entre les valeurs [0:1]. Pour ce faire, la fonction de normalisation « min-max » a été utilisée :

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (3)$$

2.5.2 Phase de calibration

Le but est de trouver ici un réseau de neurones qui représente au mieux la succession de culture d'une année à l'autre. Il nous faut donc des données qui représentent bien la succession des cultures sans que le nombre de parcelles change sur notre site d'étude. Afin d'avoir un suivi réel des parcelles dans le temps, les centroïdes des parcelles de 2015 sont intersectés avec les parcelles des années précédentes. Ainsi, chaque centroïde possède ses outputs (2015) et inputs (années précédentes). L'utilisation des centroïdes permet, contrairement aux polygones, qu'il n'y ait pas de segmentation des parcelles lors des intersections. La succession des cultures et leurs occurrences représentent donc plus la réalité avec les centroïdes que l'utilisation de morceaux de parcelles avec les polygones.

Suite à la transformation des données de code cultures, la phase de calibration utilisant seulement 1 année précédente à légèrement changé (voir Fig.16). Lors de cette phase, les inputs et outputs sont liés par un réseau de neurones qui est calculé afin de prédire au mieux les outputs de 2015. De nombreuses paramétrisations des réseaux de neurones sont testées en fonction des inputs. Dans les réseaux de neurones, les inputs et outputs sont aussi reliés par une formule simple comme :

$$Output \sim Input \quad (4)$$

Et donc dans ce cas-ci la formule f :

$$\begin{aligned} & Cult_{201_2015} + \dots + Cult_{Autre_2015} \\ & \sim \\ & Cult_{201_2014} + \dots + Cult_{Autre_2014} + Aire_{2014} + Péri_{2014} \end{aligned} \quad (5)$$

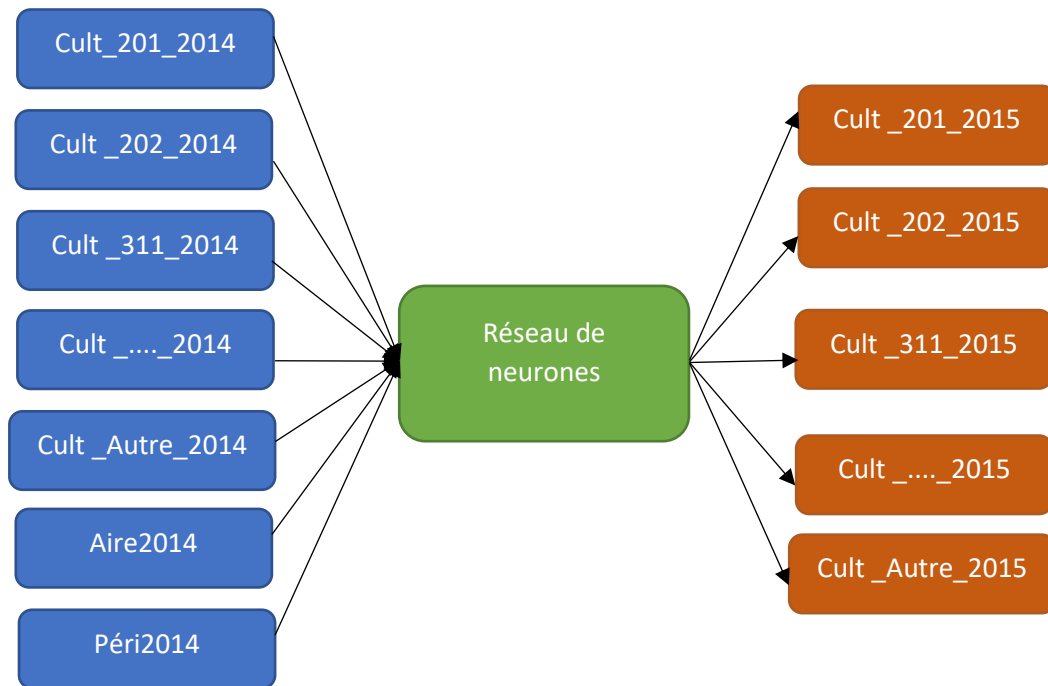


Fig. 16. Schéma de la phase de calibration réelle.

Afin de lancer le calcul des réseaux de neurones, la fonction « neuralnet » est utilisée. Celle-ci utilise dans l'ordre des arguments :

```
neuralnet(f, data = cal, hidden = c(5,5), threshold = 0.02, stepmax = 1e+06, linear.output = FALSE, lifesign = "full")
```

Où, f est la formule ci-dessus (Eq. 5), $data = cal$ est le tableau de données avec les inputs et outputs de calibration, $hidden = c(5,5)$ est le nombre de neurones (5) et de couches (2) cachés qui forment le réseau de neurones, $threshold = 0.02$ est le seuil min de dérivée partielle de l'erreur, $stepmax = 1e+06$ est le nombre d'itérations max pour calibrer le réseau de neurones, $linear.output = FALSE$ est le choix d'une fonction d'activation ou non, $lifesign = "full"$ est le choix de la quantité d'information montrée quant à l'avancement des calculs.

Les arguments qui ont été modifiés et testés sont : le nombre d'inputs à utiliser, le nombre de couches et de neurones et le seuil qui doit être adapté selon les inputs et neurones. Une fois les meilleurs réseaux de neurones formés, il suffit d'utiliser la fonction « compute » afin de prédire les outputs de 2015 à partir des inputs. Pour chaque output (code culture), ces prédictions sont des valeurs continues entre [0:1]. L'output maximal est sélectionné comme étant la culture prédite pour la parcelle. L'entièreté des prédictions pour 2015 sont ensuite comparées à la réalité de 2015 afin de connaître l'exactitude de la phase de calibration. Cette exactitude n'est qu'un aperçu de l'efficacité du modèle mais sera ensuite confirmée par la phase de validation.

2.5.3 Phase de validation

Dans cette phase, les données d'input ne sont plus les points d'intersection mais les polygones qui délimitent les parcelles (voir Fig. 14). Ces polygones peuvent correspondre à une année précédente (2015) ou à plusieurs années précédentes intersectées (2015,2014, 2013, etc.). Le schéma à la Fig.17 représente le cas qui utilise une seule année précédente comme input. Le réseau de neurone optimal trouvé lors de la phase de calibration est utilisé avec la fonction « compute » afin de prédire le code culture de 2016 pour chacun des polygones de 2015.

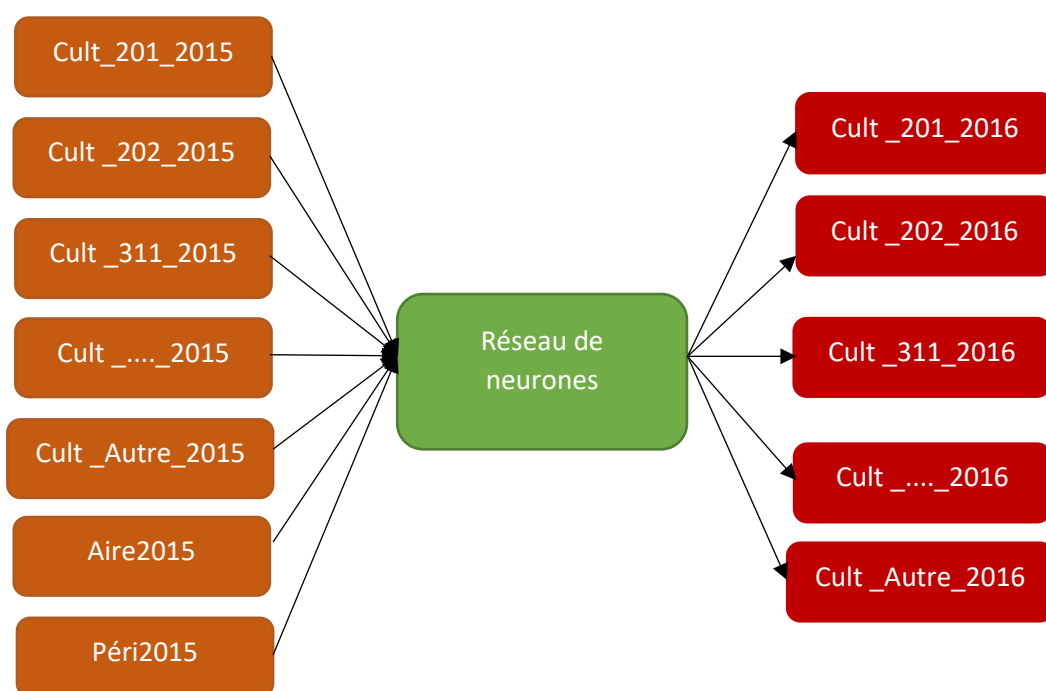


Fig. 17. Schéma de la phase de validation réelle.

Suite à la prédiction, l'exportation des polygones de 2015 possédant l'information de 2016 est réalisée grâce à la fonction « writeShapePoly ». Le shapefile créé est ensuite intersecté avec le PAA de 2016 (Fig.18). Cette intersection forme un plus grand nombre de polygones que le nombre de parcelles présentes. Mais ces polygones possèdent l'information de la prédiction de culture pour 2016 et la réalité de 2016. Ces deux informations permettent de comparer la culture prédite et réelle afin d'identifier les polygones bien prédits. Chaque polygone bien prédit ne représente cependant pas une parcelle. Pour chacun d'entre eux, la surface qu'ils représentent est calculée. Cette surface correspond à la surface bien prédite. C'est cette surface totale bien prédite qu'il faut tenter de maximiser.

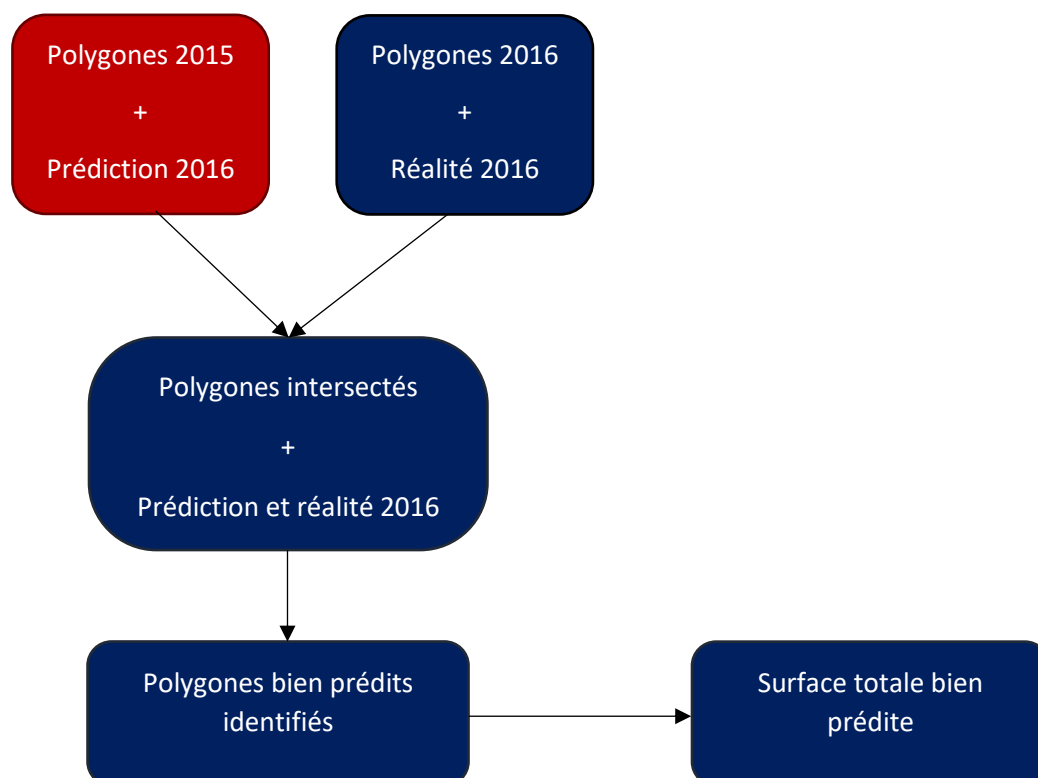


Fig. 18. Phase de validation : vérification de l'erreur

3. Résultats et discussions

3.1 Analyse et traitement des données

3.1.1 Superposition de parcelles parmi les PAA

Dans les couches annuelles des PAA il existe des erreurs de superpositions entre parcelles d'une même couche. Ces superpositions peuvent être de 3 types. Soit certains morceaux de parcelles dépassant les unes sur les autres (1), soit des parcelles dupliquées (2), soit de nouvelles parcelles entières qui superposent une partie d'autres parcelles (3). Seules les superpositions qui concernent les deux derniers cas cités ont été supprimées. En effet, le premier cas est difficile à résoudre et ne représente pas une grande erreur de surface superposée. Nous avons évalué les surfaces de parcelles qui se superposent ainsi que le nombre de superpositions pour les 3 cas au tableau 11.

Tableau 11. Superposition de parcelles parmi les PAA de 2008 à 2016 sur notre zone d'étude.

	2008	2009	2010	2011	2012	2013	2014	2015	2016
Surface (km²)									
Totale	123.53	123.27	124.28	122.60	123.21	123.34	123.26	125.92	124.47
Superposée	0.023	0.003	0.107	0.001	0.007	0.001	0.041	0.004	0.001
Corrigée			0.0075		0.0002		0.0001	0.0006	
Nombre de superpositions									
Avant	1317	1023	406	256	268	256	211	375	384
Après			390		264		206	373	

Les surfaces des superpositions peuvent être considérées comme peu importantes (max 0.1% de la surface totale). Cependant il ne faut pas négliger le nombre de superpositions. On peut compter en moyenne 400 superpositions par année de PAA. Ce chiffre doit être comparé au nombre de parcelles moyen qui est de 4600 soit 8% du nombre de parcelles touchées.

Seules certaines superpositions de 2010, 2012, 2014, 2015 ont été supprimées car elles concernent les cas (2) et (3). Pour les autres années, les seules erreurs relèvent du cas (1). Suite à cela, leurs surfaces superposées ont diminué. La plus grosse surface restante est celle de 2008 que nous n'avons pas pu modifier mais qui ne représente malgré tout que 0.02% de la surface totale. Le nombre de superpositions n'a que peu diminué cependant. Ce nombre restant et important correspond uniquement à des limites de parcelles qui se superposent, de quelques millimètres dans certains cas.

3.1.2 Évolution de la surface des PAA et du nombre de parcelles pour plusieurs zones géographiques

L'analyse des PAA de 2008 à 2016 au niveau de la Wallonie, de la Hesbaye et de la zone d'étude donne le tableau 12. L'absence de chiffres pour 2014 est due à l'indisponibilité des données pour l'ensemble de la Wallonie et la Hesbaye.

Il est à remarquer que la surface du PAA de 2008 sur notre zone d'étude est de 123.53 km². Cette surface était de 132.22 km² sur la COSW de 2007 disponible au point 2.2. Cette différence de 6.5% de surface agricole est expliquée par la prise en compte de toutes les surfaces agricoles sur la COSW. En effet, contrairement au COSW, le PAA ne prend pas en compte les terres agricoles inexploitées ou exploitées à des fins non agricoles (Grandjean, 2016). La surface agricole du PAA est appelée Surface Agricole Utile (SAU) et c'est à partir de celle-ci qu'il est le plus pertinent de faire des prédictions.

Tableau 12. Résumé des PAA de 2008 à 2016 : Wallonie, Hesbaye, Zone d'étude.

	Wallonie		Hesbaye		Zone d'étude	
	Nb parc. ¹	Surf.tot. ² (km ²)	Nb parc.	Surf.tot.(km ²)	Nb parc.	Surf.tot.(km ²)
Année						
2008	275 459	7 621	41 155	1209.9	4868	123.53
2009	275 028	7 616	40 541	1206.9	4739	123.27
2010	282 694	7 705	41 413	1218.2	4735	124.28
2011	275 673	7 642	39 420	1204.5	4487	122.60
2012	277 404	7 662	39 266	1205.1	4499	123.21
2013	278 353	7 670	39 203	1204.5	4527	123.34
2014	/	/	/	/	<u>4188</u>	123.26
2015	338 706	8 157	45 082	1241.5	4836	125.92
2016	338 341	8 119	44 178	1220.1	4822	124.47
Moyenne	292 707	7 774	41 282	1213.8	4633	123.76
Ecart-type (%)	9.7	2.9	5.4	1.1	4.8	0.80

1. Nombre de parcelles 2. Surface totale

Gras : chiffres élevés, Souligné : chiffre faible

De manière générale, d'une année à l'autre, le nombre de parcelles et la surface des cultures varie peu. Les écarts-types sont aussi 4 fois plus importants pour les nombres de parcelles que pour les surfaces. Ce phénomène peut être expliqué par les tournières enherbées et autres petites divisions de parcelles qui apparaissent et disparaissent selon les années. Les écarts-types diminuent lorsque la

surface géographique est de plus en plus petite. La zone d'étude équivaut à 10% de la surface de la Hesbaye et possède quasiment les mêmes écarts-types. Cela permet de penser que la zone d'étude représente bien la Hesbaye.

Ensuite, les années qui posent question sont 2010, 2014 et 2015. En 2010, il y a variation un peu plus importante du nombre de parcelles et de la surface. En 2014, il y a un nombre de parcelles beaucoup moins important sur la zone d'étude. En 2015 et 2016, il y a en général une augmentation importante du nombre de parcelles (+15-20%) mais l'augmentation de surface n'est que de 5% en Wallonie et peu importante ailleurs. Cela montre que le nombre de parcelles a augmenté grandement mais pas la surface. Cela peut être expliqué par l'ajout d'énormément de petites parcelles.

L'évaluation de la couche 2010 montre quelques parcelles en catégorie « Code 0 » qui ne possèdent aucune légende précise. L'évaluation de la couche 2015 montre que les nouvelles parcelles sont en majorité des catégories « Prairies et fourrage » et « Code 99 » (aussi sans légende). L'entretien téléphonique avec le responsable de l'OPW nous a expliqué deux choses :

- 1) La plupart des « Code 99 » correspondent à des parcelles non déclarées (friche) et donc non cultivées. C'était probablement pareil en 2010 pour les « Code 0 ». Ces parcelles sont apparemment ajoutées au PAA afin d'éviter la spéculation.
- 2) L'augmentation de « Prairies et fourrage » est due à la volonté des fermiers de déclarer des terres pour les valoriser. Parmi ces prairies et fourrages on retrouve par exemple les tournières enherbées qui possèdent des petites surfaces, ce qui explique une augmentation plus importante des parcelles que de la surface.

3.1.3 Évolution des types de cultures

Le but ici est de vérifier que les PAA de 2008 à 2016 possèdent une situation stationnaire qui permet de prédire les années futures. En effet, une zone qui possède une représentation de cultures différentes en fonction des années ne permet pas d'observer de rotations dans les cultures. Afin de ménager les graphes, seule l'évolution de 7 cultures les plus importantes est représentée à la Fig.19. Cette figure montre bien l'important pourcentage de surface que représente le froment d'hiver. Cette culture ne varie pas énormément d'une année à l'autre et suit une lente diminution depuis 2012. La Fig.20 s'intéresse plus précisément aux 6 autres cultures qui représentent des surfaces mutuellement plus proches.

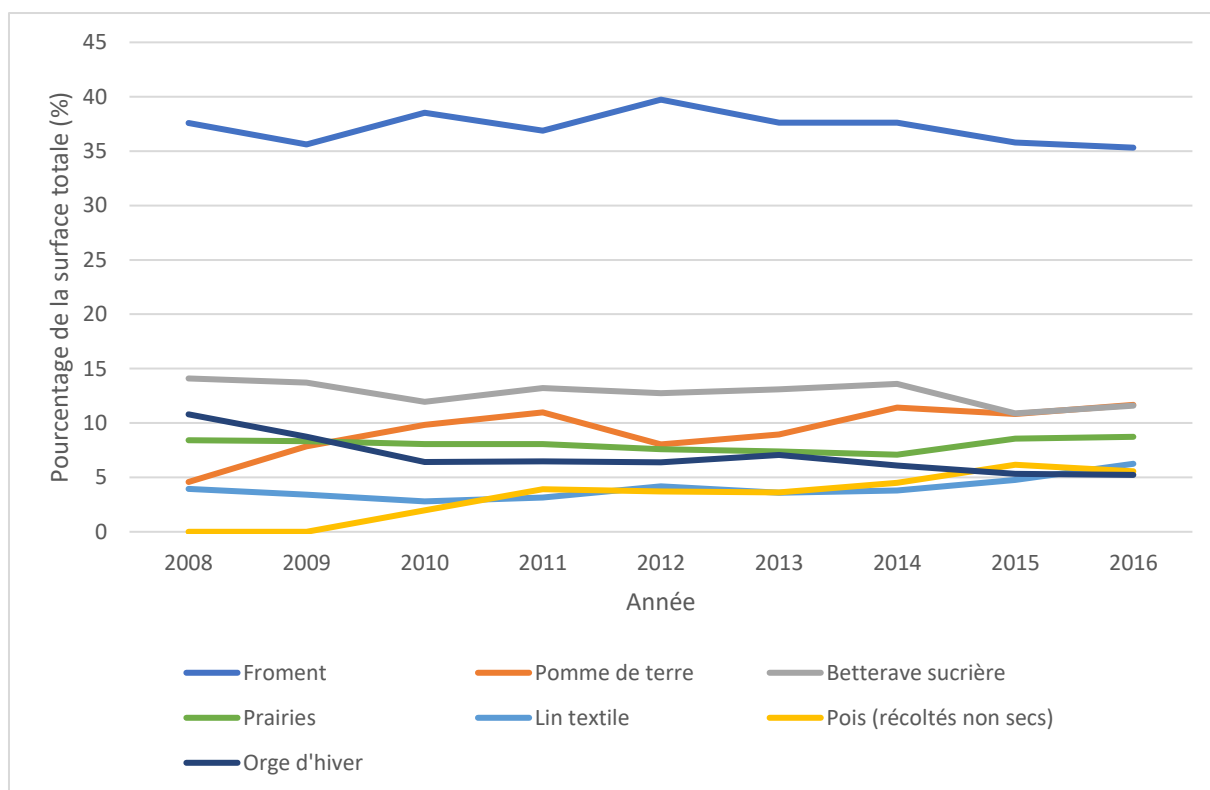


Fig. 19. Evolution des cultures principales de 2008 à 2016 (avec froment)

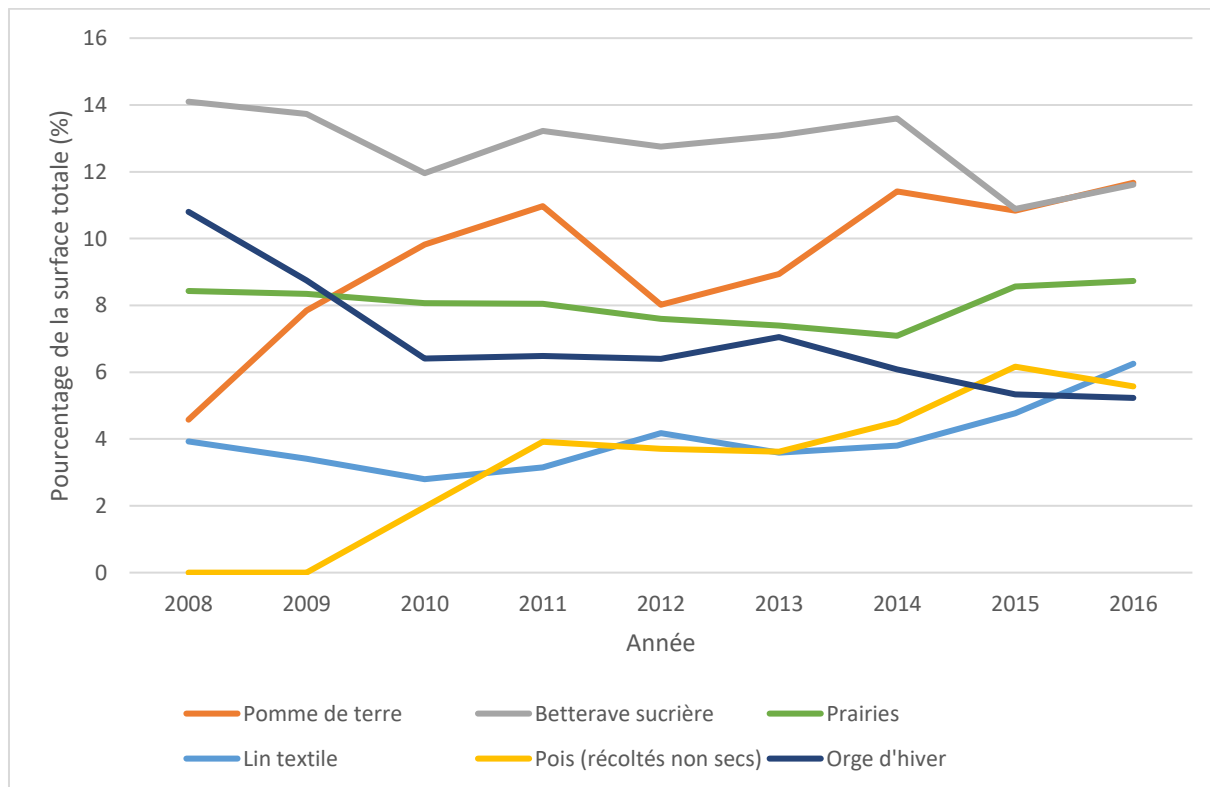


Fig. 20. Evolution des cultures principales de 2008 à 2016 (sauf froment)

Globalement, l'orge d'hiver est en grand diminution. Ajouté à la diminution du froment, les cultures d'hiver diminuent globalement. La betterave est en légère diminution, la pomme de terre augmente bien et le lin textile augmente légèrement. Les pois sont apparus en 2010 directement en grande quantité et ont fortement augmenté jusqu'à aujourd'hui. Ceux-ci ne semblaient pas être cachés dans un autre code culture auparavant car le code des pois existait déjà en 2008 et 2009. Les prairies restent plus ou moins stables dans le temps. Suite à ces observations, on peut considérer que les PAA des années précédentes peuvent être utilisés afin de prédire les prochaines années. Il pourrait être plus intéressant de s'arrêter en 2011 car les changements sont plus importants avant cette année.

3.1.4 Harmonisation et sélection des cultures

En plus des 7 cultures présentées dans ces graphes, il ne faut pas oublier toutes les autres cultures présentes. En effet, chaque année des cultures moins notables changent en superficie de façon importante. Ces changements en superficie peuvent résulter du hasard. Mais ils peuvent aussi résulter de changements dans la classification des cultures dans les PAA selon les années. Ces changements de classification regroupent certaines cultures sous un code commun et d'autres changements divisent les cultures en plusieurs codes.

Par exemple : les codes « prairies permanentes (61) », « temporaires (62) » et « tournières enherbées (751) » ont été regroupés en 2015 en un seul code « prairies et fourrages (6) ». Dans l'autre sens, en 2013, les « cultures maraichères (951) » ont été scindées en « légumes à racines (966) » et « légumes à légumineuse (967) ». Une analyse en profondeur de ces changements a été faite afin d'harmoniser les codes cultures pour toutes les années. Seule la solution d'harmonisation mise en place est présentée ci-dessous au tableau 13.

Tableau 13. Harmonisation des codes cultures

61	Prairie permanente		
62	Prairie temporaire		6 Prairie et fourrage
6	Prairie et fourrage		
751	Tournière enherbée		
951	Cultures maraichères		
966	Légumes à racine		
967	Légumes légumineuses		951 Autres légumes
9410	Haricots de conserverie		
7433	Carotte		
9515	Endives (chicons)		
971	Cultures fruitières pluri-annuelles		971 Cultures fruitières pluri-annuelles
9741	Cultures fruitière pluriannuelles-basses tiges		
85	Autres couverts semés		
80	Jachère		
82	Couvert de graminées		85 Autres couverts semés
83	Autres couvertures (droits jachère)		
84	Couvert graminées + légumineuses		
34	Avoine		34 Avoine
342	Avoine de printemps		
4111	Colza d'hiver		411 Colza et navette d'hiver
411	Colza et navette d'hiver		
52	Fèves et féveroles (récoltées en sec)		52 Fèves et féveroles (récoltées en sec)
522	Fèves et Féveroles de printemps		

Après harmonisation, 22 codes culture sont gardés. Le reste des codes cultures ainsi que les « Code 0 » et « Code 99 » (parcelles non déclarées) sont reclassifiés en « Autre ». Finalement, les années 2008 à 2016 sont représentées par une liste harmonisée de 23 codes cultures.

Dans le tableau 14, nous faisons la distinction entre les cultures qui n'ont pas subi de changement de code et les cultures qui ont subi une harmonisation.

Tableau 14. Cultures gardées pour prédiction

Cultures présentes chaque année		Cultures harmonisées	
Code	Nom de la culture	Code	Nom de la culture
201	Maïs ensilage	34	Avoine
202	Maïs grain	411	Colza et navette d'hiver
311	Froment d'hiver	52	Fèves et féveroles (en sec)
312	Froment de printemps	6	Prairie et fourrage
321	Orge d'hiver	85	Autres couverts semés
322	Orge de printemps	951	Autres légumes
36	Epeautre	971	Cultures fruitières pluriannuelles
743	Autres fourrages	/	Autre
901	Pomme de terre de consommation		
903	Pomme de terre féculière		
91	Betterave sucrière		
921	Lin textile		
931	Pois (non récoltés secs)		
9520	Pépinières de plants fruitiers		
9811	Chicorée à inuline		

3.1.5 Comparaison des types de cultures en fonction des zones géographiques

La surface couverte par ces 23 types de cultures en Wallonie, Hesbaye et dans la Zone d'étude de l'année 2016 est représentée au tableau 15. Pour chaque type de culture est ajoutée la taille moyenne d'une parcelle en Wallonie. La taille moyenne d'une parcelle tout type confondu est de 2.4 ha.

Tableau 15. Comparaison des types de cultures en 2016 en Wallonie, Hesbaye et Zone d'étude.

	Surface des cultures (%)			Taille moy. parc. ¹ (ha)
	Wallonie	Hesbaye	Zone d'étude	
Prairie et fourrage	46.07	13.60	8.73	2.31
Froment d'hiver	17.30	34.39	35.32	3.93
Autre	7.28	4.16	2.43	0.94
Maïs ensilage	6.55	4.18	2.89	2.29
Pomme de terre de consommation	4.68	9.24	11.67	3.92
Betterave sucrière	4.64	11.64	11.61	4.09
Orge d'hiver	4.55	6.07	5.23	3.70
Lin textile	1.44	3.93	6.25	5.03
Colza et navette d'hiver	1.42	1.24	0.34	4.44
Autres légumes	1.11	3.01	3.69	3.75
Epeautre	1.08	0.71	0.38	2.57
Pois (non récoltés secs)	1.02	3.42	5.57	5.49
Chicorée à inuline	0.62	2.00	2.40	4.26
Maïs grain	0.55	0.86	1.64	1.65
Autres fourrages	0.51	0.43	0.42	1.44
Avoine	0.35	0.11	0.18	1.92
Autres couverts semés	0.27	0.42	0.58	0.56
Orge de printemps	0.25	0.11	0.07	2.56
Froment de printemps	0.09	0.11	0.01	2.52
Fèves et féveroles (en sec)	0.06	0.21	0.55	4.02
Pomme de terre féculière	0.05	0.10	/	3.29
Pépinières de plants fruitiers	0.01	/	/	0.24
Cultures fruitières pluriannuelles	0.01	/	/	0.06

1. Taille moyenne des parcelles en Région wallonne

Les différences entre la situation wallonne et la zone d'étude sont nombreuses. Les deux différences les plus importantes sont « Prairies et fourrages » qui passe de 46% à 8.7% et « Autre » qui passe de

7.3% à 2.4%. La première diminution est plutôt un point négatif. En effet, la prédiction est plus ardue avec moins de prairie car une prairie a énormément de chance de rester prairie l'an prochain et d'être facilement prédite. La diminution de parcelles « Autre » est plutôt un point positif. Cela signifie que les 22 autres codes cultures sont adaptés à notre zone d'étude. En effet, ils couvrent 97.6% de la surface agricole. Ensuite, il y a une différence parmi les types de cultures dominantes. La zone d'étude possède des cultures propres à sa zone agro-géographique comme vu précédemment. Par exemple : le froment d'hiver, les pommes de terre, les betteraves sucrières, le lin textile et les pois ont l'air typiques de la région. A l'opposé : le maïs ensilage, le colza d'hiver et l'épeautre ont peu de succès.

Les différences entre la Hesbaye et la zone d'étude sont faibles en comparaison. La plus grande différence est la classe « Prairies et fourrages » qui utilise 5% de surface totale en plus en Hesbaye. À côté de ça, les pommes de terre, le lin textile et les pois utilisent environ 2% de surface totale en moins en Hesbaye. Enfin, les 22 codes cultures qui ne sont pas « Autre » permettent d'identifier 95.8% de la surface agricole. Ces faibles différences permettent de penser que la zone d'étude est assez représentative de la Hesbaye et donc que notre modèle de prédiction des cultures pourrait être applicable à la zone agro-géographique de la Hesbaye.

Les tailles moyennes des parcelles selon les cultures varient quelque peu. Les cultures caractérisées par les plus grandes parcelles sont les pois et le lin textile. Les cultures avec les plus petites parcelles sont les « Autre » et « Autres couverts semés ». Cependant, ce n'est pas renseigné ici mais les tailles de parcelles de la zone d'étude diffèrent un peu de celles en Wallonie. Si cette surface moyenne est utilisée comme facteur de prédiction, il est plus exact d'utiliser celle liée à notre site d'étude.

Enfin, les 3 dernières classes de cultures du tableau 10 posent problème. Ces cultures présentes en grande quantité toutes les années avant 2016 ne sont quasi plus représentées ici en 2016. Le choix de ces cultures avait été fait sans se rendre compte qu'elles n'apparaissaient plus en 2016. Malheureusement, la plupart du travail et des modèles réalisés ont utilisé les 23 cultures car elles étaient présentes de 2008 à 2015. Cette prise de conscience lors de la rédaction du mémoire montre bien l'importance d'une analyse de données encore plus systématique. Il faut donc garder en tête pour la suite des résultats que l'année 2016 ne possède que 20 des 23 codes cultures utilisés.

3.1.6 Assemblage des différentes couches des PAA

Afin de connaître les successions de cultures pour chaque parcelle et utiliser ces tendances pour prédire 2016, il faut pouvoir assembler les PAA de 2008 à 2015. En effet, l'année 2016 doit être prédite à partir des données précédentes. La couche de 2016 ne servira qu'à valider la prédiction. En utilisant la méthode des centroïdes avec l'option « Inside » pour l'année 2015, 4835 centroïdes sont obtenus et correspondent à autant de parcelles.

Ensuite, l'intersection entre les centroïdes 2015 et les polygones des années précédentes permet de conserver les centroïdes de 2015 qui sont présents dans tous les PAA. Les intersections entre les centroïdes 2015 et les polygones des années antérieures sont appelées des points d'intersection. L'évolution de ces points est détaillée dans le tableau 16.

Tableau 16. Evolution du nombre de points d'intersection selon l'assemblage des PAA

Année	Nombre de points d'intersection ¹	Evolution Year => Year-1
2015	4835	/
2014	4299	-536
2013	4249	-50
2012	4214	-35
2011	4150	-64
2010	4122	-28
2009	4085	-37
2008	4068	-18

1. Pour l'année 2015 cela correspond aux centroïdes originaux

Le premier fait remarquable est que dès 2014 le nombre de points d'intersection diminue de 536 parcelles. Cela signifie que 536 parcelles sont apparues de 2014 à 2015 sur des terrains antérieurement non repris dans le PAA. Finalement, après l'intersection de toutes les années, il reste 4068 parcelles. Il y a donc une diminution de quasi 20% du nombre de centroïdes de 2015 à 2008. Il faut cependant minimiser cette diminution. En effet, l'apparition de nouvelles parcelles peut être expliquée par des ajouts de parcelles non déclarées en 2015 (voir 3.1.2). C'est probablement la raison pour laquelle la plus grosse différence se trouve entre l'année 2014 et 2015.

Suite à l'intersection des différentes couches du PAA, certaines années possèdent moins de classes de cultures que les 23 choisies au tableau 14. Ce nombre de classes de cultures pour chaque année est représenté au tableau 17. Cette disparition de classes certaines années est cohérente avec la réalité. Par exemple, le pois qui n'était pas présent en 2008 est la 6ème culture la plus importante en 2015.

Tableau 17. Nombre de classe de cultures par année

	2008	2009	2010	2011	2012	2013	2014	2015	2016
Nombre de classes de cultures	20	22	23	22	23	22	22	23	20

3.1.7 Rotation de cultures

Afin d'analyser les rotations de cultures et leurs occurrences, les 20 rotations les plus fréquentes selon le nombre d'années sont représentées au tableau 18. Pour chaque nombre d'année, le pourcentage de représentativité de chaque rotation est exprimé. Celui-ci correspond au pourcentage du nombre de parcelles qui suivent une rotation donnée. En bas du tableau, la somme des pourcentages des 20 rotations les plus importantes est calculée pour se rendre compte de la part des rotations non présentes dans le tableau. Enfin, le nombre total de rotations différentes est mentionné afin de montrer que 20 rotations ne représentent qu'une infime partie de toutes les rotations possibles.

Tableau 18. Rotations les plus importantes en fonction du nombre d'années

2 années		3 années		4 années		5 années	
Rotation	% ¹	Rotation	%	Rotation	%	Rotation	%
6 6	21.7	6 6 6	20.9	6 6 6 6	20.2	6 6 6 6 6	19.5
91 311	7.8	311 91 311	4.2	311 321 91 311	2.0	311 91 311 901 311	0.9
311 321	5.7	311 901 311	3.3	91 311 901 311	1.5	311 901 311 91 311	0.9
311 91	4.9	321 91 311	2.4	901 311 91 311	1.1	91 311 321 91 311	0.7
901 311	4.6	311 321 91	2.1	311 91 311 901	1.0	311 321 91 311 321	0.6
311 311	4.6	91 311 901	1.6	311 901 311 91	1.0	91 311 901 311 91	0.5
311 901	3.6	91 311 321	1.5	311 311 91 311	0.8	202 202 202 202 202	0.5
321 91	2.6	91 311 311	1.4	321 91 311 321	0.7	201 201 201 201 201	0.4
201 311	2.4	901 311 91	1.2	311 91 311 311	0.7	321 91 311 901 311	0.4
9811 311	1.9	311 9811 311	1.2	743 743 743 743	0.6	743 743 743 743 743	0.4
311 201	1.8	311 201 311	1.2	91 311 321 91	0.6	901 311 91 311 901	0.4
311 951	1.6	311 921 311	1.1	202 202 202 202	0.6	6 6 6 6 85	0.4
311 921	1.6	311 311 321	1.1	311 91 311 321	0.6	311 321 91 311 311	0.4
202 311	1.6	311 951 311	1.1	201 201 201 201	0.5	311 311 321 91 311	0.4
951 311	1.6	311 311 91	0.9	321 91 311 311	0.5	91 311 311 91 311	0.4
921 311	1.5	743 743 743	0.9	311 321 202 311	0.5	321 91 311 321 91	0.3
311 931	1.4	311 311 311	0.8	91 311 9811 311	0.5	311 921 311 91 311	0.3
311 9811	1.3	202 202 202	0.8	311 901 311 321	0.4	311 321 91 311 901	0.3
201 201	1.3	311 202 311	0.8	311 321 901 311	0.4	311 951 311 901 311	0.3
743 743	1.2	311 931 311	0.7	311 321 201 311	0.4	311 9811 311 91 311	0.3
% Total ²	74.8		49.3		34.8		28.3
Nb rotations ³	358		1538		3313		4920

1. Pourcentage de représentativité

2. Somme des pourcentages des 20 rotations les plus importantes

3. Nombre de rotations totales

De manière générale, le nombre de rotations différentes augmente fortement au plus il y a d'années dans la rotation. Ensuite, les pourcentages totaux du top 20 diminuent avec chaque année. Il y a aussi diminution des pourcentages individuels des rotations fréquentes. Le fait d'avoir plus de rotations qui représentent toutes un pourcentage de moins en moins important peut être un point négatif car une

prédiction des cultures doit être faite à partir de toutes les rotations. Enfin, la rotation la plus importante pour toutes les années est la « monoculture » de prairies (6). Dans le reste des rotations, le froment d'hiver (311) est énormément présent car il représente la plus grande surface de culture dans la zone d'étude (35%, tableau 10).

Parmi les rotations à 2 années, il est à remarquer que certains principes agronomiques de base sont respectés. Par exemple, les cultures de printemps ne sont pas suivies d'une autre culture de printemps. C'est le cas pour la betterave (91), les pommes de terre (901), le maïs grain (202), la chicorée à inuline (9811), le lin textile (921) et les autres légumes (951) qui sont tous suivis par du froment d'hiver. Une petite exception est le maïs ensilage (201) qui est suivi de froment ou en monoculture.

Les rotations à 3 années sont aussi faciles à analyser. Rappelons que la rotation de culture typique de la Hesbaye est une rotation triennale : Betterave – Froment d'hiver – Orge d'hiver (suivi d'un engrais vert). Tout d'abord, en dessous de la monoculture de 6 (prairie permanente), apparaît la rotation 311-91-311 (froment – betterave – froment). En considérant que le premier froment d'hiver peut substituer une culture d'orge d'hiver, cette rotation respecte la règle agronomique. La suite de la liste montre la rotation 311-901-311 qui substitue la betterave et la remplace par la pomme de terre (901). Cette substitution de la betterave peut être faite aussi par plusieurs autres cultures de printemps (chicorée, maïs ensilage/grain, lin, pois). Viennent ensuite deux rotations qui représentent exactement la rotation Betterave – Froment – Orge mais dans des ordres de début et de fin différents. Finalement, il reste des rotations qui ne suivent pas la rotation triennale classique qui préconise deux cultures d'hiver parmi les 3 années. Il y a par exemple la rotation 901-311-91 (Pomme de terre -Froment- Betterave) représentée deux fois dans des sens différents et une monoculture de maïs grain (202-202-202). La rotation Pomme de terre-Froment-Betterave est déconseillée car elle peut occasionner des repousses de pomme de terre et du mildiou dans la culture de betterave (SillonBelge.be, 2017). Avec l'agriculture moderne, ces règles agronomiques ne semblent donc pas toujours respectées.

Enfin, les rotations à 4 et 5 années montrent que le nombre de rotations se diversifie de plus en plus. Le haut de la liste montre des rotations type : Printemps – Hiver – Hiver. Cela signifie qu'il y a une plus grande diversité dans les types de cultures de printemps et donc que les pourcentages sont divisés en plusieurs rotations. Cela est aussi dû au faible nombre de cultures d'hiver possibles. Cependant, de plus en plus de monocultures apparaissent dans le top 20 : 201, 202, 743 (autre fourrages). De plus dans les rotations à 5 années, une rotation apparaît avec petit changement par rapport à la monoculture 6. Un code 6 est remplacé par un code 85 (autres couverts). Suite à l'analyse de l'entièreté des rotations à 5 années non montrées ici, il est clair que les code 85 et 743 peuvent être associés aux prairies (6).

3.2 Prédiction par matrices de transition

3.2.1 Calibration

Comme vu lors de la méthodologie au point 2.4.3, la première partie de la calibration est de créer les 25 matrices de transition. Un exemple de celles-ci pour une année antérieure est représenté à la page suivante, tableau x.

Tout d’abord, en vert est représentée la culture non présente en 2011. A aucun moment, une culture de 2010 ne devient 322 en 2011. Ensuite, pour chaque culture de 2010, la plus grande probabilité qu’elle deviennent un type de culture en 2011 est signalée en rouge. C’est cette culture qui est reprise ensuite dans le tableau de prédiction. En analysant ces chiffres rouges, il est évident qu’ils ne représentent qu’un nombre restreint de classes de culture de 2011. Seules 10 classes sont prédites en 2011 en les utilisant. La plupart des classes de 2010 sont prédites en 2011 comme 311. Finalement, dans certains cas, la seconde plus grande probabilité est proche du chiffre rouge. Celle-ci est représenté en bleu. Pour certains cas de 2010 ces valeurs rouges et bleues sont identiques : 322, 36, 9520. Le choix fait entre des valeurs égales pour la table de prédiction est malheureusement fait au hasard par la fonction « max ». Dans le cas de 9520 par exemple, la prédiction pourrait être 9520 ce qui serait plus logique.

Tableau 19. Exemple de matrice de transition avec 23 cultures : 2010_2011

2011 2010	201	202	311	312	321	322	34	36	411	52	6	743	85	901	903	91	921	931	951	9520	971	9811	Autre	Total
201	29.4	2.5	49.2	0.0	0.0	0.0	0.5	3.0	0.0	0.0	0.0	0.0	0.0	7.1	0.0	5.1	0.0	0.5	0.5	0.0	0.0	1.5	0.5	100
202	2.1	31.5	57.3	1.4	0.0	0.0	0.7	0.0	0.0	0.0	0.7	0.0	0.0	2.8	0.0	2.8	0.7	0.0	0.0	0.0	0.0	0.0	0.0	100
311	5.5	2.9	15.4	0.0	18.2	0.0	2.2	1.2	0.5	0.0	1.2	0.4	0.2	13.2	0.2	18.8	5.0	5.1	4.8	0.0	0.0	4.3	0.7	100
312	0.0	0.0	18.8	0.0	6.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	12.5	0.0	25.0	0.0	0.0	31.3	0.0	0.0	6.3	0.0	100
321	8.9	4.6	3.9	0.0	10.3	0.0	3.9	0.7	8.9	0.0	0.7	0.7	0.0	7.5	0.4	35.2	5.0	1.8	1.8	0.0	0.4	3.6	1.8	100
322	0.0	22.2	22.2	0.0	11.1	0.0	0.0	0.0	5.6	5.6	0.0	0.0	0.0	11.1	0.0	11.1	0.0	0.0	5.6	0.0	0.0	5.6	0.0	100
34	3.7	0.0	59.3	0.0	3.7	0.0	7.4	3.7	0.0	0.0	3.7	0.0	0.0	7.4	0.0	7.4	0.0	0.0	3.7	0.0	0.0	0.0	0.0	100
36	13.2	13.2	10.5	0.0	5.3	0.0	0.0	0.0	0.0	0.0	0.0	2.6	0.0	15.8	0.0	7.9	2.6	0.0	13.2	0.0	0.0	0.0	15.8	100
411	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100
52	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100
6	1.1	0.7	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	95.0	0.6	0.3	0.1	0.0	0.4	0.1	0.0	0.0	0.0	0.1	0.0	0.9	100
743	3.2	1.1	4.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.5	74.7	0.0	0.0	0.0	6.3	0.0	0.0	0.0	0.0	0.0	0.0	1.1	100
85	0.0	8.3	25.0	0.0	2.8	0.0	0.0	0.0	0.0	0.0	13.9	16.7	30.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.8	100
901	1.3	0.4	91.2	0.4	0.0	0.0	0.0	0.9	0.0	0.0	0.9	0.0	0.0	0.4	0.0	1.8	0.0	0.4	1.8	0.4	0.0	0.0	0.0	100
903	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100
91	2.6	0.5	80.5	0.0	0.0	0.0	0.3	0.8	0.0	0.0	0.8	0.8	0.0	5.7	0.0	0.5	1.3	0.8	1.8	0.0	0.0	1.3	2.3	100
921	1.4	0.0	75.4	0.0	1.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.7	1.4	8.7	0.0	1.4	1.4	0.0	0.0	0.0	0.0	100
931	2.1	0.0	54.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.1	2.1	0.0	6.3	0.0	27.1	0.0	2.1	0.0	0.0	0.0	4.2	0.0	100
951	0.0	1.9	50.5	0.9	0.9	0.0	0.0	1.9	0.0	0.9	0.9	0.0	0.0	15.9	0.0	9.3	0.0	0.9	12.1	0.0	0.0	0.9	2.8	100
9520	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	33.3	0.0	13.3	0.0	20.0	0.0	33.3	0.0	0.0	0.0	100
971	0.0	14.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.1	0.0	0.0	0.0	0.0	0.0	0.0	78.6	0.0	0.0	100
9811	3.6	0.0	88.1	1.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.6	0.0	0.0	0.0	1.2	0.0	0.0	0.0	0.0	2.4	100
Autre	2.6	1.3	35.5	0.0	1.3	0.0	1.3	0.0	0.0	0.0	5.3	2.6	0.0	3.9	0.0	3.9	1.3	1.3	2.6	0.0	1.3	1.3	34.2	100

Rouge : plus grande probabilité. Bleu : seconde plus grande probabilité. Vert : culture non présente.

Pour rappel, la calibration tente de prédire la dernière année de chacune des 25 matrices de transition. Lors de la prédiction, le but est d'obtenir un Nombre de **Bonnes** Prédictions maximum (NBP). Ce nombre de **bonnes** prédictions peut être divisé par le nombre de parcelles totales et fournir ce que l'on appellera le Pourcentage de **Bonne** Prédiction (PBP).

$$PBP = \frac{NBP}{\text{Nombre de parcelles totales}} \quad (6)$$

Pour chaque nombre d'années antérieures utilisées, la moyenne des PBP augmente (Fig.21). Ici ne sont mentionné que la moyenne des matrices seules. Les écarts types montrent à quel point les prédictions peuvent varier en fonction de la matrice de transition utilisée et sur quelle année elle est utilisée.

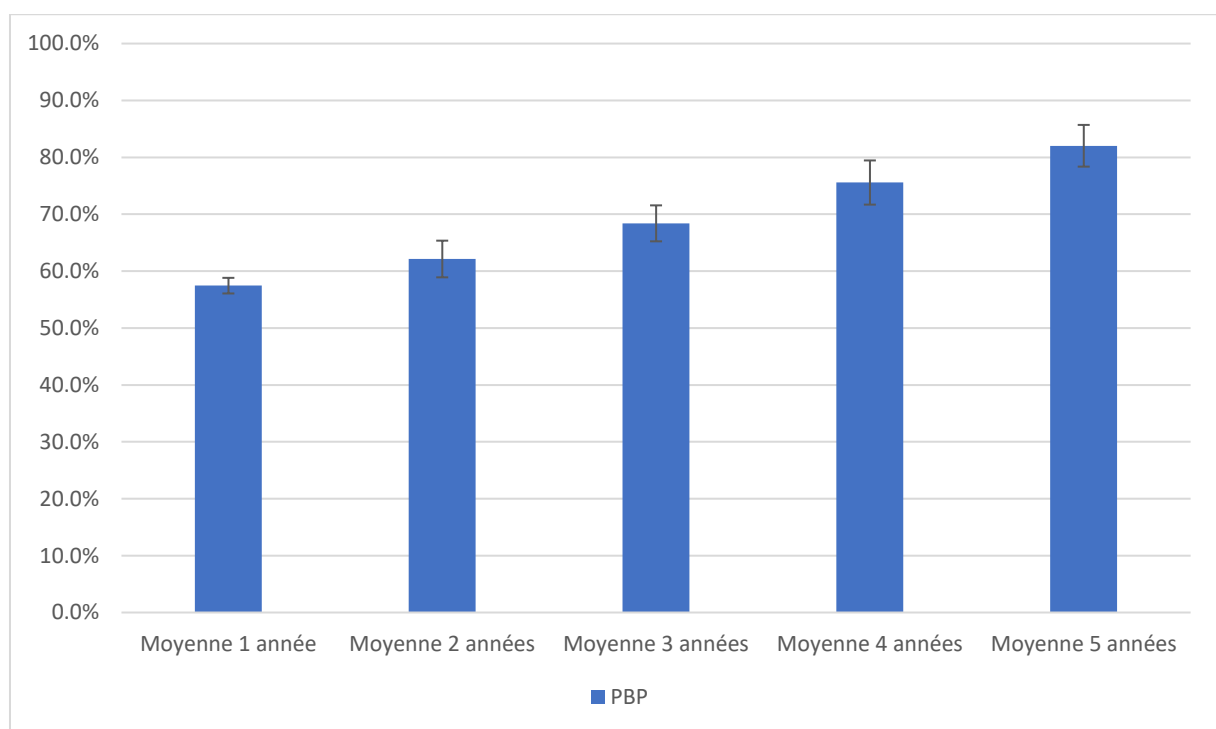


Fig. 21. Calibration : évolution du PBP en fonction du nombre d'années antérieures

De manière générale, il est normal que le PBP augmente autant à chaque ajout du nombre d'années. En effet, l'ajout d'année permet d'affiner la prédiction car ajoute des combinaisons de cultures. De plus, dans le cas de la calibration, les matrices de transitions représentent exactement la prédiction qui est faite. La calibration ne donne donc pas une idée correcte d'une prédiction faite pour une année future.

3.2.2 Validation

Pour rappel, dans notre cas la validation tente de prédire l'année 2016 à partir de 2015 et les années précédentes. Pour ce faire, la validation utilise les 25 matrices de transition vues dans la calibration. L'utilisation de ces matrices de transition peut se faire soit sur les points d'intersections pour connaître le PBP soit sur les polygones pour connaître la surface bien prédite.

Fusion des matrices de transition

L'utilisation de plusieurs années antérieures permet de prédire plus de classes de cultures et d'éviter la confusion lors du choix de la plus grande probabilité. Cependant dans le cas de la validation, un nombre plus important d'années signifie un nombre de combinaisons théoriques plus important (voir tableau x.). Or le nombre moyen de combinaisons d'un échantillon d'une seule matrice de transition ne permet pas de prédire toutes ces combinaisons théoriques de l'année suivante. Il faut pour cela un nombre plus important de parcelles dans la zone d'étude ou fusionner les matrices de transition par nombre d'années utilisées. C'est ce qui a été fait, toutes les matrices de transition à une année ont été fusionnées ensembles. Cela a aussi été réalisé pour 2 à 5 années (tableau 20).

Tableau 20. Evolution du nombre de combinaisons en fonction du nombre d'années antérieures utilisées

	Nombre d'années antérieures utilisées				
	1	2	3	4	5
Nb. Comb. théoriques ¹ (23 ⁿ)	23	529	12167	279841	6436343
Moy. des comb. obs. avec matrices seules ² (%)	22 (95%)	188 (35%)	560 (5%)	1080 (0.4%)	1640 (0.02%)
Comb. obs. avec matrices fusionnées ³ (%)	23	316 (60%)	1209 (10%)	2289 (0.8%)	2774 (0.04%)

1. Nombre de combinaisons théoriques

2. Moyenne des combinaisons observées avec matrices seules

3. Combinaisons observées avec matrices fusionnées

A première vue le nombre de combinaisons augmente significativement grâce aux matrices fusionnées et double la plupart du temps. Cependant, même avec les matrices fusionnées le nombre de combinaisons observées est loin d'approcher le nombre de combinaisons théoriques. Cela peut s'expliquer par le fait qu'il y a énormément de rotations de cultures qui ne se font pas pour des questions phytosanitaires (succession de certaines cultures de printemps par exemple). Il est dès lors plus intéressant de s'intéresser au nombre de prédictions. Le nombre de prédiction est le nombre de parcelles auquel a été assigné une culture lors de la validation. Ce n'est pas le nombre de **bonnes** prédictions (NBP), vu précédemment, pour lesquelles une prédiction a pu être réalisée.

La moyenne du nombre de prédictions en fonction du nombre d'années antérieures est représentée au tableau 21. Le nombre de prédictions diminue au plus le nombre d'années utilisées est important mais cette diminution est bien moins importante par rapport à celle des combinaisons de culture. De manière générale, les matrices fusionnées permettent de gagner en nombre de prédictions.

Tableau 21. Evolution du nombre de prédictions en fonction du nombre d'années antérieures utilisées

	Nombre d'années antérieures utilisées				
	1	2	3	4	5
Nb. de parcelles totales	4068	4068	4068	4068	4068
Moy. du Nb. de préd. avec matrices seules ¹ (%)	4058 (99.7%)	3943 (97%)	3633 (89%)	3123 (77%)	2474 (60%)
Nb. de prédictions avec matrices fusionnées (%)	4068 (100%)	4040 (99%)	3874 (95%)	3484 (85%)	2725 (67%)

1. Moyenne du nombre de prédictions avec matrices seules

Comme vu à l'instant, le nombre de prédictions n'est pas toujours égal au nombre de parcelles totales dans le jeu de données. Ainsi, un nouveau terme d'évaluation que l'on appellera l'Exactitude de prédiction est déterminé.

$$Exactitude = \frac{NBP}{Nombre\ de\ prédictions} \quad (7)$$

Appliquée aux points d'intersection

Les résultats de validation appliqués aux points d'intersection permettent de connaître les PBP et les exactitudes pour chaque approche. La première approche a été de comparer la prédiction de 2016 avec des matrices de transition d'années de plus en plus lointaines. La Fig. 22 compare les matrices de transition à 1 année appliquées à 2015 pour prédire 2016. De manière générale, il n'y a pas de tendance qui en ressort. Une matrice provenant de 2008_2009 appliquée à 2015 prédit même mieux 2016 que la matrice 2014_2015. Cette observation a été faite aussi pour les combinaisons de 2 à 5 années.

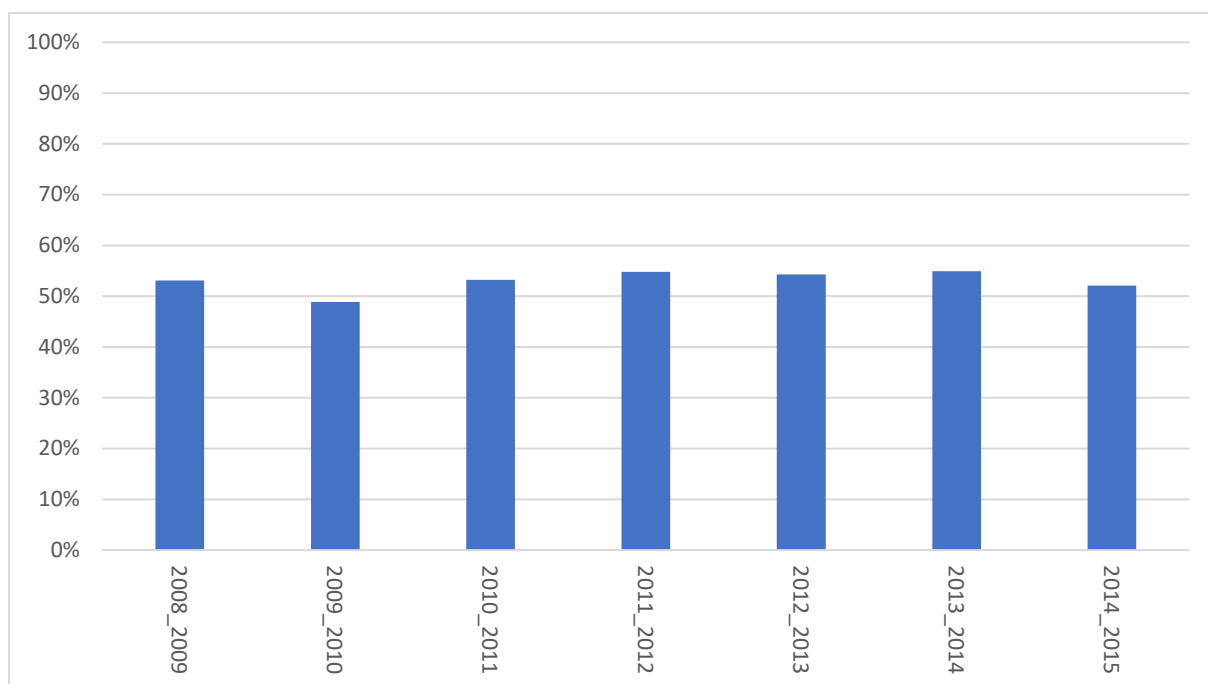


Fig. 22. PBP en utilisant des matrices seules, plus ou moins récentes, pour prédire 2016.

Ensuite, le PBP moyen de toutes les prédictions par année a été calculé. La même chose a été faite pour l'exactitude. Le PBP et l'exactitude ont été comparés à la Fig. 23.

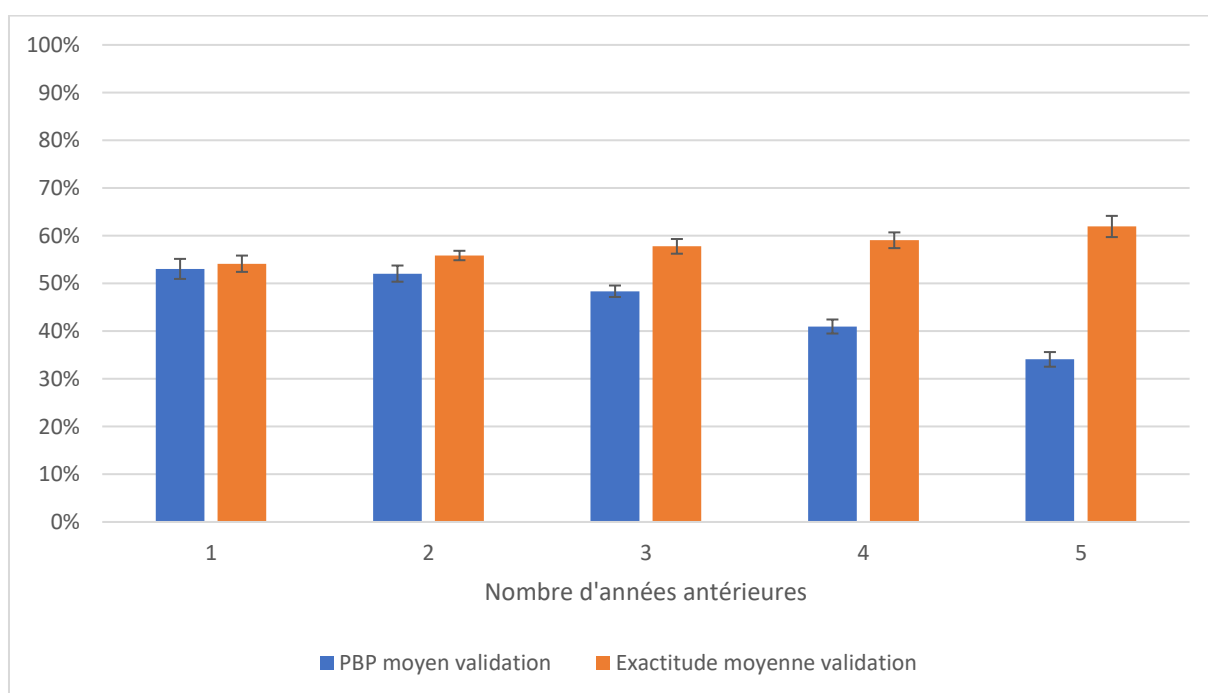


Fig. 23. Evolution du PBP et de l'Exactitude en fonction du nombre d'années antérieures

Selon le nombre d'années, il y a des tendances nettes. Tout d'abord, le PBP démarre à 1 année avec un max de 53%. Celui-ci diminue ensuite de plus en plus à mesure que le nombre d'années augmente. C'est un résultat qui était prévisible dû au nombre de prédictions chutant avec chaque année ajoutée.

Ensuite, l'Exactitude démarre au même niveau que le PBP mais contrairement à lui, celle-ci augmente avec l'ajout d'années. C'est expliqué par l'augmentation du nombre de combinaisons de cultures dans la prédiction au détriment du nombre de prédictions. La différence entre la valeur de PBP et d'Exactitude montre la marge qui pourrait être gagnée si le nombre de parcelles du jeu de données était plus important ou s'il y a utilisation des matrices fusionnées.

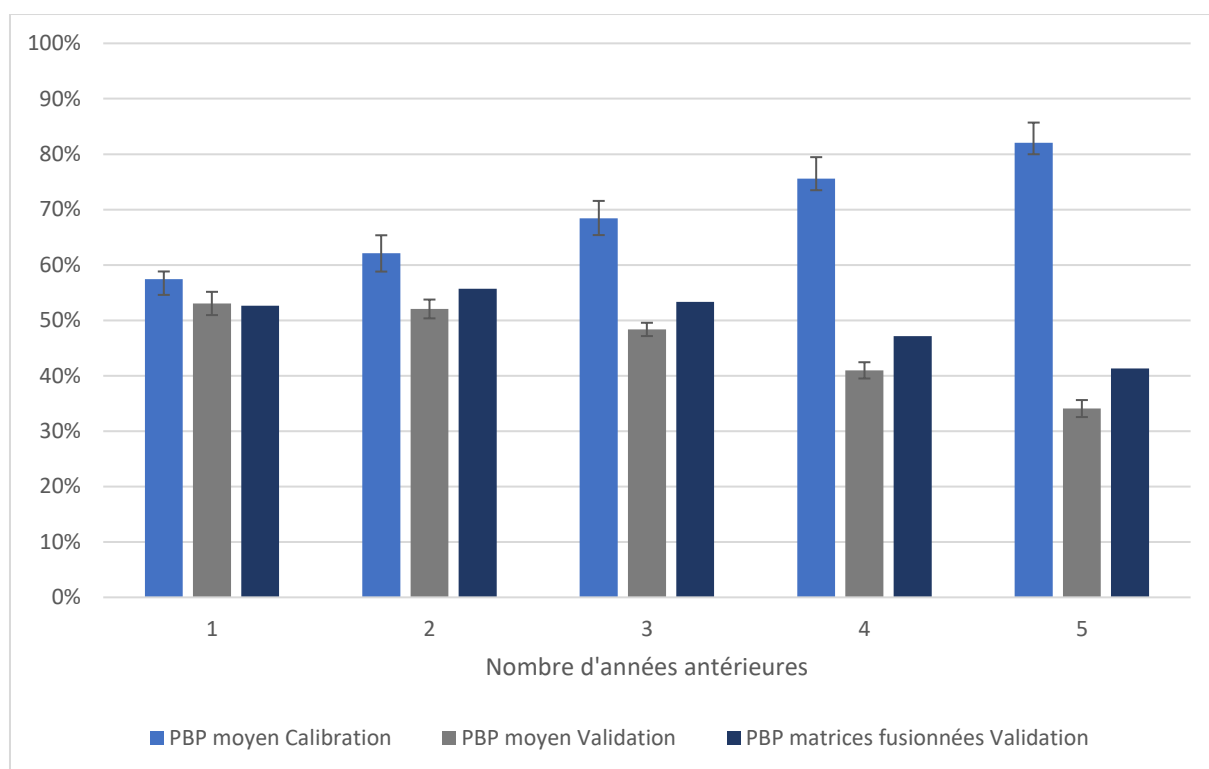


Fig. 24. Evolution du PBP : comparaison Calibration, Validation et matrices fusionnées

Les résultats du PBP entre la calibration et la validation sont fortement différents (Fig. 24). Cela démontre bien que la calibration représente mal la prédiction faite pour une année inconnue.

Ensuite, l'utilisation des matrices fusionnées permet de gagner la plupart du temps en PBP par rapport à la moyenne des matrices seules (Fig.24). Ce n'est pas le cas pour une année, mais l'apport est de plus en plus important avec l'ajout d'années supplémentaires. Le meilleur PBP est observé pour 2 années avec 55.7% suivi par 3 années avec 53.3%.

Les exactitudes des matrices seules et fusionnée sont finalement comparées à la Fig. 25.

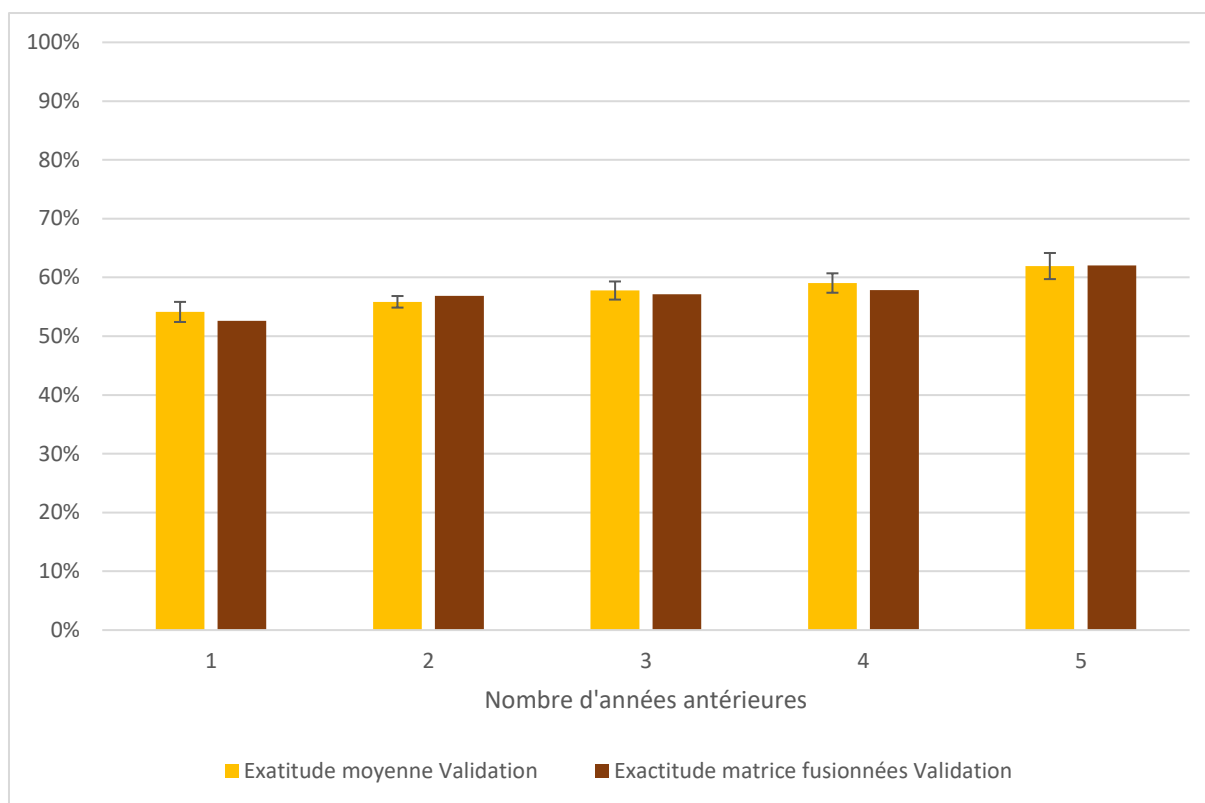


Fig. 25. Evolution de l'Exactitude : comparaison matrices seules et fusionnées

L'apport des matrices fusionnées à l'exactitude est de manière générale légèrement négative. En effet, les matrices fusionnées permettent de prédire plus de parcelles totales mais ces prédictions sont de moins bonne qualité. Parmi les nouvelles parcelles prédites grâce aux matrices fusionnées, l'exactitude était moins importante et donc l'exactitude globale a chuté par rapport aux matrices seules. Il est donc plus intéressant de garder une seule matrice si le but est d'être certain d'avoir une bonne exactitude. Ceci aurait tendance à appuyer le fait que les rotations ne sont pas stables dans le temps.

Matrices de confusion et autres méthodes

Il est difficile d'analyser plus précisément le PBP car certaines parcelles ne sont pas prédites dû au manque de combinaisons de cultures. Il est possible cependant d'analyser plus en détails les résultats d'exactitude. Pour ce faire, 2 et 5 années ont été comparées : l'une étant la plus intéressante au niveau du PBP et l'autre possédant la meilleure exactitude. C'est l'utilisation de la matrice de confusion qui permet d'analyser l'exactitude des prédictions. Les matrices de confusion de 2 et 5 années sont présentées aux tableaux 22 et 23 et possèdent 57% et 62% d'exactitude.

Pour chaque code, la sensibilité est calculée à partir du nombre de bonne prédiction (NBP, en **gras**) et du nombre de fois ou la culture est actuellement observée (voir tableau 22).

$$Sensibilité = \frac{NBP \text{ de la culture}}{Observations \text{ totales de la culture}} \quad (8)$$

La précision est quant à elle le NBP divisée par le nombre de prédictions de la culture.

$$Précision = \frac{NBP \text{ de la culture}}{Prédictions \text{ totales de la culture}} \quad (9)$$

Enfin, le pourcentage du NBP est le NBP de la culture divisé par le NBP total.

$$\% \text{ du NBP} = \frac{NBP \text{ de la culture}}{NBP \text{ total}} \quad (10)$$

Tableau 22. Matrice de confusion sur l'Exactitude de la prédiction par matrices fusionnées de 2 années.

Prédiction Observation	311	6	91	901	321	201	921	202	85	Autre	931	951	9811	743	36	411	34	52	322	312	971	Observations Totales	Sensibilité	% du NBP
311	84%	1%	4%	3%	2%	3%	0%	2%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1114 (28.0%)	84%	41.3%
6	1%	94%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	3%	0%	0%	0%	0%	0%	0%	0%	869 (21.8%)	94%	35.9%
91	24%	2%	55%	3%	11%	4%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	329 (8.3%)	55%	7.9%
901	24%	0%	25%	31%	15%	3%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	278 (7.0%)	31%	3.8%
321	5%	2%	33%	20%	36%	3%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	211 (5.3%)	36%	3.4%
201	26%	6%	10%	10%	7%	31%	0%	4%	1%	1%	1%	0%	0%	3%	1%	0%	0%	0%	0%	0%	0%	194 (4.9%)	26%	2.3%
921	14%	0%	50%	21%	11%	2%	0%	1%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	159 (4.0%)	0%	0.0%
202	14%	3%	21%	4%	23%	3%	0%	25%	1%	1%	0%	0%	0%	5%	1%	0%	0%	0%	0%	0%	0%	148 (3.7%)	25%	1.6%
85	9%	70%	0%	0%	1%	2%	0%	0%	9%	4%	0%	0%	0%	3%	0%	0%	0%	0%	1%	0%	0%	135 (3.4%)	9%	0.5%
Autre	26%	11%	7%	2%	11%	8%	0%	1%	0%	29%	0%	3%	0%	2%	0%	0%	0%	0%	2%	0%	1%	133 (3.3%)	29%	1.7%
931	20%	1%	50%	13%	10%	2%	0%	0%	0%	0%	0%	2%	0%	0%	2%	0%	0%	0%	0%	0%	0%	127 (3.2%)	0%	0.0%
951	30%	2%	26%	24%	9%	1%	0%	0%	0%	3%	0%	3%	0%	1%	0%	0%	0%	0%	0%	0%	0%	90 (2.3%)	3%	0.1%
9811	14%	0%	41%	30%	5%	2%	0%	0%	0%	0%	0%	7%	0%	0%	0%	0%	0%	0%	0%	0%	0%	56 (1.4%)	0%	0.0%
743	15%	15%	6%	4%	0%	4%	0%	0%	0%	4%	0%	0%	0%	52%	0%	0%	0%	0%	0%	0%	0%	48 (1.2%)	52%	1.1%
36	36%	5%	5%	14%	16%	14%	0%	0%	0%	2%	5%	5%	0%	0%	0%	0%	0%	0%	0%	0%	0%	44 (1.1%)	0%	0.0%
411	0%	0%	67%	0%	22%	0%	0%	0%	0%	6%	0%	6%	0%	0%	0%	0%	0%	0%	0%	0%	0%	18 (0.5%)	0%	0.0%
34	14%	7%	7%	7%	50%	0%	7%	7%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	14 (0.4%)	0%	0.0%
52	60%	0%	20%	10%	0%	0%	0%	10%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10 (0.3%)	0%	0.0%
322	43%	29%	0%	0%	29%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	7 (0.2%)	0%	0.0%
312	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1 (0.03%)	0%	0.0%
971	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	NA	NA
Prédictions Totales	1331	981	631	299	301	154	4	78	20	63	4	23	0	85	5	1	0	0	3	0	2	3985 (100%)		100%
Précision (%)	70	83	29	28	25	40	0	47	60	60	0	13	NA	29	0	0	NA	NA	0	NA	0		2266 (57%)	

Tableau 23. Matrice de confusion sur l'Exactitude de la prédiction par matrices fusionnées de 5 années

Prédiction Observation	311	6	91	901	321	201	921	202	85	Autre	931	951	9811	743	36	411	34	52	322	312	971	Observations Totales	Sensibilité	% du NBP
311	80%	1%	3%	3%	2%	3%	1%	2%	1%	1%	1%	0%	0%	0%	2%	0%	0%	0%	0%	0%	0%	649 (24.0%)	80%	31.10%
6	0%	96%	0%	0%	0%	0%	0%	0%	1%	1%	0%	0%	0%	2%	0%	0%	0%	0%	0%	0%	0%	767 (28.3%)	96%	44.00%
91	15%	4%	42%	9%	7%	2%	5%	4%	0%	1%	5%	2%	3%	1%	0%	0%	0%	0%	0%	0%	0%	238 (8.8%)	42%	6.00%
901	21%	1%	18%	37%	13%	2%	1%	1%	1%	0%	1%	2%	2%	0%	0%	2%	1%	0%	0%	0%	0%	175 (6.5%)	37%	3.80%
321	22%	1%	10%	9%	35%	3%	5%	7%	0%	2%	3%	1%	0%	0%	0%	3%	0%	0%	0%	0%	0%	153 (5.7%)	35%	3.20%
201	22%	5%	13%	2%	13%	32%	3%	2%	1%	1%	4%	0%	1%	0%	0%	1%	1%	0%	0%	0%	0%	114 (4.2%)	32%	2.20%
921	13%	0%	30%	20%	12%	2%	14%	2%	0%	0%	4%	2%	0%	0%	0%	0%	1%	0%	0%	0%	0%	125 (4.6%)	14%	1.10%
202	14%	5%	9%	2%	12%	5%	0%	42%	0%	3%	1%	0%	0%	4%	0%	0%	4%	0%	0%	0%	0%	104 (3.8%)	42%	2.60%
85	3%	25%	0%	0%	0%	6%	0%	2%	59%	4%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	97 (3.6%)	59%	3.40%
Autre	18%	12%	7%	4%	4%	5%	2%	4%	4%	33%	2%	5%	0%	2%	0%	0%	0%	0%	0%	0%	0%	57 (2.1%)	33%	1.10%
931	11%	0%	32%	16%	16%	1%	4%	3%	0%	1%	15%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	75 (2.8%)	15%	0.70%
951	23%	0%	11%	23%	6%	0%	2%	2%	0%	0%	15%	11%	2%	0%	0%	0%	2%	0%	0%	0%	2%	47 (1.7%)	11%	0.30%
9811	20%	0%	30%	28%	5%	3%	3%	3%	0%	0%	10%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	40 (1.5%)	0%	0.00%
743	17%	22%	11%	0%	0%	11%	0%	0%	6%	0%	0%	6%	0%	28%	0%	0%	0%	0%	0%	0%	0%	18 (0.7%)	28%	0.30%
36	29%	10%	19%	0%	24%	0%	0%	0%	0%	0%	0%	5%	0%	0%	14%	0%	0%	0%	0%	0%	0%	21 (0.8%)	14%	0.20%
411	10%	0%	40%	0%	0%	10%	10%	10%	0%	0%	10%	0%	0%	0%	0%	10%	0%	0%	0%	0%	0%	10 (0.4%)	10%	0.10%
34	33%	17%	17%	0%	0%	0%	0%	33%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	6 (0.2%)	0%	0.00%
52	29%	0%	14%	0%	0%	0%	14%	0%	0%	14%	0%	29%	0%	0%	0%	0%	0%	0%	0%	0%	0%	7 (0.3%)	0%	0.00%
322	25%	50%	0%	0%	25%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	4 (0.1%)	0%	0.00%
312	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	NA	NA
971	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	NA	NA
Prédictions Totales	738	805	284	182	175	89	55	94	72	43	61	26	17	30	15	10	10	0	0	0	1	2707 (100%)		100%
Précision (%)	71	92	36	35	31	42	33	47	79	44	18	19	0	17	20	10	0	NA	NA	NA	0		1680 (62%)	

Tout d'abord, 2 cultures manquent (9520 et 903) tout simplement car non prédites et non observées en 2016. Ensuite, certaines cultures ne sont pas prédites mais sont observées. Pour 2 années c'est le cas pour les codes 34, 9811, 52, 312. Et pour 5 années, c'est le cas pour les codes 322, 52, 312. C'est dû au manque de combinaisons de cultures et dû à la sélection du plus grand pourcentage de transition dans les matrices de transition. Par exemple, le code 311 (froment d'hiver) est prédit 1331 fois au total alors qu'il n'est actuellement observé que 1114 fois. Enfin, certaines cultures sont prédites mais pas observées en 2016. C'est le cas du code 971 par exemple.

La culture la mieux prédite est la 6 (Prairies et fourrages) avec 94% (2 ans) et 96% (5 ans) de sensibilité. Elle est talonnée par la culture 311 (Froment d'hiver) avec une sensibilité de 84% et 80%, respectivement. Ces deux cultures représentent des grands '% du NBP'. Pour 2 années cela correspond à 77% du NBP, et pour 5 années à 75% du NBP.

La culture 91 (Betterave) est la 3^e la plus présente. Elle n'est cependant pas très bien prédite avec seulement 55% et 42% de sensibilité. Elle est souvent prédite comme 311 ou 321.

La culture 901 (Pomme de terre non hâtives) est la 4^e la plus présente. Elle est assez mal prédite avec 31% et 37% de sensibilité. Elle est souvent prédite comme 91 ou 311.

La culture 321 (Orge d'hiver) est la 5^e la plus présente. Elle est assez mal prédite avec 36% et 35% de sensibilité. Elle est souvent prédite comme 91 ou 901 pour 2 années et comme 311 pour 5 années.

La culture 201 (Mais ensilage) est la 6^e la plus présente. Elle est assez mal prédite avec 31% et 32% de sensibilité. Elle est souvent prédite comme 311.

La culture 921 (Lin textile) est la 7^e la plus présente. Elle n'est jamais bien prédite avec 2 années et très peu prédite pour 5 années avec 14% de sensibilité. Elle est souvent prédite comme 91 ou 901.

La culture 202 (Mais grain) est la 8^e la plus présente. Elle est très peu bien prédite avec 25% et 42% de sensibilité. Elle est souvent prédite comme 321 ou 91 pour 2 années.

La culture 85 (Autres couverts) est la 9^e la plus présente. Elle n'est quasiment jamais bien prédite pour 2 années avec 9%. Mais bien prédite pour 5 années avec 59% de sensibilité. Les parcelles mal prédites sont considérées comme 6 (Prairie).

Enfin, la culture Autre est la 10^e la plus présente. Elle est très peu bien prédite avec 29% et 33% de sensibilité. Elle est souvent prédite comme 311.

D'autres observations sur les restes de cultures ne sont pas à négliger. La culture 36 (Epeautre) n'est jamais bien prédite au profit de 311 et 321. La culture 951 (Autres légumes) n'est quasi jamais bien prédite mais ne suit pas de logique quant à sa prédiction en tant qu'autre code. La culture 931 (Pois

récoltés non secs) n'est jamais bien prédite au profit de 91. Et la culture 9811 (Chicorée à inuline) n'est jamais bien prédite au profit de 91 et 901.

L'autre valeur calculée à partir de la matrice de confusion est la précision (aussi appelé valeur prédictive positive ou exactitude de l'utilisateur). Elle est présentée en bas des tableaux 22 et 23. Celle-ci est utile pour calculer le score F1 pour chacune des cultures. Le score F1 est la moyenne pondérée de la précision et de la sensibilité.

$$Score F_1 = 2 \frac{(Sensibilité \times Précision)}{Sensibilité + Précision} \quad (11)$$

Tableau 24. Score F1 pour chaque culture pour 2 (haut) et 5 années (bas).

311	6	91	901	321	201	921	202	85	Autre	931	951	9811	743	36	411	34	52	322	312	971
76%	88%	38%	29%	30%	35%	NA	33%	15%	39%	NA	5%	NA	38%	NA	NA	NA	NA	NA	NA	NA
75%	94%	39%	36%	33%	37%	20%	45%	67%	38%	16%	14%	NA	21%	17%	10%	NA	NA	NA	NA	NA

Les valeurs NA apparaissant lorsque la sensibilité ou la précision égalent zéro. Le score F1 de chaque culture augmente globalement en passant de 2 à 5 années comme prévu. Cependant, dans notre cas, les exactitudes entre 2 et 5 années sont déjà fort différentes. Ce score F1 aurait été plus utile afin de comparer des résultats avec des exactitudes plus proches. Par exemple, il n'est pas évident de choisir entre des résultats avec une grande précision d'un côté et une grande sensibilité de l'autre. Le score F1 aurait été utile dans ce cas-là.

Enfin une dernière valeur qui permet de comparer les deux méthodes est le coefficient Kappa. Kappa utilise l'exactitude et ajuste cette valeur avec l'Exactitude aléatoire. L'exactitude aléatoire est l'exactitude qui pourrait être espérée par la chance seule. Elle est la somme des produits des effectifs marginaux divisée par le carré de la taille de l'échantillon.

$$Kappa = \frac{Exactitude - Exactitude Aléatoire}{1 - Exactitude Aléatoire} \quad (12)$$

Le coefficient Kappa augmente comme prévu de 2 à 5 années fusionnées. Celui-ci passe de 0.48 à 0.54. Ces valeurs de coefficient Kappa sont considérées comme des accords modérés car se retrouvant entre 0.41 et 0.60.

Appliquée aux polygones intersectés

L'utilisation de polygones intersectés au lieu des points d'intersection permet de connaître les surfaces bien prédites en 2016. Cette surface prédite est comparée aux 124.47 km² total de SAU sur le PAA de 2016 afin de connaître le pourcentage de surface bien prédite. Elle a été calculée pour les meilleurs PBP, c'est-à-dire lorsque 1, 2 et 3 années antérieures étaient utilisées (voir Fig.24). Le pourcentage de surface bien prédite a été comparé aux valeurs de PBP (voir Fig. 26) pour voir si un lien existe entre les deux valeurs.

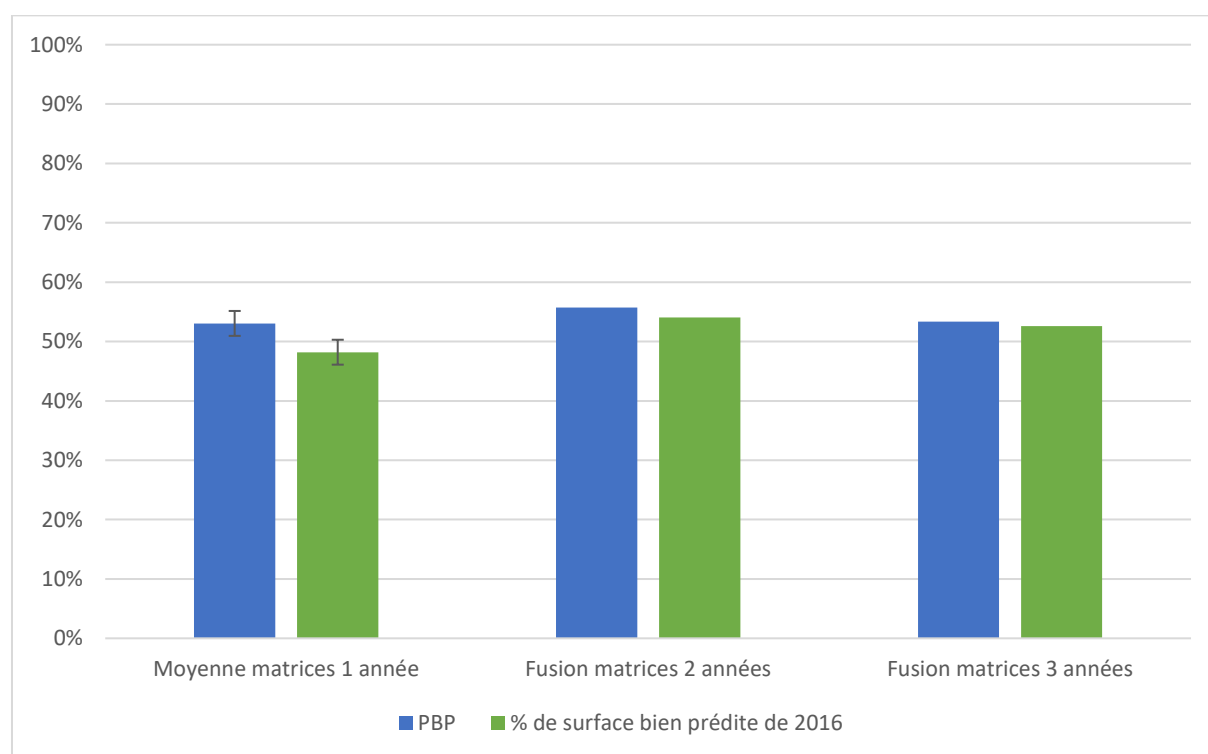


Fig. 26. Comparaison PBP et % de surface bien prédite de 2016

Les deux valeurs n'ont a priori aucune raison d'être comparées car la taille des parcelles varie et le PBP n'est pas forcément représentatif de la surface bien prédite. Cependant, il y a tout de même une tendance qui montre que le pourcentage de surface bien prédite est généralement inférieur au PBP. Cette différence semble moins importante lors de l'utilisation de plusieurs années antérieures. Un pourcentage de surface bien prédite moins important que le PBP signifie que ce sont plutôt les parcelles de petites tailles qui sont bien prédites.

Enfin, c'est donc l'utilisation de matrices fusionnées à deux années antérieures qui permet d'avoir le meilleur PBP ainsi que le plus grand % de surface bien prédite avec 54.1%.

3.2.3 Application aux groupes de cultures

La prédiction des cultures ne s'est pas avérée assez précise par notre méthode qui prend en compte 23 cultures précises. Le regroupement de ces cultures a donc été fait sous 4 termes : culture d'Hiver, de Printemps, Prairies et Autre. La transformation pour chaque culture est présente au tableau 25.

Ces regroupements peuvent être réalisés de deux manières qui seront testées. Soit un regroupement après prédiction des 23 types de cultures par les matrices de transition. Soit un regroupement avant prédiction et avant la création des matrices de transition. Les résultats n'ont été réalisés que sur la partie Validation et sur 1, 2 et 3 années antérieures.

Tableau 25. Regroupement des cultures sous les groupes : Hiver, Printemps, Prairie, Autre

Code culture	Groupe
91 (Betterave)	Printemps
951 (Autres légumes)	Printemps
901 (Pomme de terre de consommation)	Printemps
921 (Lin textile)	Printemps
9811 (Chicorée à inuline)	Printemps
201 (Mais ensilage)	Printemps
202 (Mais grain)	Printemps
322 (Orge de printemps)	Printemps
34 (Avoine)	Printemps
312 (Froment de printemps)	Printemps
931 (Pois)	Printemps
903 (Pomme de terre féculière)	Printemps
52 (Fève et fèverole)	Printemps
6 (Prairie et fourrage)	Prairie
743 (Autre fourrages)	Prairie
85 (Autre couvert)	Prairie
311 (Froment)	Hiver
321 (Orge d'hiver)	Hiver
36 (Epeautre)	Hiver
411 (Colza et navette d'hiver)	Hiver
9520 (Pépinière)	Autre
971 (Culture fruitière pluriannuelle)	Autre
Autre	Autre

Regroupement après prédiction

Les résultats en utilisant ces 4 groupes de cultures sont déjà plus intéressants que lors de l'utilisation de 23 codes cultures (Fig.27). En utilisant 2 et 3 années, le PBP augmente de 17% après le regroupement. Celui-ci n'augmente que de 12% avec 1 année. Les meilleurs résultats restent ceux à deux années avec 72.5% de PBP. Les différences entre le PBP et l'exactitude restent les mêmes.

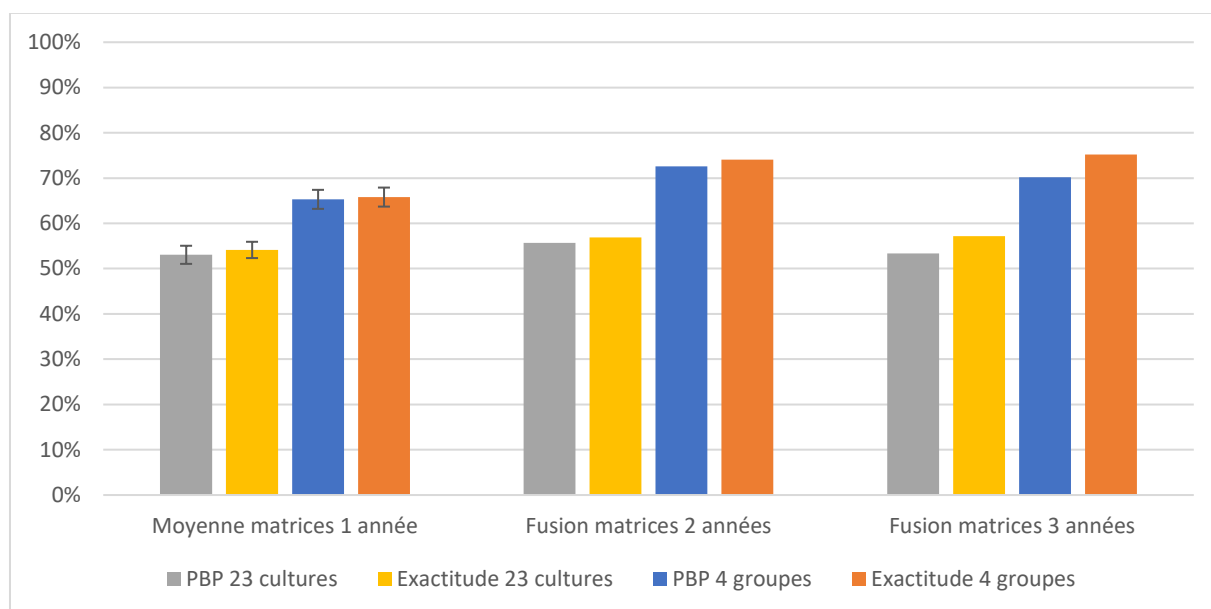


Fig. 27. Comparaison résultats 23 cultures et 4 groupes de cultures après prédiction

Regroupement avant prédiction

Dans ce cas-ci, il faut réaliser de nouvelles prédictions à partir des 4 nouveaux groupes créés. Cela signifie donc de nouvelles matrices de transition. Les matrices de transition de 1 et 2 années sont présentées ci-dessous mais elles ont été réalisées aussi pour 3 à 5 ans. Le pourcentage du nombre de parcelles totales de chaque culture dans le « présent » est aussi renseigné.

Tableau 26. Groupes de cultures : Matrice de transition 1 an

Futur Présent					% du nombre de parcelles
	Printemps	Hiver	Prairie	Autre	
Printemps	25%	73%	1%	1%	29%
Hiver	59%	39%	1%	1%	42%
Prairie	6%	3%	91%	0%	28%
Autre	15%	10%	17%	58%	1%

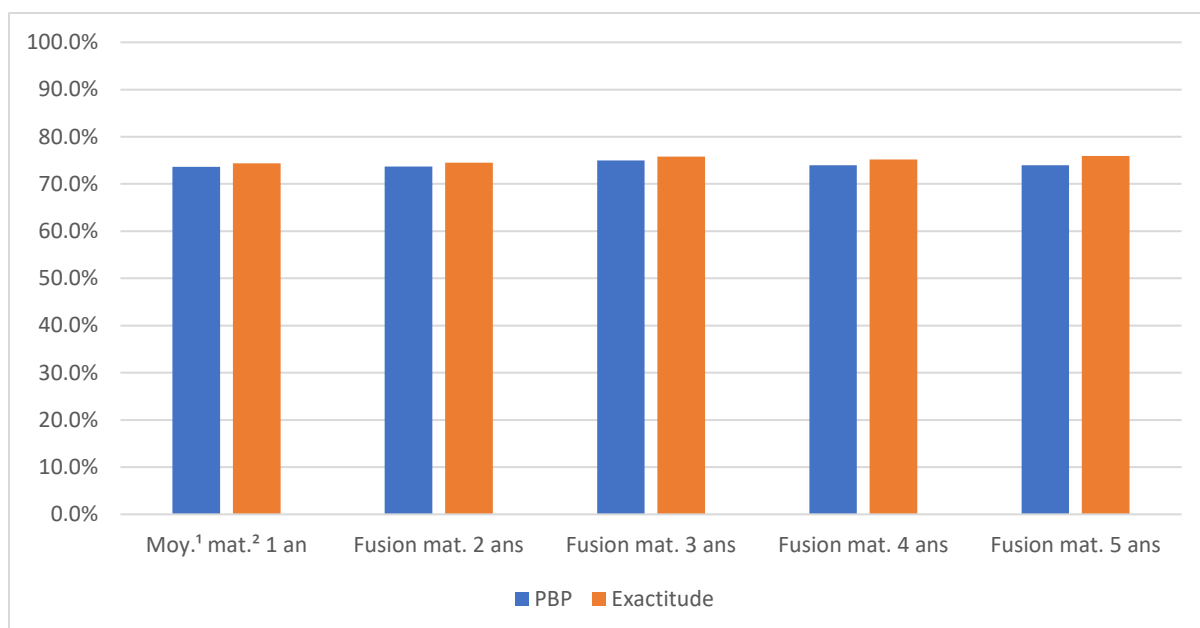
Dans cette matrice de 1 an, il faut tout d'abord remarquer qu'une culture de printemps sera suivie 73% du temps par une culture d'hiver. C'est donc cette succession qui sera retenue pour la prédiction.

Cependant, 25% du temps la culture de printemps est suivie d'une culture de printemps (typiquement les monocultures de maïs). Dans ces cas-là malheureusement la culture prédite sera hiver ce qui donnera une certaine erreur. C'est le même principe pour les cultures d'hiver qui sont suivies à 59% de cultures de printemps. Ensuite, les prairies sont à 91% suivies par d'autres prairies, l'erreur de prédiction devrait donc être faible ici. Enfin, les cultures « Autre » sont suivies de « Autre » mais sont assez diverses dans leurs successions.

Tableau 27. Groupes de culture : Matrice de transition 2 ans fusionnée

Futur Présent					
	Printemps	Hiver	Prairie	Autre	% du nombre de parcelles
Prairie Prairie	3%	1%	95%	1%	25.0%
Hiver Printemps	17%	81%	1%	1%	23.9%
Printemps Hiver	64%	32%	2%	2%	23.6%
Hiver Hiver	60%	35%	3%	2%	13.7%
Printemps Printemps	44%	50%	4%	3%	7.8%
Prairie Printemps	46%	33%	19%	3%	1.0%
Autre Autre	20%	3%	4%	72%	1.0%
Hiver Prairie	11%	19%	69%	1%	0.7%
Printemps Prairie	15%	3%	74%	8%	0.6%
Hiver Autre	30%	48%	1%	20%	0.6%
Autre Printemps	35%	50%	6%	8%	0.5%
Prairie Hiver	46%	36%	15%	3%	0.5%
Printemps Autre	40%	24%	2%	34%	0.4%
Autre Hiver	48%	39%	3%	9%	0.4%
Prairie Autre	8%	10%	26%	56%	0.2%
Autre Prairie	7%	0%	80%	13%	0.1%

Les 5 premiers types de rotations représentent 94% du nombre de parcelles. Il y a tout d'abord la monoculture de prairie et la rotation hiver-printemps-hiver qui possèdent des pourcentages de succession qui permettent de ne pas faire trop d'erreur de prédiction. Ensuite, la rotation printemps-hiver et hiver-hiver peuvent être suivies par hiver 1 fois sur 3. Il y aura donc des erreurs lors de la prédiction. Enfin, la rotation printemps-printemps est assez bien représentée et sa succession est fortement partagée entre hiver et printemps. La prédiction d'une culture hiver succédant à cette rotation sera très peu exacte. Ces successions de cultures de printemps sont expliquées par les monocultures de maïs qui dans certains cas reviennent à une culture d'hiver.



1. Moyenne. 2. matrices

Fig. 28. Regroupement avant prédiction : résultats.

Les résultats du regroupement avant prédiction sont légèrement plus intéressants que ceux après prédiction (Fig.28). Ceux-ci trouvent des PBP entre 73.5% et 75% alors que le PBP max dans la méthode après prédiction est de 72.5%. De manière générale, les PBP sont relativement semblables et l'utilisation de 3 années semble être le plus intéressant. L'exactitude de 3 années (75.8%) est analysée dans la matrice de confusion ci-dessous (tableau 28).

Tableau 28. Regroupement avant prédiction : Matrice de confusion de l'exactitude de prédiction par matrices de 3 ans fusionnée

	Printemps	Hiver	Prairie	Autre	Observations totales	Sensibilité	% du NBP
Printemps	70%	25%	4%	1%	1428 (35.5%)	70.4%	33.0%
Hiver	27%	70%	3%	0%	1395 (34.7%)	70.4%	32.2%
Prairie	2%	1%	96%	1%	1061 (26.4%)	95.6%	33.3%
Autre	35%	16%	14%	34%	140 (3.5%)	34.3%	1.6%
Prédictions Totales	1450	1376	1121	77	4024 (100%)		100%
Précision	69.3%	71.4%	90.5%	62.3%		75.8%	

Les cultures de printemps et hiver représentent à peu près le même nombre d'observations totales ainsi que la même sensibilité. Leur sensibilité est inférieure à l'exactitude moyenne. Celle-ci est plus importante grâce à la sensibilité des prairies qui sont prédites à 95%. Les 3 groupes printemps, hiver et prairie représentent alors le même '% du NBP' alors que le nombre d'observations de prairies est

moins important. En s'intéressant aux cultures mal prédites (pas dans un carré rouge), il est clair que les cultures de printemps mal prédites sont surtout prédites en culture d'hiver et inversement. Ensuite les cultures « Autre » sont très peu représentées et généralement mal prédites. Enfin le coefficient Kappa de cette matrice est de 0.64 ce qui signifie un accord fort.

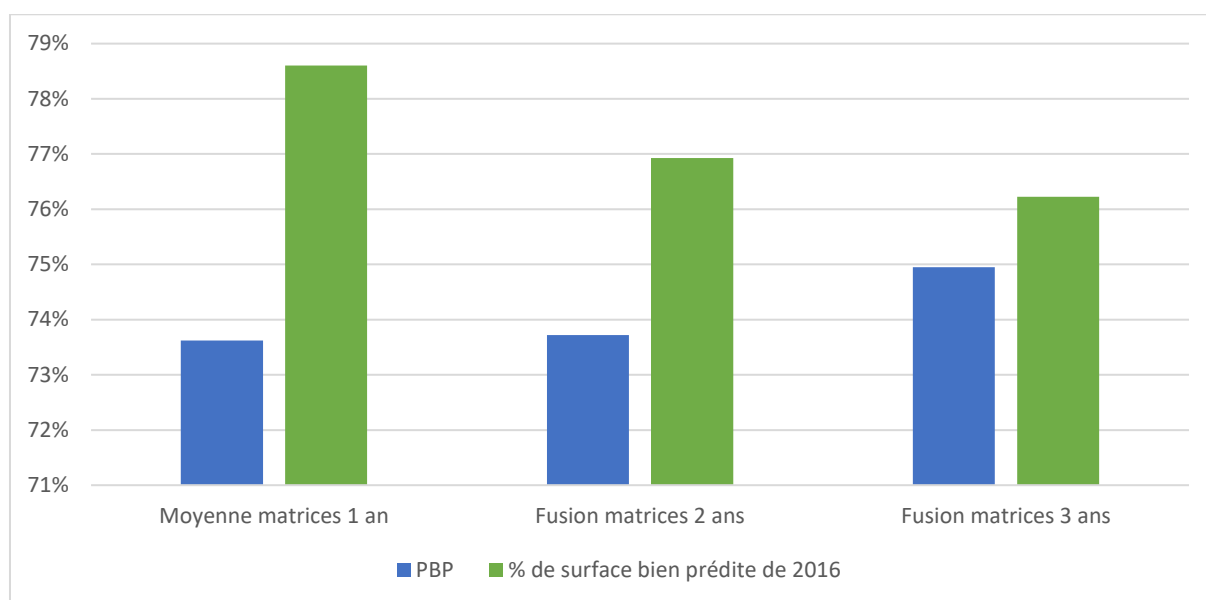


Fig. 29. Groupes de cultures : comparaison PBP et % de surface bien prédite de 2016

Finalement, c'est la matrice à 1 année antérieure qui permet de bien prédire le plus grand pourcentage de surface de 2016 avec 78.6% (Fig.30). Étonnamment, les matrices fusionnées à 3 ans ne donnent que le 3^e meilleur pourcentage de surface alors elles possédaient les meilleurs PBP. Enfin, contrairement à l'application à 23 cultures, l'application à 4 groupes donne de meilleurs pourcentages de surface bien prédite que de PBP.

3.3 Prédiction par réseau de neurones

Cette partie présente les résultats obtenus via la méthode des réseaux de neurones. Cette méthode est détaillée au point 3.1.3. Dans cette partie nous présenterons tout d'abord les soucis et autres résultats de la mise en place du réseau de neurones. Nous présenterons ensuite les résultats de la phase de calibration. Ceux-ci seront suivis par les résultats de la phase de validation. Et nous finirons par les résultats de la méthode appliquée aux 4 groupes de cultures.

3.3.1 Mise en place du réseau de neurones

Les années qui ont été utilisées comme input dans nos résultats sont principalement 1 année antérieure. Dû à leur fonctionnement comme une boîte noire, les réseaux de neurones mettent sur un pied d'égalité tous les inputs (voir Fig1.). Il n'y a donc pas de succession ou de hiérarchie entre les données selon leur année. Ainsi, la calibration s'est faite principalement en utilisant 2014 comme input et 2015 comme output. Et la validation s'est faite principalement en utilisant 2015 comme input et 2016 comme output. Les ajouts des années 2013 pour la calibration et 2014 pour la validation ont tout de même été analysés afin d'évaluer l'impact de deux années précédentes sur la prédiction.

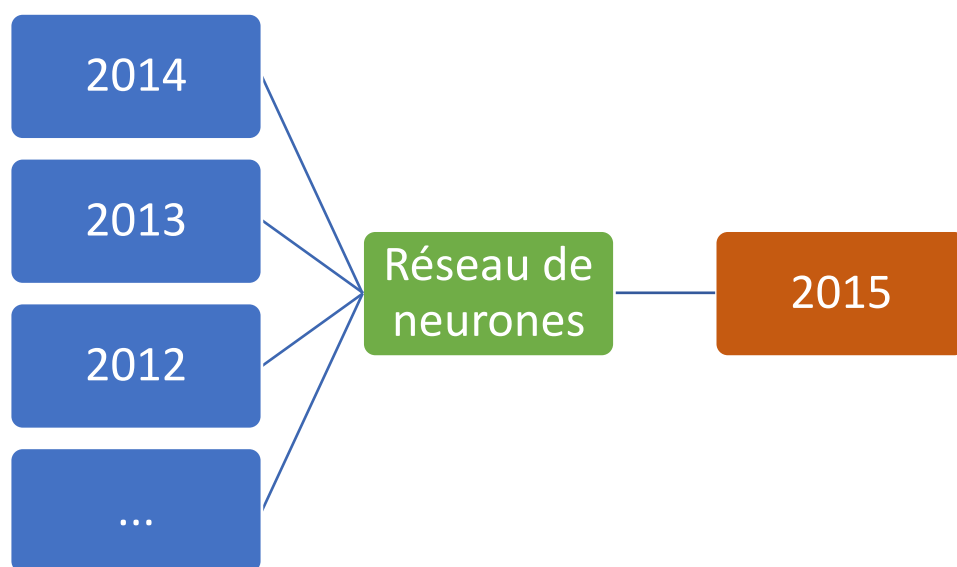


Fig. 30. Réseau de neurone à plusieurs années d'input mises sur un pied d'égalité

Tout comme pour la méthode des matrices de transition, ce sont 23 codes cultures qui sont retenus afin de réaliser la prédiction. Cependant, certains codes cultures ne sont pas toujours présents chaque année dû à leur faible nombre de parcelles dans notre zone d'étude. Malheureusement, ces cultures n'avaient pas été reclassifiés lors de la méthode des matrices de transition. Cependant, la méthode des réseaux de neurones ne fonctionne pas de manière aussi modulable que celle des matrices de

transition. Les réseaux de neurones fonctionnent comme une boîte noire. Et cette boîte noire a besoin d'exactly le même nombre d'inputs lors de la calibration que lors de la validation. Dans notre cas, c'est le code 903 (pomme de terre féculière) qui a dû être changé car il n'était pas présent en 2014. Celui-ci a été reclassifié en code 901 (pomme de terre non hâtive).

Finalement les inputs utilisés pour une année précédente sont au nombre de 24. Parmi ceux-ci, il y a tout d'abord les 22 codes cultures restants représentés de façon binaire. Il y a ensuite l'aire et le périmètre de l'année précédente qui ont été normalisés. Enfin, il est possible d'utiliser les codes cultures de l'année d'il y a 2 ans. Dans ce cas, il faut rajouter 22 inputs ce qui porte le total d'inputs à 46.

3.3.2 Phase de calibration

Impact individuel des types d'inputs sur l'erreur

Les 3 types d'inputs principaux sont donc l'aire, le périmètre et les cultures de 2014. Chaque input a tout d'abord été testé individuellement (Tableau 29.). Ces inputs ont tous été analysés à partir de mêmes paramètres : 1 couche de 5 neurones et un seuil de 0.01. Cependant, il faut savoir que de nombreux paramètres différents ont été testés pour les neurones et le seuil. Mais dans ce cas-ci, l'exactitude était toujours similaire, seul le temps de calcul était différent.

Tableau 29. Phase de calibration : impact individuel des types d'inputs.

Input (total)	Output (total)	Neurones	Seuil	NBP	NPP	Exactitude (%)
Péri2014 (1)	Cult2015 (22)	5	0.01	1328	4299	30.91
Aire2014 (1)	Cult2015 (22)	5	0.01	1446	4299	33.63
Cult2014 (22)	Cult2015 (22)	5	0.01	2178	4299	50.66

Les cultures de 2014 seules permettent de prédire les cultures de 2015 avec une exactitude moins importante que l'utilisation de matrices de transition (voir Fig.23). Le périmètre seul et l'aire seule permettent de prédire en partie la culture de 2015. L'aire semble avoir un impact plus important cependant. Ces deux inputs semblent donc pouvoir assister la prédiction à partir des cultures de 2014.

Combinaison de types d'inputs

L'impact de plusieurs inputs sont analysés lorsqu'ils sont mutuellement combinés. Pour chaque combinaison, de nombreux paramètres de neurones et de seuil ont été testés. Les résultats d'exactitude variaient selon les paramètres mais seuls les meilleurs résultats pour chaque type d'inputs sont renseignés dans le tableau 30.

Les paramètres du nombre de neurones doivent en général se situer entre le nombre d'input et le nombre d'output. Il n'existe ensuite pas de règle fixe, ceux-ci doivent être testés par essai/erreur. Au niveau du nombre de couches, il est généralement suffisant de se limiter à 1 couche. Dans notre cas, 2 et 3 couches ont aussi été testées. Cependant, au plus le nombre de neurones et de couches augmentent, au plus il y a de combinaisons à calculer. Ajouter des couches et des neurones prend donc plus de temps afin de trouver des bons résultats. Ici, pour la combinaison « Aire2014+Péri2014 » par exemple, le nombre de neurones qui donne les meilleurs résultats est « 10 » et se trouve compris entre le nombre d'inputs (2) et outputs (22). C'est pareil pour les autres exemples qui utilisent 23 neurones (entre 23 et 22) et 40 neurones (entre 46 et 22).

Le seuil minimum possible à atteindre augmente avec le nombre d'inputs, le nombre de neurones et le nombre de couches de neurones. C'est ce qui explique aussi que l'ordinateur prenne plus de temps afin de trouver des résultats en dessous d'un certain seuil. Le seuil est un paramètre qu'il faut donc diminuer petit à petit après avoir fixé les paramètres de neurones. Mettre un seuil trop bas fera que l'ordinateur mettra trop de temps pour arriver à des résultats, et mettre un seuil trop élevé augmentera l'erreur sur l'exactitude.

Tableau 30. Phase de calibration : Combinaison de type d'inputs.

Input (total)	Outputs	Neurones	Seuil	NBP	NPP	Exactitude (%)
Aire2014 + Péri2014 (2)	22	10	0.016	1498	4299	34.84
Péri2014 + Cult2014 (23)	22	23 23	0.12	2373	4299	55.2
Aire2014 + Cult2014 (23)	22	23 23	0.11	2433	4299	56.6
Aire2014 + Péri2014 + Cult2014 (24)	22	23 23 23	0.24	2609	4299	60.7
Aire2014 + Péri2014 + Cult2014 + Cult2013 (46)	22	40	0.1	2916	4244	68.7

Tout d'abord, le tableau 30 montre que lorsque deux types d'inputs sont utilisés ensembles, l'exactitude augmente de manière intéressante. Le périmètre associé à l'aire permet d'augmenter un petit peu l'exactitude par rapport à l'utilisation de l'aire seule. L'exactitude obtenue en associant les cultures de 2014 et l'aire ou le périmètre est proche des meilleurs résultats avec matrice de transition (56%). Et c'est l'aire qui améliore le plus l'exactitude de la prédiction.

Ensuite, lorsque 3 types d'inputs sont utilisés, l'exactitude augmente encore. Utiliser le périmètre et l'aire avec les cultures de 2014 semble donner des résultats a priori plus intéressants que les matrices de transition.

Enfin, cette exactitude augmente encore de manière très importante lorsque les cultures de 2013 sont ajoutées. Afin de vérifier si l'ajout d'input permet vraiment de gagner en bonne prédiction, il faut vérifier le fonctionnement de ces réseaux de neurones par la phase de validation.

3.3.3 Phase de validation

La phase de validation donne des résultats qui ne sont à priori pas en adéquation avec la phase de calibration (Tableau 31.). Premièrement, lors de l'utilisation d'uniquement la culture précédente comme input, le pourcentage de surface bien prédite est inférieur au pourcentage de l'exactitude de calibration. La comparaison n'est pas la plus pertinente vu que l'un représente le nombre de parcelle et l'autre la superficie. Cependant, cela montre que des corrélations existent entre ces deux résultats. Soit les surfaces des parcelles prédites sont des parcelles plus petites que la moyenne de toutes les parcelles, soit l'application du réseau de neurone donne des résultats moins bons car celui-ci a été calibré sur une année connue.

Tableau 31. Phase de validation

Input de calibration	Input de validation	Surface ¹ (km ²)	Surface ¹ (%)
Cult_2014	Cult_2015	59.55	47.84
Péri2014 + Cult2014	Péri2015 + Cult2015	59.55	47.84
Aire2014 + Cult2014	Aire2015 + Cult2015	59.55	47.84
Aire2014 + Péri2014 + Cult2014	Aire2015 + Péri2015 + Cult2015	59.55	47.84
Aire2014 + Péri2014 + Cult2014 + Cult2013	Aire2015 + Péri2015 + Cult2015 + Cult2014	50.15	40.29

1. Surface bien prédite

Deuxièmement, la surface prédite par validation n'augmente pas malgré l'utilisation de réseaux de neurones qui augmentaient l'exactitude lors de la calibration. Cette surface prédite diminue même lorsque la culture de 2013 est utilisée. Ce phénomène peut être expliqué par de l'overfitting. En effet, le modèle arrive à trouver les meilleurs paramètres pour la calibration mais ceux-ci ne sont pas appropriés pour la validation. L'aire et le périmètre sont ici des inputs qui permettent de trouver les meilleurs paramètres mais seulement pour l'année en question lors de la calibration. Lorsque les paramètres sont réutilisés pour l'année inconnue lors de la validation, les inputs aire et périmètre possèdent des valeurs trop différentes et ne permettent pas de mieux prédire. Une explication de cet overfitting peut être la différence de type de variables utilisées. Les variables de cultures étant binaires et les valeurs d'aire et de périmètre étant des variables continues.

Enfin, le pourcentage de surface maximale bien prédite de 2016 est de 48%. C'est moins bon que la meilleure méthode qui utilisait les matrices de transition à 2 années et qui obtenait 54% de surface bien prédite.

Les pistes d'amélioration sont nombreuses. Tout d'abord, de nombreux inputs différents pourraient être utilisés comme : le type de sol et sa texture, la pente moyenne de la parcelle, l'agriculteur qui cultive la parcelle, etc. Ensuite, il faudrait trouver un moyen d'éviter l'overfitting afin de pouvoir utiliser d'autres informations que la culture précédente. Pour ce faire, une méthode appelée « régularisation » peut fonctionner. Cependant, il faudrait s'intéresser plus profondément au fonctionnement des réseaux de neurone afin de pouvoir l'appliquer.

3.3.4 Application aux groupes de cultures

Enfin, tout comme pour la méthode des matrices de transition, une méthode d'application à seulement 4 groupes de culture a été réalisée afin de prédire une plus grande surface de culture (Tableau 4.) Celle-ci a été réalisée seulement avec regroupement avant l'utilisation de réseaux de neurones.

Tableau 32. Groupes de cultures : calibration.

Input (total)	Outputs	Neurones	Seuil	NPBP	NPP	Exactitude (%)
Cult2014 (4)	4	4	0.001	3031	4299	70.5
Cult2014 + Aire2014 (5)	4	5 5	0.02	3055	4299	71.06
Cult2014 + Aire2014 + Péri2014 (6)	4	5 5 5	0.05	3074	4299	71.5

Premièrement, comme pour les 22 cultures précises, l'exactitude de calibration augmente avec le nombre d'inputs utilisés. Cependant cette augmentation est minime. C'est sûrement parce que parmi les 4 groupes de cultures (Hiver, Printemps, Prairie et Autre), les spécificités liées aux aires et périmètres des parcelles sont réduites.

Tableau 33. Groupes de cultures : validation.

Input de calibration	Input de validation	Surface ¹ (km ²)	Surface(km ²)
Cult2014	Cult2015	97.83	78.60
Cult2014 + Aire2014	Cult2015 + Aire2015	97.59	78.41
Cult2014 + Aire2014 + Péri2014	Cult2015 + Aire2015 + Péri2015	97.55	78.37

1. Surface bien prédite

Deuxièmement, le pourcentage de surface bien prédite diminue avec le nombre d'inputs utilisés. L'aire et le périmètre ont donc un impact négatif sur la prédiction dans ce cas-ci. Le pourcentage maximal de surface bien prédite est celui qui utilise seulement la culture de l'année précédente. Celui-ci est identique au pourcentage trouvé grâce à la méthode des matrices de transition de 1 année. L'hypothèse serait qu'ici le réseau de neurone fonctionne de la même manière que la matrice de transition.

Enfin, le pourcentage de surface bien prédite lors de la validation est plus important que l'exactitude de calibration. La raison principale est forcément que les parcelles bien prédites possèdent une surface plus importante que la surface moyenne des parcelles. L'autre raison serait que le réseau de neurone arrive à mieux prédire 2016 avec 2015 que 2015 avec 2014.

Conclusion

Dans la perspective de prédire les inondations boueuses de plus en plus fréquentes en Belgique centrale, ce mémoire a évalué deux approches de prédiction de l'occupation du sol. Ces approches basées sur les rotations de cultures sont les matrices de transitions et les réseaux de neurones.

Dans un premier temps, l'analyse de l'état de l'art nous a appris que le domaine de la prédiction de cultures au niveau de la parcelle en est encore à ses débuts. Peu de méthodes à ce jour obtiennent des résultats significatifs. La plupart de ces approches se basent sur les chaînes de Markov et ne s'intéressent pas aux réseaux de neurones. Ces approches utilisent les informations de l'occupation du sol datant de plusieurs années précédentes afin de prédire les occupations culturelles qui suivront.

Deuxièmement, l'analyse des données reprenant les années antérieures a mis en avant certains points importants. Tout d'abord, ces données doivent être utilisées en étant limitées à leurs régions agricoles et agro-géographiques spécifiques. Cependant, ces données doivent en parallèle représenter la plus grande zone géographique possible afin de posséder un échantillon de variables le plus représentatif. Ensuite, la classification de ces données par la Région devrait être plus constante suivant les années. En effet, un des freins à la bonne identification des rotations est un manque d'harmonisation dans la classification de ces cultures. Enfin, l'assemblage des données de plusieurs années ne doit pas être négligé. Afin d'observer une logique de rotation, la meilleure méthode semble d'intersecter les centroïdes des parcelles de la dernière année connue avec les parcelles des années précédentes.

Ensuite les résultats des méthodes de prédiction varient selon l'approche. Les matrices de transition et réseaux de neurones semblent coïncider plus ou moins au niveau des résultats lorsqu'une seule année précédente est utilisée. Cependant, si plusieurs années précédentes sont utilisées, les matrices de transition sont plus faciles à mettre en place et donnent de meilleurs résultats. De plus, l'aspect stochastique des matrices de transition permet de mieux comprendre la logique dans la prédiction tandis que les réseaux de neurones fonctionnent plutôt comme des boîtes noires. Le principal intérêt des réseaux de neurones était d'intégrer des données comme l'aire ou le périmètre des dernières parcelles connues pour prédire la prochaine culture. Cette tentative a échoué mais cette approche reste une possibilité si le problème d'overfitting est résolu.

Parmi les résultats des matrices de transition, ceux-ci permettent une prédiction de 54% de la surface agricole en se basant sur 23 types de cultures et 79% de la surface en se basant sur 4 groupes de cultures (hiver, printemps, prairie, autre). Ces prédictions pourraient être réalisées par l'administration dès la clôture des demandes d'aide via les déclarations de superficie à la PAC (fin mai de l'année précédente à celle que l'on souhaite prédire). Malheureusement le parcellaire agricole ne semble

disponible publiquement qu'à partir de janvier de l'année même de prédiction. Cela limite donc l'intérêt de cette démarche.

Dans le cas des 23 types de cultures, la meilleure surface prédite a été réalisée en utilisant 2 et 3 années précédentes. Cependant, cette prédiction n'a pu se faire qu'en fusionnant les différentes matrices de transition des années précédentes. En effet, il a été mis en évidence que si le nombre de classes de cultures est important et que plusieurs années précédentes sont combinées, il est nécessaire de fusionner les matrices de transition.

Dans le cas des 4 groupes de cultures, la meilleure surface prédite a été réalisée en utilisant une seule année précédente. Néanmoins, l'utilisation de 2 et 3 années permet d'obtenir des résultats proches. Ensuite, deux approches ont été comparées. D'une part, une prédiction avec matrices de transition reprenant les cultures selon les 4 groupes d'une part. Et d'autre part, une méthode qui reprend les résultats de prédiction des 23 cultures et les reclasse ensuite selon les 4 groupes. L'approche regroupant les cultures avant prédiction donne dans tous les cas de meilleurs résultats.

Pour finir, un travail mériterait d'être mené en complexifiant la méthode des matrices de transition. Osman et al. (2015) et les Markov Logic Network semblent aller dans ce sens et de meilleurs résultats pourraient être obtenus. De plus, la prise en compte d'images satellites pourrait également aider à améliorer la prédiction. Celles-ci pourraient, en hiver, permettre de déjà identifier les sol nus et couverts. Enfin, l'impact de la prédiction de la culture sur les inondations boueuses doit être déterminé.

Bibliographie

- Albrecht, H., 2003. Suitability of arable weeds as indicator organisms to evaluate species conservation effects of management in agricultural ecosystems. *Agric. Ecosyst. Environ.* 98, 201–211.
- Aurbacher, J., Dabbert, S., 2011. Generating crop sequences in land-use models using maximum entropy and Markov chains. *Agric. Syst.* 104, 470–479. <https://doi.org/10.1016/j.agsy.2011.03.004>
- Auzet, A.-V., 1987. L'érosion des sols cultivés en France sous l'action du ruissellement. *Ann. Géographie* 96, 529–556. <https://doi.org/10.3406/geo.1987.20627>
- Bachinger, J., Zander, P., 2007. ROTOR, a tool for generating and evaluating crop rotations for organic farming systems. *Eur. J. Agron.* 26, 130–143. <https://doi.org/10.1016/j.eja.2006.09.002>
- Biélders, C., 2016. LBIRE2105 Erosion et conservation du sol - 8.Méthodes de conservation du sol.
- Biélders, C., Maignard, A., Feltz, N., Cordonnier, H., Degré, A., Destain, M.-F., Colard, F., 2011. Convention GISER - Convention de recherche d'intérêt général gestion intégrée sol érosion ruissellement.
- Boardman, J., Stammers, R., 1984. Soil erosion and flooding on downland areas. *Surveyor* 164, 8–11.
- Boardman, J., Verstraeten, G., Biélders, C., 2006. Muddy Floods, in: *Soil Erosion in Europe*. John Wiley & Sons, Ltd, pp. 743–755. <https://doi.org/10.1002/0470859202.ch53>
- Bogaert, P., Waldner, F., Defourny, P., 2016. An information-based criterion to measure pixel-level thematic uncertainty in land cover classifications. *Stoch. Environ. Res. Risk Assess.* <https://doi.org/10.1007/s00477-016-1310-y>
- Boller, E.F., Malavolta, C., Jörg, E., 1997. Guidelines for Integrated Production of Arable Crops in Europe: IOBC Technical Guideline III. *IOBC Wprs Bull.* 20.
- Borselli, L., Cassi, P., Torri, D., 2008. Prolegomena to sediment and flow connectivity in the landscape: A GIS and field numerical assessment. *CATENA* 75, 268–277. <https://doi.org/10.1016/j.catena.2008.07.006>
- Boryan, C., Yang, Z., Mueller, R., Craig, M., 2011. Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. *Geocarto Int.* 26, 341–358. <https://doi.org/10.1080/10106049.2011.562309>
- Campbell, J.B., Wynne, R.H., 2011. *Introduction to Remote Sensing*, Fifth Edition. Guilford Press.

- Castellazzi, M.S., Wood, G.A., Burgess, P.J., Morris, J., Conrad, K.F., Perry, J.N., 2008. A systematic representation of crop rotations. *Agric. Syst.* 97, 26–33.
- Climate-data.org, 2012. Climat Orp-Jauche: Diagramme climatique, Courbe de température, Table climatique pour Orp-Jauche - Climate-Data.org [WWW Document]. URL <https://fr.climate-data.org/location/214013/> (accessed 1.12.18).
- Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* 37, 35–46.
- Congalton, R.G., Oderwald, R.G., Mead, R.A., 1983. Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques. *Photogramm. Eng. Remote Sens.* 49, 1671–1678.
- Cox, D.R., Miller, H.D., 1977. *The theory of stochastic processes*. CRC Press.
- Dahikar, S.S., Rode, S.V., 2014. Agricultural crop yield prediction using artificial neural network approach. *Int. J. Innov. Res. Electr. Electron. Instrum. Control Eng.* 2, 683–686.
- Dehghan, H., Ghassemian, H., 2006. Measurement of uncertainty by the entropy: Application to the classification of MSS data. *Int. J. Remote Sens.* - INT J REMOTE SENS 27, 4005–4014. <https://doi.org/10.1080/01431160600647225>
- Dendoncker, N., Van Wesemael, B., Rounsevell, M.D., Roelandt, C., Lettens, S., 2004. Belgium's CO₂ mitigation potential under improved cropland management. *Agric. Ecosyst. Environ.* 103, 101–116.
- Detlefsen, N., 2004. Crop rotation modelling, in: *Proceedings of the EWDA-04 European Workshop for Decision Problems in Agriculture and Natural Resources*. Silsoe Research Institute, England. pp. 5–14.
- Dewalque, B., 2016. Caractérisation du risque d'inondation boueuse sur les récepteurs de risque lors d'événements pluvieux intenses.
- Dogliotti, S., Rossing, W.A.H., Van Ittersum, M.K., 2003. ROTAT, a tool for systematically generating crop rotations. *Eur. J. Agron.* 19, 239–250.
- El-Nazer, T., McCarl, B.A., 1986. The Choice of Crop Rotation: A Modeling Approach and Case Study. *Am. J. Agric. Econ.* 68, 127. <https://doi.org/10.2307/1241657>
- EOEdu, 2016. EOEdu - Petit guide de la télédétection - La résolution [online] [WWW Document]. URL <http://eoeu.belspo.be/fr/guide/resolu.asp?section=2.5.2> (accessed 7.17.17).

- ESA, 2017. Spatial - Resolutions - Sentinel-2 MSI - User Guides - Sentinel Online [WWW Document]. URL <https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/spatial> (accessed 7.17.17).
- Evrard, O., Biielders, C.L., Vandaele, K., van Wesemael, B., 2007. Spatial and temporal variation of muddy floods in central Belgium, off-site impacts and potential control measures. *CATENA* 70, 443–454. <https://doi.org/10.1016/j.catena.2006.11.011>
- FAO, 2006. European Union accession and land tenure data in Central and Eastern Europe. Food and Agriculture Organization of the United Nations, Rome.
- Fleiss, J.L., 1981. Statistical Methods for Rates and Proportions. Wiley.
- Fritsch, S., Guenther, F., Guenther, M.F., 2016. Package ‘neuralnet.’
- Gerowitt, B., 2003. Development and control of weeds in arable farming systems. *Agric. Ecosyst. Environ.* 98, 247–254.
- Grandjean, M., 2016. Le foncier agricole face à l’artificialisation des terres en Wallonie.
- Gregorio, A.D., FAO, 2005. Land Cover Classification System: Classification Concepts and User Manual : LCCS. Food & Agriculture Org.
- Guo, W.W., Xue, H., 2014. Crop Yield Forecasting Using Artificial Neural Networks: A Comparison between Spatial and Temporal Models [WWW Document]. *Math. Probl. Eng.* <https://doi.org/10.1155/2014/857865>
- Hamza, M.A., Anderson, W.K., 2005. Soil compaction in cropping systems: a review of the nature, causes and possible solutions. *Soil Tillage Res.* 82, 121–145.
- Haneveld, W.K.K., Stegeman, A.W., 2005. Crop succession requirements in agricultural production planning. *Eur. J. Oper. Res.* 166, 406–429. <https://doi.org/10.1016/j.ejor.2004.03.009>
- Inglada, J., Vincent, A., Arias, M., Marais-Sicre, C., 2016. Improved Early Crop Type Identification By Joint Use of High Temporal Resolution SAR And Optical Image Time Series. *Remote Sens.* 8, 362. <https://doi.org/10.3390/rs8050362>
- Jankauskas, B., Jankauskiene, G., 2003. Erosion-preventive crop rotations for landscape ecological stability in upland regions of Lithuania. *Agric. Ecosyst. Environ.* 95, 129–142. [https://doi.org/10.1016/S0167-8809\(02\)00100-7](https://doi.org/10.1016/S0167-8809(02)00100-7)

- Klöcking, B., Ströbl, B., Knoblauch, S., Maier, U., Pfützner, B., Gericke, A., 2003. Development and allocation of land-use scenarios in agriculture for hydrological impact studies. *Phys. Chem. Earth Parts ABC* 28, 1311–1321. <https://doi.org/10.1016/j.pce.2003.09.007>
- Labatut, V., Cherifi, H., 2012. Accuracy measures for the comparison of classifiers. *ArXiv Prepr. ArXiv12073790*.
- Le Ber, F., Benoît, M., Schott, C., Mari, J.-F., Mignolet, C., 2006. Studying crop sequences with CarrotAge, a HMM-based data mining software. *Ecol. Model.* 191, 170–185. <https://doi.org/10.1016/j.ecolmodel.2005.08.031>
- Leteinturier, B., Tychon, B., Oger, R., 2007. Diagnostic agronomique et agro-environnemental des successions culturales en Wallonie (Belgique). *BASE*.
- Lobell, D.B., Asner, G.P., 2004. Cropland distributions from temporal unmixing of MODIS data. *Remote Sens. Environ.* 93, 412–422. <https://doi.org/10.1016/j.rse.2004.08.002>
- Na-udom, A., Rungrattanaubol, J., 2015. A Comparison of Artificial Neural Network and Regression Model for Predicting the Rice Production in Lower Northern Thailand, in: Kim, K.J. (Ed.), *Information Science and Applications*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 745–752. https://doi.org/10.1007/978-3-662-46578-3_88
- Osman, J., Inglada, J., Dejoux, J.-F., 2015. Assessment of a Markov logic model of crop rotations for early crop mapping. *Comput. Electron. Agric.* 113, 234–243. <https://doi.org/10.1016/j.compag.2015.02.015>
- Paoletti, M.G., 1999. Using bioindicators based on biodiversity to assess landscape sustainability. *Agric. Ecosyst. Environ.* 74, 1–18.
- Paswan, R.P., Begum, S.A., 2013. Regression and Neural Networks Models for Prediction of Crop Production 1.
- Pontius, R.G., Millones, M., 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *Int. J. Remote Sens.* 32, 4407–4429. <https://doi.org/10.1080/01431161.2011.552923>
- R Development Core Team, 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Ranjeet, T., Armstrong, L., 2014. An Artificial Neural Network for Predicting Crops Yield in Nepal. *ECU Publ. Post* 2013.
- Renard, K.G., Foster, G.R., Weesies, G.A., Porter, J.P., 1991. RUSLE: Revised universal soil loss equation. *J. Soil Water Conserv.* 46, 30–33.

- Renard, K.G., USA, USA (Eds.), 1997. Predicting soil erosion by water: a guide to conservation planning with the revised universal soil loss equation (RUSLE), Agriculture handbook. Washington, D. C.
- Ressources naturelles Canada, 2008a. Polarimétrie [online] [WWW Document]. Rncan.gc.ca. URL <http://www.rncan.gc.ca/sciences-terre/geomatique/imagerie-satellitaire-photos-aeriennes/imagerie-satellitaire-produits/ressources-educatives/9276> (accessed 7.17.17).
- Ressources naturelles Canada, 2008b. Résolution temporelle [online] [WWW Document]. URL <http://www.rncan.gc.ca/sciences-terre/geomatique/imagerie-satellitaire-photos-aeriennes/imagerie-satellitaire-produits/ressources-educatives/9366> (accessed 1.8.18).
- Ressources naturelles Canada, 2008c. Classification et analyse des images [online] [WWW Document]. URL <https://www.rncan.gc.ca/sciences-terre/geomatique/imagerie-satellitaire-photos-aeriennes/imagerie-satellitaire-produits/ressources-educatives/9362> (accessed 7.20.17).
- Revet, S., 2017. Anthropologie d'une catastrophe: Les coulées de boue de 1999 au Venezuela. Presses Sorbonne Nouvelle.
- Salmon-Monviola, J., Durand, P., Ferchaud, F., Oehler, F., Sorel, L., 2012. Modelling spatial dynamics of cropping systems to assess agricultural practices at the catchment scale. *Comput. Electron. Agric.* 81, 1–13. <https://doi.org/10.1016/j.compag.2011.10.020>
- Schönhart, M., Schmid, E., Schneider, U.A., 2011. CropRota – A crop rotation model to support integrated land use assessments. *Eur. J. Agron.* 34, 263–277. <https://doi.org/10.1016/j.eja.2011.02.004>
- Seutin, B., Vancutsem, F., Bodson, B., 2011. 9. Environnement - 2. La culture en association de froment et de légumineuses.
- SillonBelge.be, 2017. Pommes de terre: lutter contre les repousses, c'est essentiel! [WWW Document]. SillonBelge.be. URL <http://www.sillonbelge.be/903/article/2017-06-16/pommes-de-terre-lutter-contre-les-repousses-cest-essentiel> (accessed 1.13.18).
- Supit, I., van Diepen, C.A., de Wit, A.J.W., Wolf, J., Kabat, P., Baruth, B., Ludwig, F., 2012. Assessing climate change effects on European crop yields using the Crop Growth Monitoring System and a weather generator. *Agric. For. Meteorol.* 164, 96–111. <https://doi.org/10.1016/j.agrformet.2012.05.005>
- Van der Wel, F.J., Van der Gaag, L.C., Gorte, B.G., 1998. Visual exploration of uncertainty in remote-sensing classification. *Comput. Geosci.* 24, 335–343.

- Verstraeten, G., Poesen, J., 1999. The nature of small-scale flooding, muddy floods and retention pond sedimentation in central Belgium. *Geomorphology* 29, 275–292. [https://doi.org/10.1016/S0169-555X\(99\)00020-3](https://doi.org/10.1016/S0169-555X(99)00020-3)
- Wechsung, F., Krysanova, V., Flechsig, M., Schaphoff, S., 2000. May land use change reduce the water deficiency problem caused by reduced brown coal mining in the state of Brandenburg? *Landsc. Urban Plan.* 51, 177–189.
- Wibberley, J., 1996. A brief history of rotations, economic considerations and future directions. *Asp. Appl. Biol. U. K.*
- Wischmeier, W.H., C, D.D., 1978. Predicting rainfall erosion losses - A guide to conservation planning, in: Wischmeier, W.H., Smith, D.D. (Eds.), *Paper. Department of Agriculture, USA*, pp. 1–69.
- Xiao, Y., Mignolet, C., Mari, J.-F., Benoît, M., 2014. Modeling the spatial distribution of crop sequences at a large regional scale using land-cover survey data: A case from France. *Comput. Electron. Agric.* 102, 51–63. <https://doi.org/10.1016/j.compag.2014.01.010>

Évaluation de deux méthodes de prédiction de l'occupation du sol basées sur les rotations de cultures dans une perspective de prédiction d'inondations boueuses

Présenté par Miguel Urdiain-Arraiza

Résumé

La Belgique centrale est de plus en plus touchée par des inondations boueuses ces dernières années. L'une des causes de ces inondations boueuses est la présence de certaines cultures favorisant l'érosion. La prédiction des types de cultures plusieurs mois avant les événements orageux permettrait de prédire ces inondations boueuses et permettrait aux communes d'identifier les sites à risques afin d'anticiper les problèmes.

L'objectif de cette étude est de prédire cette occupation du sol en se basant sur les rotations de cultures. L'étude s'intéresse aux méthodes probabilistes comme les chaînes de Markov et les réseaux de neurones. Ce travail analyse tout d'abord les données d'occupations du sol des années précédentes au niveau temporel et géographique. Il prédit ensuite l'occupation du sol selon 23 types de cultures et selon 4 groupes de cultures par les deux méthodes citées.

Il ressort de travail que l'utilisation des chaînes de Markov est plus facile à mettre en place et donne de meilleurs résultats. De plus, celle-ci possède un aspect stochastique qui permet de mieux comprendre la logique de prédiction. Cette méthode permet de prédire, dès janvier, 54% de la surface agricole lorsque classifiée en 23 cultures et 79% lorsque classifiée en 4 groupes de cultures (hiver, printemps, prairie, autres).

A partir de ces conclusions, il faut reconnaître que les résultats de ce mémoire semblent peu exploitables pour être intégrés dans un système d'alerte d'inondations boueuses. Il faudrait les améliorer et continuer la recherche dans le domaine des chaînes de Markov. Ainsi, une méthode plus complexe comme les Markov Logic Network pourrait être envisagée. Enfin, la prise en compte d'images satellites pourrait également aider à améliorer la prédiction.