

WALOUS

D5.1 : Guide méthodologique finalisé et recommandations d'usages pour la fusion des données de la carte d'occupation du sol de Wallonie 2018

Subvention financée par le Service Public de Wallonie, DGO3 et le Département de
la Géomatique du Secrétariat Général.

Février 2019

OBJET DU RAPPORT

Ce rapport est composé en trois parties :

- La première partie du rapport évalue la qualité des données de classification (précision globale) utilisées dans le cadre de la fusion de données pour la création de la carte d'occupation des sols ;
- La seconde partie décrit brièvement les différentes approches utilisées pour la fusion des données, l'approche choisie pour généraliser la méthode sur la Région wallonne et la méthodologie de cette dernière ;
- La troisième partie évalue la qualité de la carte d'occupation des sols de Wallonie, issue de la fusion de différentes sources de données par machine learning (classification Random Forest).

1 ANALYSE DE LA QUALITÉ DES DONNÉES POUR LA FUSION

Les données utilisées pour créer la carte d'occupation du sol sont :

- Une classification de l'occupation du sol complète obtenue par une approche orientée objet sur base des orthophotos 2018 (OBIA) ;
- Une classification de l'occupation du sol complète obtenue par une approche orientée pixel sur base des orthophotos 2018 ;
- Une classification de l'occupation du sol complète obtenue par une approche orientée pixel sur base des images Sentinel 1&2 2018 ;
- Un masque forestier fourni par l'Université Agro-Bio Tech de Gembloux comprenant 3 classes d'occupations du sol (feuillus, résineux et sols nus en forêts) de 2016 ;
- Deux classifications des terres arables sur base des images Sentinel 2 de 2016 obtenues en utilisant le logiciel Sent2Agri : un masque des terres arables et une classification des types de terres arables ;
- Le SIGEC qui est une donnée de référence fournissant une information sur les parcelles de terres arables et les herbages déclarés en Wallonie pour 2018 ;
- Une information de stratification ; cette donnée a été générée sur base des classifications par pixel et par objet des orthophotos 2018 en distinguant 3 strates : une première strate lorsque les classifications étaient en accord, une seconde lorsqu'elles étaient en désaccord et une troisième basée sur la distance géographique à chaque classe (buffer autour des classes).

Pour rappel, la légende utilisée pour la carte d'occupation du sol est reprise à la Figure 1.

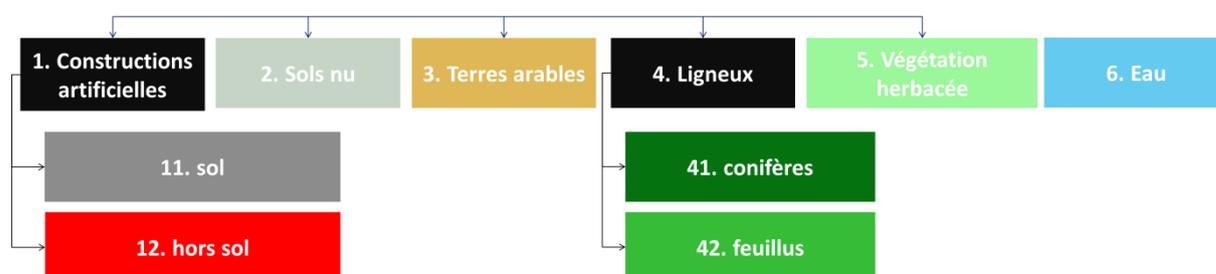


Figure 1 - Légende de la carte d'occupation du sol fusionnée

1.1 MÉTHODOLOGIE D'ANALYSE DE LA QUALITÉ DES DONNÉES

La qualité est déterminée sur base de la précision globale de chaque classification calculée via la matrice de confusion construite par rapport à un set de 1460 points de validation photointerprétés.

1.1.1 POINTS DE VALIDATION

Le plan d'échantillonnage est stratifié et aléatoire. Les strates correspondent aux classes d'occupation du sol de la classification par pixel des orthos 2018. Le nombre de points par strates a été optimisé selon les « bonnes pratiques » préconisées par Olofsson et al (2014). L'optimisation se base sur une estimation de l'exactitude par classe obtenue lors de l'analyse préliminaire des quatre zones d'étude.

Ces points ont été photo-interprétés par 3 experts pour être utilisés dans le calcul des matrices pour chaque donnée. Un poids a été attribué à chaque point, correspondant au rapport entre la surface de la strate dans laquelle il a été sélectionné sur la surface totale des strates et sur le nombre de points sélectionnés dans cette strate (Card, 1982). Ceci revient à pondérer la validation en fonction de la proportion de l'espace régional occupé par les différentes classes d'occupation du sol.

Le tableau 1 reprend le nombre de points optimal et le poids de chaque point, par strate :

Tableau 1 : Nombre de points optimal et poids de chaque point, par strate, choisis sur base des exactitudes des classes en sortie de fusion sur les 4 zones d'étude.

Classe	Nombre de points	Poids de chaque point de la classe
Constructions artificialisées au sol	186	1,47
Constructions artificialisées hors sol	98	2,46
Eau	62	0,66
Herbage	533	5,17
Terres arables	220	14,3
Sols nus	92	1,52
Feuillus	104	22,52
Résineux	205	5,18

1.1.2 MATRICE DE CONFUSION

Pour créer les matrices de confusion, nous avons utilisé l'outil d'analyse spatiale « extract values to points » de ArcGIS. Un équivalent de cet outil existe également dans QGIS : « add raster values to points ». Le calcul de précision globale a ensuite été effectué via un tableau dynamique croisé dans Excel pour chaque classification.

Le calcul de la précision globale de chaque donnée a été effectué sur base du rapport entre la somme des valeurs de la diagonale de la matrice de confusion et la somme des valeurs de toutes les cellules de la matrice de confusion.

1.2 EVALUATION DE LA QUALITÉ DONNÉES D'ENTRÉE DE LA FUSION

1.2.1 CLASSIFICATION OBIA

La classification OBIA ne contient pas la classe « terres arables ». Dès lors, les terres arables classées en herbage ou en sol nu sont prises pour *vrais positifs*. La précision globale de la classification est de 86,8%.

En analysant les classifications visuellement, il ressort que certaines classes se confondent entre elles. Nous avons comparé ces différentes classes entre elles, et les taux de confusions que nous retrouvons sont les suivants :

- Confusion de 10% entre les classes feuillus et résineux ;
- Confusion de 8,3% entre les classes constructions artificialisées au sol et hors sol ;
- Confusion de 18,8% entre les classes constructions artificialisées au sol et sols nus.

La majorité des erreurs systématiques (pas en terme de surface) de cette classification vient donc de la confusion entre les sols nus et les constructions artificialisées au sol.

1.2.2 CLASSIFICATIONS PIXELS

1.2.2.1 Classification sur base des orthophotos 2018

La classification par pixels prenant en compte la classe « terres arables » atteint une précision globale de 73,6%. En omettant cette classe et en considérant que les terres arables classées en herbage sont prises pour *vrais positifs*, la classification atteint une précision de 88,3%.

De même que pour la classification par objet, nous avons comparé plusieurs classes entre elles et les résultats de confusion sont les suivants :

- Confusion de 10,4% entre les classes feuillus et résineux ;
- Confusion de 5,5% entre les classes constructions artificialisées au sol et hors sol ;
- Confusion de 22,1% entre les classes constructions artificialisées au sol et sols nus ;
- Confusion de 26,7% entre les classes terres arables et herbage.

La majorité des erreurs systématiques (pas en terme de surface) vient de la confusion terres arables-herbage et de la confusion sols nus-construction artificialisées au sol.

1.2.2.2 Classification sur base des images Sentinel

La classification sur base des images Sentinel a une résolution de 10m. L'échantillonnage des points a été construit pour la validation de données à une résolution de 1m. Il n'est donc pas adapté à la validation de cette classification. Nous utiliserons tout de même ce jeu de validation afin d'avoir une estimation de la qualité, en restant cependant prudents sur les chiffres de précision globale et l'interprétation de la qualité de la donnée.

La précision globale de la classification de l'occupation du sol sur base des images Sentinel est de 65,3%.

Plusieurs classes ont été comparées afin d'analyser le taux de confusion entre elles :

- Confusion de 8,5% entre les classes feuillus et résineux ;
- Confusion de 12,4% entre les classes constructions artificialisées au sol et sols nus ;

- Confusion de 27,5% entre les classes terres arables et herbage.

1.2.3 CLASSIFICATIONS VIA SENT2AGRI

Pour les mêmes raisons que la classification de l'occupation du sol par les images Sentinel, le jeu de validation n'est pas adapté à l'évaluation des classifications issues de Sent2Agri, mais va être utilisé afin d'avoir une estimation de la qualité des données.

1.2.3.1 Masque des terres arables

Le masque des terres arable est un produit binaire séparant les terres arables annuelles des autres occupations du sol. La précision globale de cette classification au niveau de la classe « terres arables » est de 93%.

1.2.3.2 Typologie des terres arables

Cette donnée reprend 11 classes de terres arables en Wallonie, basées sur la temporalité des images Sentinel 2 pour 2018. La précision globale de la classification pour les classes terres arables vaut 93%.

1.2.4 MASQUE FORESTIER

Le masque forestier prend en compte 3 classes d'occupation du sol : les feuillus, les résineux et les sols nus en forêt. Etant donné que la classification date de 2016, les sols nus en forêts, classe fortement variable sur le moyen terme, n'ont pas été pris en compte pour le calcul de la précision globale ni pour la fusion des données. En ne considérant alors que les feuillus et les résineux du masque forestier, sa précision globale est de 91%. Les points de validation ne tombant pas dans les données du masque mais étant annotées comme feuillus ou résineux sont pris en compte dans le calcul de précision. Le taux de confusion entre ces deux classes est de 10%.

1.2.5 DONNÉES DU SIGEC

Les données du SIGEC ont été utilisées sous forme de deux classes : l'herbage d'une part comprenant les prairies permanentes et les terres arables d'autre part comprenant les cultures et les prairies temporaires. La précision globale de cette information est de 84,9% en tenant compte des points de validation photointerprétés en terres arables mais ne tombant pas dans les données du SIGEC (incomplétude de la donnée).

1.2.6 RÉSUMÉ DES PRÉCISIONS GLOBALES PAR CLASSIFICATION

Le tableau 2 reprend les différentes classifications utilisées dans la fusion avec leur précision globale.

Tableau 2 : Précision globale des données de la fusion

Classification	Précision globale [%]
Orthophotos 2018 par objets – sans distinction des terres arables	86,8
Orthophotos 2018 par pixels – sans distinction des terres arables	88,3
Orthophotos 2018 par pixels – avec distinction des terres arables	73,6
Sentinel - occupation du sol 2018 (10 m de résolution)	65,3
Sent2Agri – masque des terres arables 2017	93,0
Sent2Agri – typologie des terres arables 2017	93,0
Masque forestier 2016 en deux classes	91,0
Données du SIGEC 2018 en deux classes	84,9

2 MÉTHODOLOGIE DE FUSION DES DONNÉES

Trois approches différentes ont été testées afin de fusionner les données de classification en un résultat unique :

- Approche par ensemble de règles de décision (ISSEP). Le détail de la méthode est disponible à l'annexe 1.
- Approche stochastique selon la théorie de « Dempster Shafer » (DS) (UCL). Le détail de la méthode est disponible à l'annexe 2.
- Approche Machine Learning selon la méthode « Random Forest » (RF) (ULB). Le détail de la méthode est développé au point 2.2.

La fusion par la méthode Random Forest a été choisie par l'équipe Walous sur base d'une analyse visuelle. Afin de conforter ce choix, la qualité des trois méthodes de fusion a été évaluée sur les 4 zones test où elles ont été produites

2.1 ANALYSE DE LA QUALITÉ DES FUSIONS SUR LES 4 ZONES D'ÉTUDES

Le tableau 3 reprend le nombre de points par classe parmi le set de 1460 points de validation se trouvant dans les 4 zones d'étude. La stratification des points pour l'évaluation des données ayant été opérée sur toute la Région wallonne, seuls 102 points se situent dans les 4 zones d'étude. Il est donc question de comparer la qualité relative des fusions entre elles plutôt que d'interpréter, sans grande précision, leur exactitude estimée.

Tableau 3 : Nombre de points du jeu de validation par classe dans les 4 zones d'étude

Classe	Nombre de points
Constructions artificialisées au sol	13
Constructions artificialisées hors sol	10
Eau	8
Herbage	37
Terres arables	7
Sols nus	7
Feuillus	12
Résineux	8
Total	102 points

Des premiers résultats de validation de données ont été calculés sur base de ces points et sont présentés ci-après.

2.1.1 FUSION PAR DEMPSTER-SHAFER (SUR LES 4 ZONES TESTS)

Sur base de ces 102 points, la précision globale de la fusion Dempster-Shafer est de 90,6% sur les 4 zones d'étude.

2.1.2 FUSION PAR MACHINE LEARNING (RANDOM FOREST) SUR LES 4 ZONES TESTS

La précision globale de la fusion Random Forest sur les 4 zones test est de 91,9%.

2.1.3 FUSION PAR RÈGLES DE DÉCISION (SUR UNE ZONE TEST)

A cause de la forte dépendance des règles de décision à la zone étudiée, la fusion par règles de décision n'a été effectuée que sur la zone de Marche. Le nombre de points du set de validation sur cette zone est de 17, ce qui ne permet pas d'évaluer correctement la qualité de la fusion. De plus, toutes les classes ne sont pas représentées dans ces 17 points. Cette approche de fusion a été écartée car, étant non automatisable, elle est jugée laborieuse. Certaines règles seront cependant éventuellement reprises lors de la consolidation de la carte d'occupation du sol.

2.2 MÉTHODOLOGIE DE L'APPROCHE DE FUSION PAR LA MÉTHODE RANDOM FOREST

L'approche de fusion par la méthode Random Forest repose sur un paradigme orienté-objet. Dans celui-ci, les segments issus de la segmentation automatisée de la classification d'occupation du sol OBIA (WP3) sont utilisés comme unités spatiales de base pour l'ensemble des traitements.

La méthode repose sur l'algorithme d'apprentissage automatique (Machine Learning) « Random Forest » (RF). Il a été démontré dans la littérature que cette méthode était robuste vis-à-vis du bruit dans les données d'entraînement ainsi qu'à l'autocorrélation entre les données d'entrées. Cette méthode est dite supervisée, car elle nécessite un jeu d'entraînement labélisé pour pouvoir construire le modèle. Dans le cas présent, le jeu d'entraînement est constitué de 3000 points labélisés manuellement au sein des 4 zones tests.

2.2.1 CRÉATION DU JEU D'ENTRAÎNEMENT

Au sein des 4 zones test, un premier jeu de 1500 points labélisés a été produit, répartis aléatoirement selon un échantillonnage stratifié sur base des zones d'incertitude : la première strate représentant les zones où les classifications OBIA et pixel des orthos 2016 étaient en accord, la seconde où elles étaient en discordance et la troisième représentant les zones à proximité des bordures de classe.

Quatre « types » de zones polygonales ont été produites :

- Les zones où la classification pixels et la classification OBIA était en accord (A) ;
- Les zones d'ombres (B) ;
- Les zones à proximité des bordures des différentes classes (C) ;
- Les zones dans des classes mixtes/en désaccord entre la classification pixels et OBIA (D).

Une méthodologie générale a été appliquée pour obtenir toutes ces zones polygonales. Une étape supplémentaire a été nécessaire pour la création des zones à la limite des classes, à savoir une expansion de 3 pixels appliquée à chaque classe.

Les 1500 points ont été sélectionnés aléatoirement comme suit :

- 100 points dans les zones A et 100 points dans les zones B, regroupés en une strate ;
- 100 points dans chacune des zones C, à savoir :
 - Zone « bâti-arbre » regroupant les pixels mixtes composés de bâti d'une part et de feuillus ou résineux d'autre part ;
 - Zone « bâti-sol » regroupant les pixels mixtes composés de bâti d'une part, de sol artificialisé, de sol nu, d'herbage ou de terres arables d'autre part ;
 - Zone « eau » regroupant les pixels mixtes composés d'eau et d'une autre classe ;
 - Zone « feuillus-résineux » regroupant les pixels mixtes de feuillus et de résineux ;

- Zone « herbage-culture » regroupant les pixels mixtes d’herbage et de terres arables ;
- Zone « herbage-sol artificialisé » regroupant les pixels mixtes d’herbage et de sol artificialisé ;
- Zone « sol nu-sol artificialisé » regroupant les pixels mixtes composés de sol nu d’une part et de sol artificialisé ou d’herbage d’autre part.
- Zone « sol-arbre » regroupant les pixels mixtes composés de feuillus ou de résineux d’une part, et d’une autre classe d’autre part, à l’exception du bâti ;

NB : ces zones ont été définies après analyse de la surface relative et de la surface totale de chaque groupe de pixels mixtes.

- 100 points dans chacune des zones D, à savoir dans les limites des classes bâti, forêt (feuillus et résineux regroupés dans cette même classe), sol nu, sol artificialisé et eau.

Dû au passage à la couverture 2018, l’intégralité des points a dû être soumise à nouveau à interprétation visuelle. Ce premier jeu de 1500 points s’est avéré insuffisant pour construire des modèles suffisamment robustes pour les méthodes DS et RF. Un second jeu de 1500 points a donc été produit. Celui-ci a été créé selon un échantillonnage aléatoire stratifié par classe, au sein des 4 zones tests. Ces 1500 points ont été photo-interprétés par les 3 équipes, mis en commun et plus de 100 points ont été interprétés consensuellement lors d’une réunion de visu à 5 personnes. La plupart des divergences étaient issues de l’interprétation des zones ouvertes, et la réunion a permis d’obtenir un consensus sur la typologie. Il est important de préciser qu’il existe une réelle difficulté à décrire l’occupation du sol sans ambiguïté à cette échelle.

2.2.2 MÉTHODE GÉNÉRALE DE FUSION PAR LE MODÈLE RANDOM FOREST

Les données d’entrée utilisées comme « features », sur lesquelles se base le modèle RF pour prédire une classe d’occupation du sol dans le résultat de la fusion, viennent des sources de données suivantes :

- La classification d’occupation par objet (OBIA) 2018 :
 - Probabilité d’appartenance des segments à chaque classe d’occupation du sol ;
 - Classe majoritaire ;
 - L’Equivalent Reference Probability (ERP), fournissant une indication du niveau de confiance de la classification (Bogaert & al. 2017).
- La classification d’occupation par pixel sur les orthophotos 2018 :
 - Proportion des différentes classes d’occupation du sol de la classification ‘pixel’ dans les segments issus de l’approche OBIA;
 - Classe majoritaire de la classification ‘pixel’ dans les segments issus de l’approche OBIA ;
 - L’ERP, calculé sur base des proportions de chaque classe d’occupation du sol ‘pixel’.
- Le modèle numérique de surface (MNH) :
 - Valeur moyenne par segment.
- La classification d’occupation par pixel sur les séries temporelles Sentinel 1/2
 - Classe modale par segment.
- La classification « Sen2Agri » avec le type de culture :
 - Classe modale par segment.
- Parcelles agricoles du SIGEC :
 - Classe modale par segment (information binaire).
- Masque forestier :
 - Classe modale par segment (information binaire).
- Stratification basée sur les classifications des orthos 2018 :
 - Classe modale par segment.

L'ensemble des traitements est réalisé dans le logiciel GRASS GIS et entièrement automatisé via une chaîne de traitement en Python (implémentée dans un Jupyter notebook). L'implémentation de la méthode est conçue pour maximiser au mieux les capacités de traitement en cluster (lancement de multiples processus en parallèle, réduisant significativement le temps de production). Les grandes lignes de l'implémentation sont reprises à la Figure 2 et ensuite expliquées en détail.

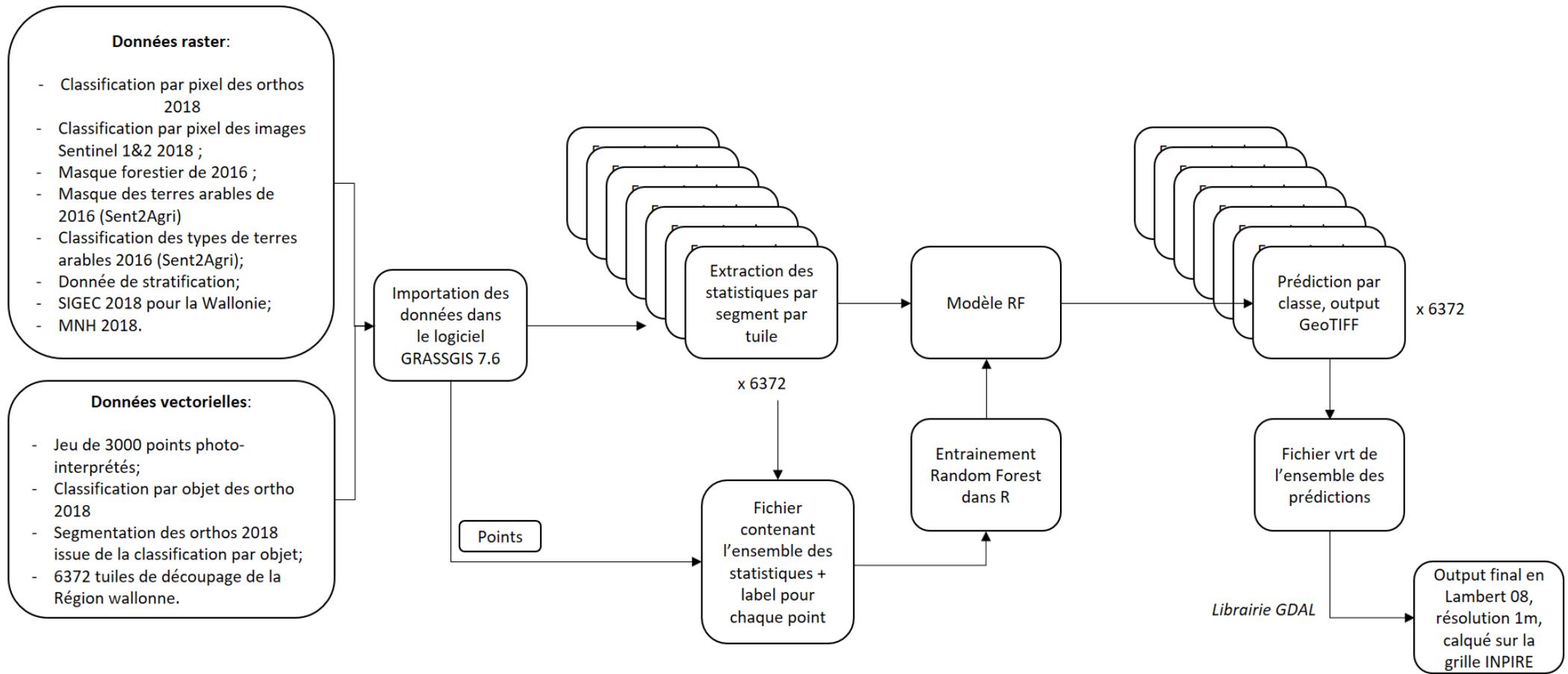


Figure 2 : Chaîne de traitement de la fusion via l'approche Random Forest

1. **Configuration des chemins des logiciels et des données sources** : l'ensemble des traitements a été réalisé dans GRASS GIS 7.6 fonctionnant avec une version de Python 2. Un fichier de configuration permet d'actualiser les chemins et variables d'environnement vers les différents logiciels en fonction de la configuration de la machine sur laquelle sera lancé le script. Ce fichier permet également de déterminer tous les chemins vers les données à utiliser par la suite.
2. **Importation de l'ensemble des sources de données d'entrée** dans GRASS GIS. Les données vectorielles (SIGEC) ont été préalablement rastérisées via la librairie GDAL. Les données raster sont toutes importées et puis tuilées en plusieurs centaines (voir milliers) de tuiles virtuellement mosaïquées sous une couche VRT (Virtual Raster – Standard GDAL). Ce tuilage permet d'optimiser les accès en lecture écriture sur le disque, dans le cadre d'une recherche de parallélisation importante des traitements. Cette étape d'importation des données n'est pas parallélisée actuellement, car elle n'est pas la plus longue, mais pourrait l'être, en fonction de la taille et de la quantité de données nécessaires.
3. **Extraction des statistiques des segments**, qui seront utilisées comme « features » pour l'entraînement du RF. Cette étape est parallélisée selon le tuilage des orthophotos (Fig.3, 6372 tuiles). L'extraction des statistiques est effectuée en parallèle pour les différentes tuiles. Il n'y a pas de limite théorique à cette parallélisation et son niveau ne dépend donc que des caractéristiques du cluster disponible dans l'infrastructure HPC (High Performance Computing system). Un fichier .csv contenant les statistiques des segments est créé pour chaque tuile d'ortho et stocké dans un dossier externe à GRASS.

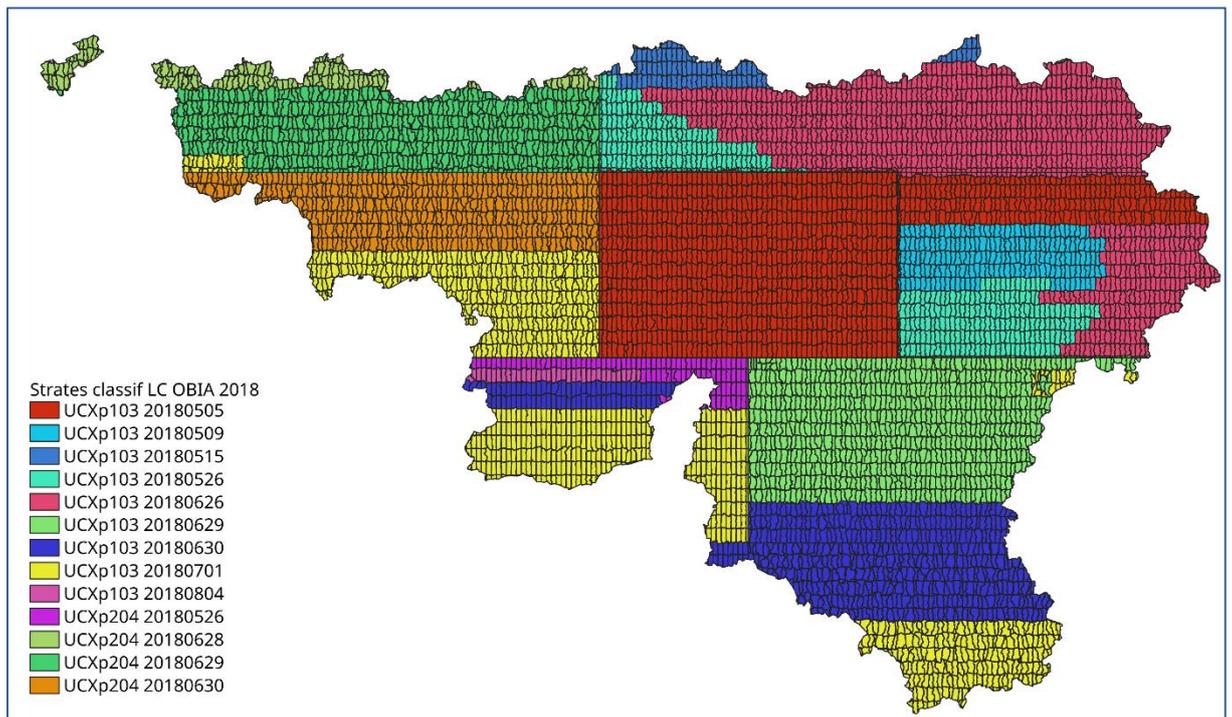


Figure 3: Regroupement des tuiles de la mosaïque des orthophotos par "strates"

4. **Entraînement du modèle Random Forest** : le fichier des 3000 points labélisés est importé dans GRASS GIS et un double identifiant unique (id segment + id tuile d'ortho) est attribué à chaque point en fonction de sa localisation. Les statistiques des segments situés sous les points d'entraînement sont extraites des fichiers .csv contenant ces statistiques des segments de chaque tuile d'ortho. Un fichier .csv contenant l'ensemble des statistiques et des labels associés est ainsi créé pour les 3000 points. Ce fichier est utilisé pour entraîner le modèle RF via le logiciel R (un script R est créé et lancé automatiquement via Python). Le modèle entraîné

est sauvegardé physiquement sur le disque (par défaut, R ne travaille qu'en mémoire RAM), afin de pouvoir être chargé ultérieurement dans l'étape de prédiction. L'étape d'entraînement du modèle n'est pas parallélisable.

5. **Prédiction des classes d'occupation du sol** : La prédiction est réalisée à nouveau au moyen du logiciel R, via Python. Le modèle sauvegardé à l'étape précédente est chargé dans R et la prédiction est réalisée pour chaque fichier .csv contenant les statistiques des segments des tuiles. La sortie de la classification fournit un nouveau fichier .csv qui contient la probabilité d'appartenance à chaque classe, l'indice ERP ainsi que la classe avec la plus forte probabilité d'appartenance. Une règle supplémentaire a été ajoutée afin de reclasser les segments feuillus ou résineux dont le MNH moyen était inférieur à 3m en arbustif feuillus ou arbustif résineux. Toutes ces informations sont ensuite utilisées pour reclasser les couches raster de segmentation dans GRASS GIS pour produire l'output final, à savoir un fichier .tif avec le résultat de la fusion pour chaque tuile qui correspond à la classe avec la plus forte probabilité d'appartenance pour chaque segment. Cette étape est parallélisable sur autant de nœuds/cœurs disponibles et il n'y a pas de limite théorique à la parallélisation.
6. **Un mosaïquage virtuel (VRT)** de l'ensemble des résultats produits par tuile d'ortho est ensuite réalisé.
7. **Transformation en Lambert 08 et calquage sur la grille INSPIRE** : afin d'utiliser le résultat pour la consolidation de la carte, ce dernier a subi une transformation de Lambert 72 à Lambert 08, un rééchantillonnage à 1m de résolution via un paramètre de mode et a ensuite été ajusté sur la grille INSPIRE à 1m fournie par le SPW, le tout via l'utilisation de la librairie gdal (`gdalwarp -srs 'EPSG:31370' -t_srs 'EPSG:3812' -tr 1 1 -r mode -ot Byte -co COMPRESS=LZW -co TILED=YES`). Il est important de garder le résultat en format « byte » afin de limiter la taille des fichiers.

L'algorithme RF n'est pas complètement une « black-box » car il permet de connaître le poids (l'importance) relatif des différentes variables dans le modèle. Ainsi, le graphique de l'importance des variables (Fig.4) montre bien que parmi les variables les plus importantes du modèle, on retrouve à la fois les classifications des orthophotos par pixel et par objet, mais également l'information de hauteur du MNH ou encore la classification Sentinel.

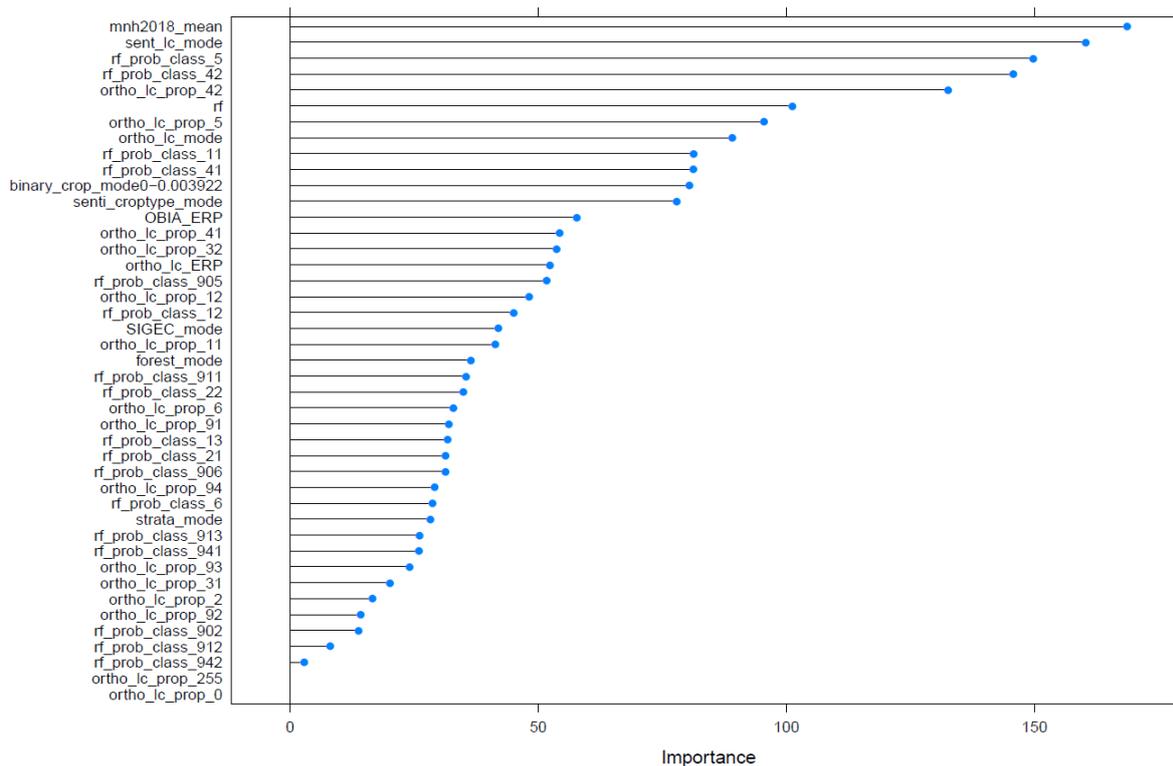


Figure 4 : Importance des variables utilisés dans le modèle Random Forest pour la fusion. Les 3 variables les plus importances sont, dans l'ordre : la hauteur du MNH, la classification LC de Sentinel, la classe herbage de l'OBIa

Le passage d'un logiciel à l'autre (GRASS GIS, R) au sein de la chaîne de traitement se fait automatiquement sans arrêt du processus.

Les 5 premières étapes de la chaîne de traitement développée sur les 4 zones tests par l'ULB ont été reprises en main par l'équipe de l'UCL qui s'est chargée de l'application sur l'ensemble de la Région wallonne. Ce transfert s'est opéré sans souci majeur mais a pris un peu de temps à cause notamment du temps nécessaire pour la compréhension et l'adaptation aux machines et logiciel. Cela représente un bon test au regard de la reproductibilité de la méthode et la migration vers une autre infrastructure informatique.

Le temps théorique nécessaire pour chaque étape réalisée à l'UCL après prise en main du script et des logiciels, sans bug, est approximativement le suivant :

- **Importation de l'ensemble des sources de données d'entrée** : 1 à 2 jours, sans compter le temps de préparation des données qui n'est pas négligeable (rasterisation, transformation en Lambert 72, etc.) ;
- **Extraction des statistiques des segments** : 4 jours sur 16 cœurs
- **Entraînement du modèle Random Forest** : moins d'une demi-heure
- **Prédiction des classes d'occupation du sol** : 1-2 jours sur 16 cœurs.

Le mosaïquage et l'utilisation de gdal pour la transformation ont pris moins de quelques heures. Le poids du fichier contenant le résultat final de fusion est d'environ 1Go.

Il est important de préciser que la fusion n'a pas été correctement réalisée pour une vingtaine de tuiles. Ces erreurs ont été corrigées lors de la phase de consolidation de la carte.

3 EVALUATION DE LA QUALITÉ DE LA FUSION

Les premiers résultats de validation de la fusion et des classifications ont été estimés sur base de la même méthodologie (et des mêmes points) que les premiers résultats d'évaluation de la qualité des données d'entrée. L'exactitude globale obtenue (87,7 %) présente une nette amélioration (un peu moins de 10% en valeur absolue) par rapport au résultat de classification avec toutes les classes et surpasse même la qualité de produits spécialisés (Tableau 4). Comme le montre le tableau 4, les résultats de la fusion sont meilleurs (moins de confusion) que les différentes données spécialisées utilisées en entrée. Par exemple, le taux de confusion de la fusion WALOUS en forêt est de près de 3% inférieur à celui du masque forestier et, en zones agricoles, de 0,2% meilleur que le SIGEC. Ceci démontre l'intérêt de valoriser la complémentarité des différentes données d'entrées ainsi que des deux approches de classification.

Tableau 4 : taux de confusions entre plusieurs classes des classifications de la fusion et de la fusion

	Confusion feuillus - résineux	Confusion sols nus - constructions artificialisées au sol	Confusion terres arables - herbage	Confusion constructions artificialisées au sol et hors sol
Ortho 2018 – pixels	10,4	22,1	26,7	5,5
Ortho 2018 – OBIA	10	18,8	/	8,3
Sentinel	8,5	12,4	27,5	/
Masque forestier	10	/	/	/
SIGEC	/	/	8,4	/
FUSION	7,1	12,1	8,2	7,5

Au niveau des deux classifications des orthophotos 2018, il existe des différences minimales entre elles qui ne sont pas systématiques pour les deux jeux de données. La fusion permet de corriger ces erreurs. On notera par ailleurs que la classification des données Sentinel a une faible exactitude globale (mesurée sur base des points interprétés à 1m), mais qu'elle présente un faible taux de confusion entre les feuillus et les résineux d'une part, et entre les sols nus et les constructions artificialisées au sol d'autre part.

Le tableau 5 reprend les exactitudes globales focalisées sur les zones agricoles et les zones boisées afin de comparer les résultats de la fusion avec les données de référence (SIGEC) ou auxiliaires (masque forestier). Etant donné que ces deux couches thématiques ne sont pas complètes, cette comparaison s'effectue soit sur l'étendue de la fusion WALOUS (en ajoutant une classe « autres »), soit sur l'étendue des masques respectifs :

- Les types d'arbres (feuillus/résineux), avec une classe « autres » reprenant les non ligneux, en prenant en compte tous les points de validation ;
- Les types d'arbres (feuillus/résineux), avec une classe « autres » reprenant les non ligneux, en ne prenant en compte que les points de validation inclus dans le masque forestier ;
- La végétation basse (terres arables et herbages), avec une classe « autres » pour toutes les autres classes, en prenant en compte tous les points de validation ;
- La végétation basse (terres arables et herbages), avec une classe « autres » pour toutes les autres classes, en ne prenant en compte que les points situés dans les polygones du SIGEC ;

Tableau 5 : comparaison des exactitudes globales selon les points utilisés

	Types d'arbres Tous les points	-Types d'arbres Points du masque forestier	-Végétation basse Tous les points	-Végétation basse Points du SIGEC
Ortho 2018 – pixels	93,3	89,5	78,6	72,5
Ortho 2018 – OBIA	92,6	84,6	/	/
Sentinel	88,6	72,9	57,3	74,1
Masque forestier	91	86	/	/
SIGEC	/	/	84,9	90,1
FUSION	95	89,5	91,4	90

Le tableau 5 confirme le fait que la fusion des données contribue à améliorer les résultats de la classification. Par ailleurs, on remarque de grosses différences suivant que la matrice de confusion soit construite avec la totalité ou seulement une partie des points.

Dans le cas du masque forêt, cette différence est principalement due à une différence de définition. En effet, alors que le masque forêt se focalise sur les parcelles forestières, les données WALOUS cartographient tous les arbres. Il est donc normal que les arbres isolés, les haies, les petits bosquets et les arbres des zones urbaines ne soient pas présents dans le masque forestier. L'usage de ce dernier n'est donc pertinent que là où il a été conçu.

Le produit fusionné présente une exactitude globale supérieure au SIGEC pour tous les points, et une exactitude légèrement inférieure quand on se base uniquement sur les points à l'intérieur du SIGEC. Il faut toutefois noter que les définitions de classes du SIGEC sont également différentes des définitions de classes dans les données WALOUS. Ceci explique un certain nombre de divergences entre les jeux de données. D'une part, le SIGEC ne reprend que les herbages agricoles et omet ainsi tous les autres herbages (bords de routes, jardins, terrains de sports...) ce qui réduit fortement son exactitude globale. Il faut toutefois noter que les omissions de zones cultivées y sont presque nulles (moins de 0.5%), ce qui sera utile pour la consolidation. Par ailleurs, les parcelles incluent parfois des arbres isolés. Ces arbres isolés constituent la principale source d'erreur de contamination dans les herbages du SIGEC (3.3 %). Enfin, certains points devront être vérifiés sur le terrain car l'obtention d'une base de référence parfaite des herbages ne peut se faire uniquement sur base d'une photo-interprétation.

4 ANNEXES

Annexe 1 – Détails techniques de l’approche de fusion par règles de classification.

1. MÉTHODOLOGIE GÉNÉRALE

La fusion par règles de classification repose sur l’analyse experte par interprétation visuelle des données. Interactivement, l’expert définit des règles permettant d’associer à chaque pixel de la scène une classe d’occupation du sol. La littérature renseigne de nombreux exemples d’approches de classification de l’occupation du sol par règles de classification (Beaumont et al., 2017 ; Belgiu et al., 2014 ; Voltersen et al., 2014 ; O’Neil-Dunne et al., 2013). Dans sa section discussion, Beaumont et al. (2017) détaille les avantages/inconvénients de cette approche vis-à-vis des méthodes de cartographie par algorithme de classification (i.e. machine learning). L’approche par règles expertes permet d’améliorer la qualité de classification dans des contextes spécifiques à la scène. La définition de règles est un processus fastidieux et peu automatisable ; l’ordre de chaque règle - la définition des seuils de classification impactent le résultat en sortie.

La variante de fusion par règles de classification développée dans WALOUS comprend 23 règles conditionnelles regroupées en 9 volets. Ces règles s’appliquent au niveau des pixels (25x25 cm). Au départ d’opérateurs logiques et conditionnels (e.g. si ... > ... et ... < ... alors ... ; si non ...), l’occupation du sol de chaque pixel est ainsi définie en tenant compte de l’accord entre les classifications OBIA et pixel, de la présence-absence dans des bases de données ancillaires et de règles contextuelles.

Les règles développées sont très génériques, ce qui en facilite l’implémentation dans différents outils. Par exemple :

- ArcGIS 10.6 ou ArcGIS Pro > fonctions :
 - o Con : <https://pro.arcgis.com/fr/pro-app/tool-reference/spatial-analyst/con-.htm>
 - o Expand/Développer : <https://pro.arcgis.com/fr/pro-app/tool-reference/spatial-analyst/expand.htm>
 - o Shrink/Réduire : <https://pro.arcgis.com/fr/pro-app/tool-reference/spatial-analyst/shrink.htm>
- GRASS GIS:
 - o r.map.calc: <https://grass.osgeo.org/grass77/manuals/r.mapcalc.html>
 - o r.grow.shrink : <https://grass.osgeo.org/grass76/manuals/addons/r.grow.shrink.html>
- Orfeo ToolBox (OTB): band math <https://www.orfeo-toolbox.org/packages/archives/Doc/CookBook-4.0.0.pdf>
- eCognition Developer 9: <http://www.ecognition.com/>

Pour rappel, la légende fusionnée est reprise à la Figure 1.

Références:

- Beaumont B., T. Grippa, M. Lennert, N. Stephenne, S. Zarougui, J.-C. Jasselette and E. Wolff, 2019. Vers une méthodologie opérationnelle de cartographie à très haute résolution spatiale de l’occupation et de l’utilisation du sol en Wallonie. *Dynamiques Régionales* N°8, p15-33.
- Beaumont, B., T. Grippa, M. Lennert, S. Vanhuysse, N. Stephenne and E. Wolff, 2017. Toward an operational framework for fine-scale urban land-cover mapping in Wallonia using submeter remote sensing and ancillary vector data. *J. Appl. Remote Sens.* 11(3), 036011 (2017), doi: 10.1117/1.JRS.11.036011.
- Belgiu, M., L. Draguț, and J. Strobl. 2014. “Quantitative Evaluation of Variations in Rule-Based Classifications of Land Cover in Urban Neighbourhoods Using WorldView-2 Imagery.” *ISPRS Journal of Photogrammetry and Remote Sensing* 87: 205–15. doi:10.1016/j.isprsjprs.2013.11.007.

O'Neil-Dunne, J. P.M., S. W. MacFaden, A. R. Royar, and K. C. Pelletier. 2013. "An Object-Based System for LiDAR Data Fusion and Feature Extraction." *Geocarto International* 28 (3): 227–42. doi:10.1080/10106049.2012.689015.

Voltersen, M., Berger, C., Hese, S., & Schmullius, C. (2014). Object-based land cover mapping and comprehensive feature calculation for an automated derivation of urban structure types at block level. *Remote Sensing of Environment*, 154, 192–201. <https://doi.org/10.1016/j.rse.2014.08.024>

2. LISTE DES DONNÉES UTILISÉES

Neuf données sont utilisées pour la fusion par règles :

- Orthophotos 2018 pour l'interprétation visuelle menant à la définition des règles ;
- Classification OBIA des orthophotos/MNH 2018 ;
- Classification pixel des orthophotos/MNH 2018 ;
- Modèle Numérique de Hauteur 2018 ;
- Parcellaire agricole anonyme (divisé en deux sous-catégories : terres arables et prairies) 2017 ;
- Squelette vectoriel WALOUS (surfaces d'eau [IGN], réseau routier [PICC], bâtiment [PICC]) ;
- Classification pixel de la série temporelle Sentinel 2018 ;
- Masque forestier 2016 ;
- Classification des espaces forestiers en trois classes 2016 (feuillus, conifères et coupes).

3. RÈGLES DE CLASSIFICATION DÉTAILLÉES

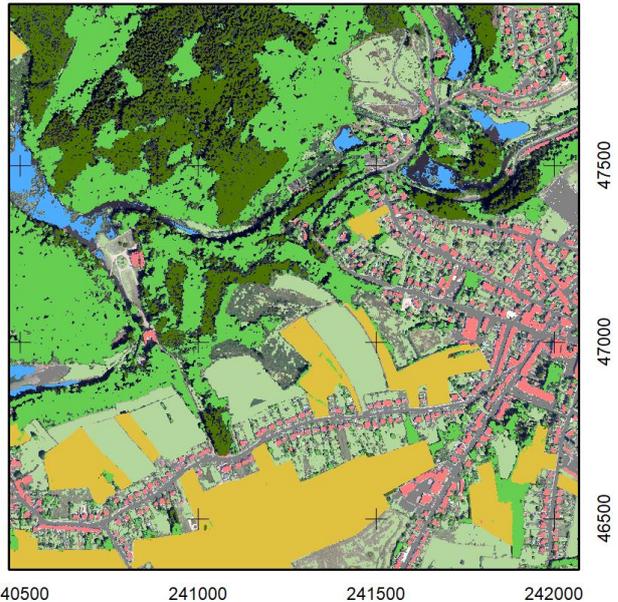
Une vision détaillée des règles est proposée dans le Tableau 7. Un zoom est proposé pour une vision plus fine de l'impact des règles.

Tableau 7 - Détail des règles de classification.

Règles	Illustration (Coordonnées 49.727976, 5.634354)
Fond de plan Orthophotos 2018 (©SPW)	
1. Règle conditionnelle permettant d'isoler les pixels (25x25 cm) communément classés par les deux approches OBIA et PIXEL.	
Les pixels non-isolés se voient attribuer une valeur temporaire de « 99 ».	
<pre>outCon1 = Con((OBIA==PIXEL), PIXEL, 99)</pre>	

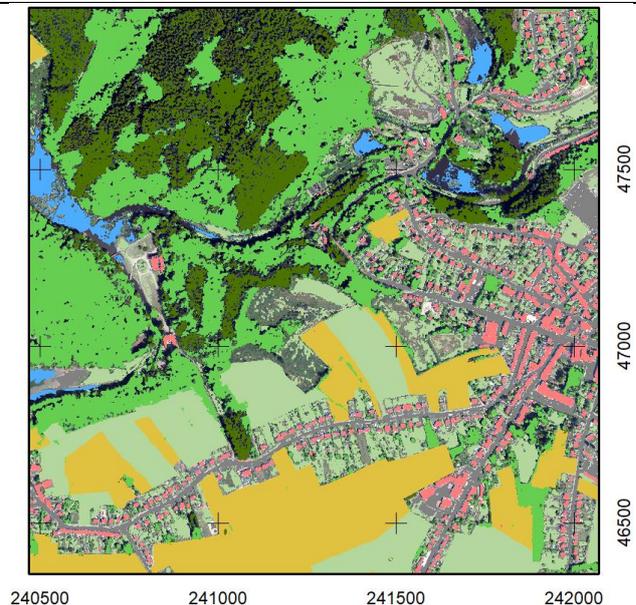
2. Reclassification en « 3 - Terres arables » des pixels classés comme « 99 », « 11 », « 5 » OU « 2 » ET croisant les parcelles de terres arables au sein du parcellaire agricole anonyme (SIGEC). Les éléments ligneux ne sont pas forcés par cette règle afin de permettre la gestion ultérieure du double label sur ces éléments en hauteur.

```
outCon2 =
Con ((TerresArableSIGEC==1) & ((outCon1==99) | (o
uCon1==11) | (outCon1==5) | (outCon1==2)), 3, outC
on1)
```



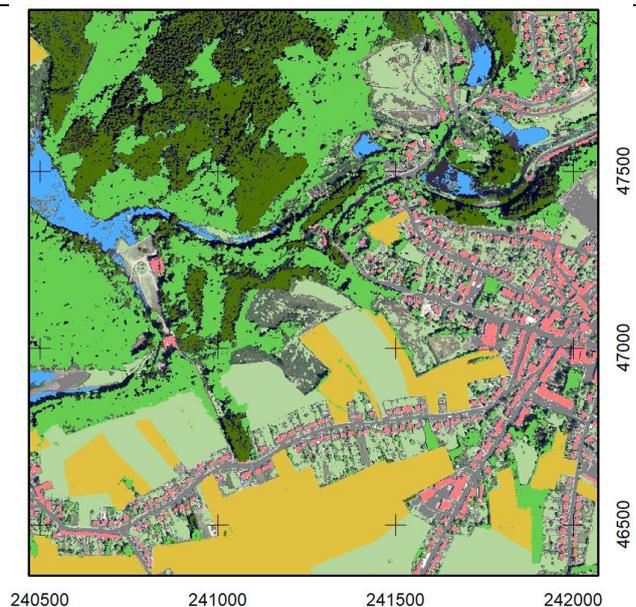
3. Reclassification en « 5 - Végétation herbacée » des pixels classés comme « 99 », « 11 », « 5 » OU « 2 » ET croisant les parcelles de prairies au sein du parcellaire agricole anonyme (SIGEC). Les éléments ligneux ne sont pas forcés par cette reclassification afin de permettre la gestion ultérieure du double label sur ces éléments en hauteur.

```
outCon3 =
Con ((PrairiesSIGEC==1) & ((outCon2==99) | (outCo
n2==11) | (outCon2==5) | (outCon2==2)), 5, outCon1
)
```



4. Reclassification des pixels classés « 99 » en « 6 – Eau » si ces pixels sont classés « Eau » par la classification OBIA OU par la classification PIXEL ET qu'ils sont repris dans le squelette vectoriel WALOUS comme surfaces d'eau.

```
outCon4 =
Con ((outCon3==99) & (SqueletteEAU==1) & (PIXEL==
6), 6, Con ((outCon3==99) & (SqueletteEAU==1) & (OB
IA==6), 6, outCon3))
```

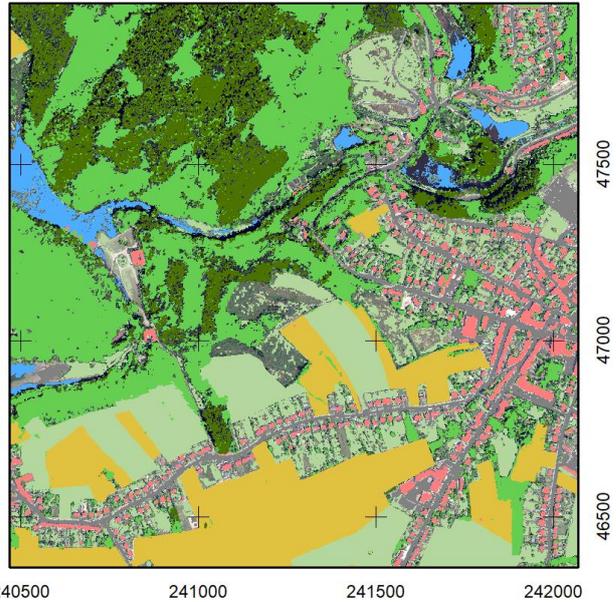


5. Cinq règles conditionnelles permettent de reclasser les pixels classés « 99 » pour lesquels la classification OBIA OU la classification PIXEL attribue un label « Ombre » (OBIA = 902-905-911-912-913-941 ; PIXEL = 91-92-93-94). La reclassification profite alors de l'information contenue dans OBIA OU PIXEL pour attribuer un nouveau label à ces pixels.

```

outCon5 =
Con ( (outCon4==99) & (PIXEL==91) , Con ( (OBIA==911
) | (OBIA==913) , 11, outCon4) , outCon4)
outCon51 =
Con ( (outCon5==99) & (PIXEL==92) & (OBIA==902) , 2,
outCon5)
outCon52 =
Con ( (outCon51==99) & (PIXEL==93) & (OBIA==905) , 5
, outCon51)
outCon53 =
Con ( (outCon52==99) & (PIXEL==94) , Con ( (OBIA==94
1) , 41, 42) , outCon52)
outCon54 =
Con ( (outCon53==99) & (OBIA==912) , 12, outCon53)

```



6. L'intégration de règles conditionnelles sur le MNH (< ou >3m), le masque forestier (FORET) et la classification en trois classes (FORET3CL) permettent de classer et de corriger la classification des « Ligneux ». La fonction « Expand » est utilisée afin de combler les trous au sein des espaces forestiers lorsque les classifications OBIA et PIXEL ont un résultat de sortie contradictoire (« 41 » vs « 42 »), pixels classés temporairement en « 88 ». Cette fonction permet d'étendre d'un nombre défini de pixels (ici 50) les zones « 41 » et « 42 » sur les autres classes dans leur voisinage. Une règle conditionnelle reclasse finalement les pixels « 88 » selon leur classification dans le raster résultant de la fonction « Expand ».

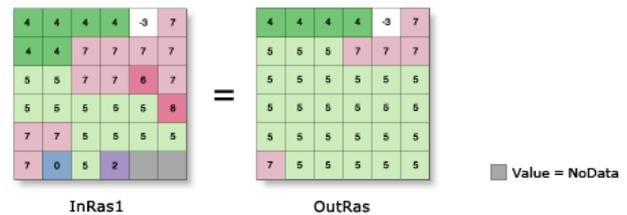
```

outCon6 = Con ( (outCon54==99) & (MNH >=
3) & (FORET==1) , Con ( (PIXEL==41) & (FORET3CL==3) ,
41, Con ( (PIXEL==42) & (FORET3CL==1) , 42, outCon54
) ) , outCon54)
outCon7 = Con ( (outCon6==99) & (MNH >=
3) & (FORET==1) , Con ( (OBIA==41) & (FORET3CL==3) , 4
1, Con ( (OBIA==42) & (FORET3CL==1) , 42, outCon6) ) ,
outCon6)
outCon8 =
Con ( (outCon7==99) & (PIXEL==41) & (OBIA==42) & (FO
RET3CL==1) , 42, Con ( (outCon7==99) & (PIXEL==41) &
(OBIA==42) & (FORET3CL==3) , 41,
Con ( (outCon7==99) & (PIXEL==42) & (OBIA==41) & (FO
RET3CL==1) , 42, Con ( (outCon7==99) & (PIXEL==42) &
(OBIA==41) & (FORET3CL==3) , 41, outCon7) ) ) )
outCon9 =
Con ( (outCon8==99) & (PIXEL==41) & (OBIA==42) , 88,
Con ( (outCon8==99) & (PIXEL==42) & (OBIA==41) , 88,
outCon8) )
Expand1 = Expand(outCon9, 50, [41, 42])
outCon10 =
Con ( (outCon9==88) , Expand1, outCon9)

```

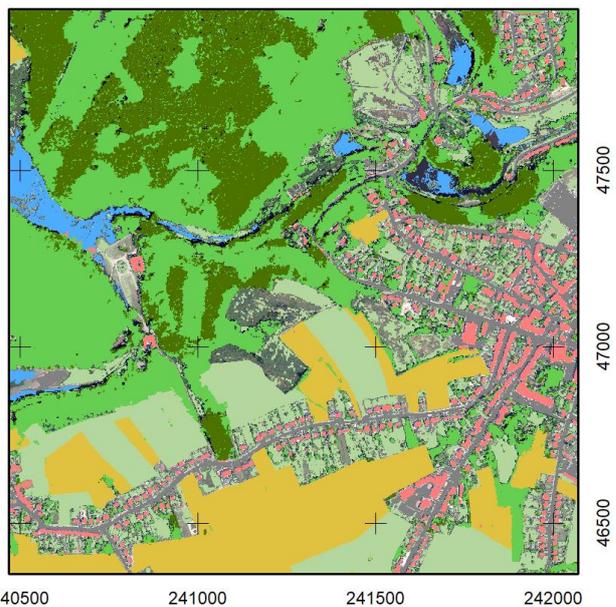
La fonction « Expand/Développer » telle qu'expliquée dans l'aide ArcGIS Pro :

Illustration



`OutRas = Expand(InRas1, 1, [5])`

<https://pro.arcgis.com/fr/pro-app/tool-reference/spatial-analyst/expand.htm>

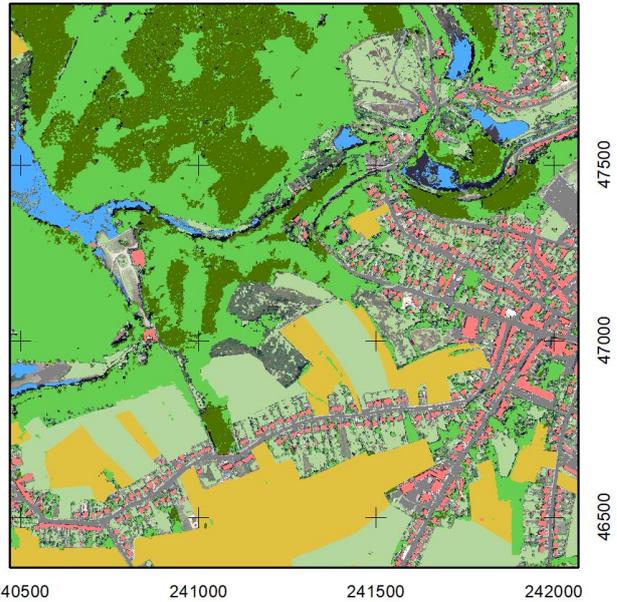


7. Le squelette vectoriel bâtiment (provenant du PICC) est utilisé pour forcer la représentation des bâtiments dans le résultat final d'occupation du sol (classe « 12 »). Les fonctions « Expand » et « Shrink » (fonction inverse) sont utilisées afin de conserver les bâtiments et annexes non repris dans le PICC qui sont caractérisés par une unité minimale de cartographie de 6,25m² (bâtiment carré). Les éléments linéaires le long de bâtiments résultant du décalage entre le PICC et les orthophotos sont éliminés par ces règles.

```

outCon11 =
Con ( (BATIPICC==0) & (outCon10==12) , 77, Con ( (BAT
IPICC==1) & (outCon10 != 41) & (outCon10 !=
42) , 12, outCon10) )
outShrink = Shrink(outCon11, 5, [77])
outExpand = Expand(outShrink, 5, [77])
outCon111 = Con ((outExpand==77) , 12, outCon11)

```

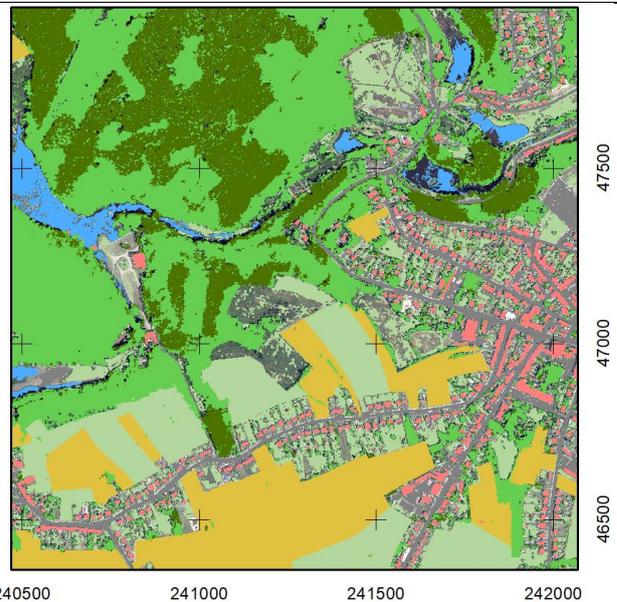


8. Règle conditionnelle de reclassification des routes par intégration du squelette vectoriel rasterisé à l'exception des classes « 41 » et « 42 » pour la gestion ultérieure du double label.

```

outCon12 = Con ((ROUTE==1) & (outCon111 !=
41) & (outCon111 != 42) , 11, outCon111)

```



9. Reclassification des pixels toujours classés « 99 » :

- selon des règles de proximité basées sur l'utilisation d' « Expand »,
- selon un seuil de $>$ ou $<$ 3 m défini sur le MNH,
- selon l'occupation du sol provenant de la classification de la série temporelle d'images Sentinel-2.

```
outCon13 = Con((outCon12==99) & (OBIA ==  
5) & ((PIXEL == 31) | (PIXEL == 32)), 5, outCon12)  
Expand2 = Expand(outCon13, 50, [2, 5, 6, 11])  
Expand3 = Expand(outCon13, 50, [12, 41, 42])  
outCon14 = Con((outCon13==99) & (MNH >=  
3) & ((Expand3==41) | (Expand3==42)), Expand3, out  
Con13)  
outCon15 =  
Con((outCon14==99) & (MNH < 3), Expand2, outCon14)  
outCon16 =  
Con(((outCon15==99) | (outCon15==88) | (outCon15  
==77)), Expand2, Con(IsNull(outCon15), Expand2,  
outCon15))  
outCon17 =  
Con(((outCon16==99) | (outCon16==88) | (outCon16  
==77)), Con((PIXEL != 9), PIXEL, LCSSENTINEL), outC  
on16)  
outCon18 =  
Con((outCon17==99), 5, Con((outCon17==91),  
11, Con((outCon17==92), 2, Con((outCon17==93), 5  
, Con((outCon17==31) | (outCon17==32)), 3, outCo  
n17))))))
```

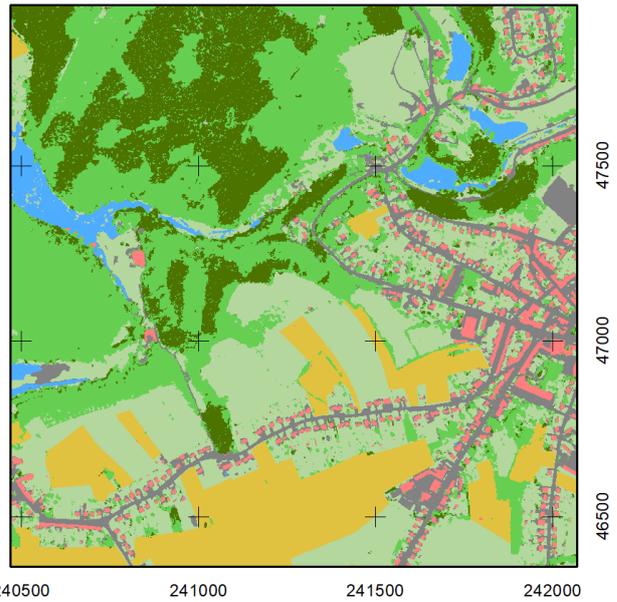
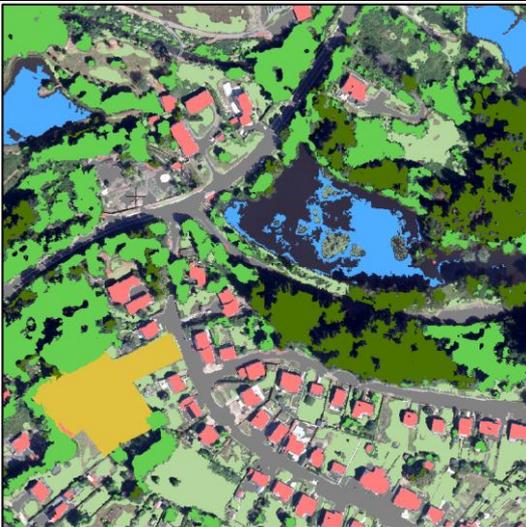


Tableau 3 - Zoom illustrant les 9 étapes de classification (Coordonnées 49.731908, 5.639478).



1

Pixels communément classés



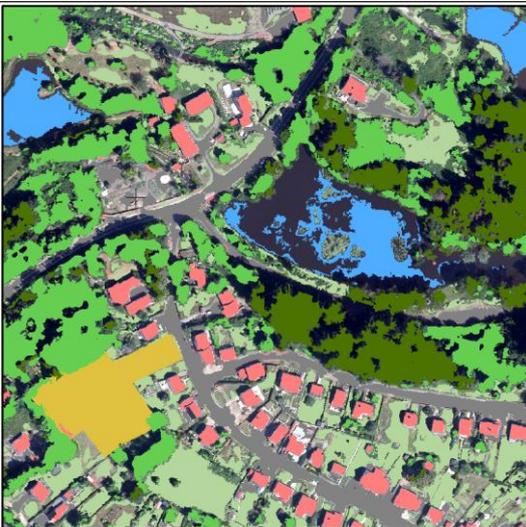
2

Forçage d'une terre arable depuis le SIGEC



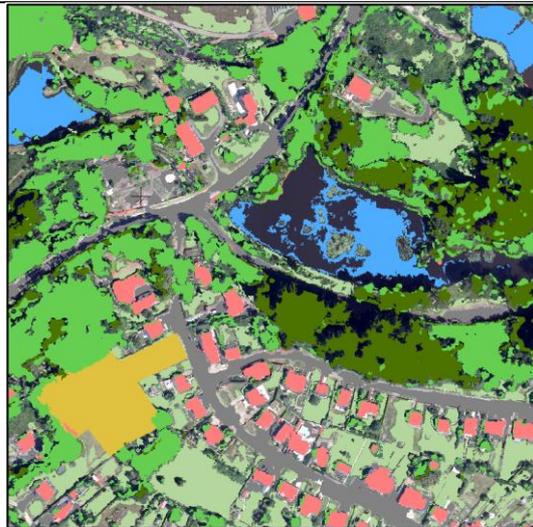
3

> pas de prairie dans cette zone



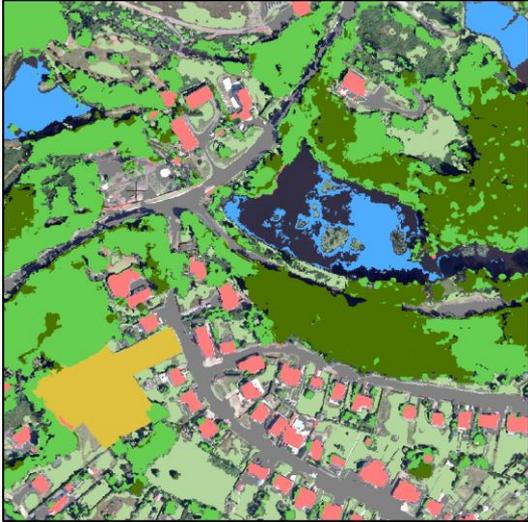
4

> pas de surfaces d'eau dans le squelette dans cette zone



5

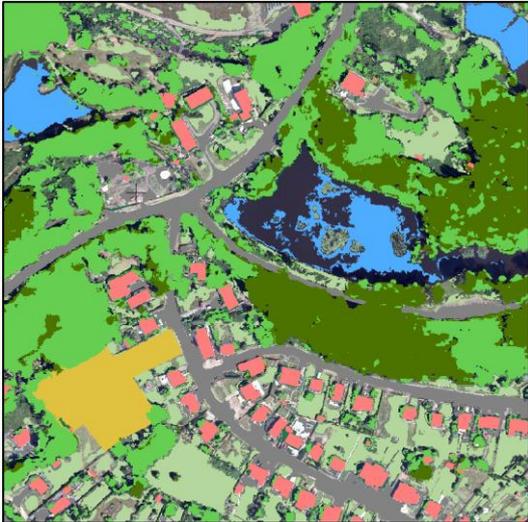
Labellisation des ombres.



6
Labellisation des espaces ligneux



7
Forçage PICC et détection de changement des bâtiments



8
Forçage squelette réseau routier



9
Classification occupation du sol fusionnée

4. SCRIPT PYTHON-ARCGIS

Le script est implémenté dans l'interface PyCharm et fait appel à des fonctions ArcGIS PRO en langage Python (ArcPY scripting).

```
### Import system modules
import arcpy
from arcpy import env
from arcpy.sa import *

# Set environment settings
env.workspace = "H:/Fusion"

# Check out the ArcGIS Spatial Analyst extension license
arcpy.CheckOutExtension("Spatial")

# Overwrite pre-existing files
arcpy.env.overwriteOutput = True

### A. Data Inputs

OBIA = Raster("Classif_OBIA2018/V2/Clip/OBIA_Z2.tif")
PIXEL = Raster("RasterZ2/PixCl1_25.tif")
SIGECTA = Raster("RasterZ2/10/SIGECTA10.tif")
SIGECP = Raster("RasterZ2/10/SIGECP10.tif")
LCSEN = Raster("RasterZ2/WALOUS_ML_recoded.tif")
BATI = Raster("RasterZ2/10/Bati10.tif")
ROUTE = Raster("RasterZ2/10/Route10.tif")
FORET = Raster("RasterZ2/10/Foret10.tif")
FORET3CL = Raster("RasterZ2/10/3cl_201610.tif")
EAU = Raster("RasterZ2/10/EAU10.tif")
MNH = Raster("RasterZ2/MNH.tif")
print("input data loaded")

### B. Reclassification rules

## 1. Identify commonly classified pixels in OBIA_25 and Pixel_25
(reclassified from 50)
outCon1 = Con((OBIA==PIXEL),PIXEL,99)
outCon1.save("Results/Z2_v1/1_Common.tif")
outCon1 = Raster("Results/Z2_v1/1_Common.tif")
print("End of step 1")

## 2. Reclass arable lands (Land Cover definition)
outCon2 =
Con((SIGECTA==1) & ((outCon1==99) | (outCon1==11) | (outCon1==5) | (outCon1==2)),
3,outCon1)
outCon2.save("Results/Z2_v1/2_Agri.tif")
outCon2 = Raster("Results/Z2_v1/2_Agri.tif")
print("End of step 2")

## 3. Reclass arable lands (Land Cover definition)
outCon3 =
Con((SIGECP==1) & ((outCon2==99) | (outCon2==11) | (outCon2==3) | (outCon2==2)),5
,outCon2)
outCon3.save("Results/Z2_v1/3_Grasslands.tif")
outCon3 = Raster("Results/Z2_v1/3_Grasslands.tif")
print("End of step 3")

## 4. Reclass water surfaces
outCon4 =
Con((outCon3==99) & (EAU==1) & (PIXEL==6),6,Con((outCon3==99) & (EAU==1) & (OBIA=
```

```

=6), 6, outCon3)
outCon4.save("Results/Z2_v1/4_Water.tif")
outCon4 = Raster("Results/Z2_v1/4_Water.tif")
print("End of step 4")

## 5. Reclass shadow
outCon5 =
Con ( (outCon4==99) & (PIXEL==91) , Con ( (OBIA==911) | (OBIA==913) , 11, outCon4) , out
Con4)
outCon51 = Con ( (outCon5==99) & (PIXEL==92) & (OBIA==902) , 2, outCon5)
outCon52 = Con ( (outCon51==99) & (PIXEL==93) & (OBIA==905) , 5, outCon51)
outCon53 =
Con ( (outCon52==99) & (PIXEL==94) , Con ( (OBIA==941) , 41, 42) , outCon52)
outCon54 = Con ( (outCon53==99) & (OBIA==912) , 12, outCon53)
outCon54.save("Results/Z2_v1/5_Shadow.tif")
outCon54 = Raster("Results/Z2_v1/5_Shadow.tif")
print("End of step 5")

## 6. Reclass forest lands
outCon6 = Con ( (outCon54==99) & (MNH >=
3) & (FORET==1) , Con ( (PIXEL==41) & (FORET3CL==3) , 41, Con ( (PIXEL==42) & (FORET3CL=
=1) , 42, outCon54) ) , outCon54)
outCon6.save("Results/Z2_v1/outCon6.tif")
outCon7 = Con ( (outCon6==99) & (MNH >=
3) & (FORET==1) , Con ( (OBIA==41) & (FORET3CL==3) , 41, Con ( (OBIA==42) & (FORET3CL==1
) , 42, outCon6) ) , outCon6)
outCon7.save("Results/Z2_v1/outCon7.tif")
outCon8 =
Con ( (outCon7==99) & (PIXEL==41) & (OBIA==42) & (FORET3CL==1) , 42, Con ( (outCon7==9
9) & (PIXEL==41) & (OBIA==42) & (FORET3CL==3) , 41,
Con ( (outCon7==99) & (PIXEL==42) & (OBIA==41) & (FORET3CL==1) , 42, Con ( (outCon7==9
9) & (PIXEL==42) & (OBIA==41) & (FORET3CL==3) , 41, outCon7) ) ) )
outCon8.save("Results/Z2_v1/outCon8.tif")
outCon9 =
Con ( (outCon8==99) & (PIXEL==41) & (OBIA==42) , 88, Con ( (outCon8==99) & (PIXEL==42)
& (OBIA==41) , 88, outCon8) )
outCon9.save("Results/Z2_v1/outCon9.tif")
outCon9 = Raster("Results/Z2_v1/outCon9.tif")
Expand1 = Expand(outCon9, 50, [41, 42])
Expand1.save("Results/Z2_v1/Expand1.tif")
outCon9 = Raster("Results/Z2_v1/outCon9.tif")
Expand1 = Raster("Results/Z2_v1/Expand1.tif")
outCon10 = Con ( (outCon9==88) , Expand1, outCon9)
outCon10.save("Results/Z2_v1/6_Foret.tif")
outCon10 = Raster("Results/Z2_v1/6_Foret.tif")
print("End of step 6")

## 7. Reclass building
outCon11 = Con ( (BATI==0) & (outCon10==12) , 77, Con ( (BATI==1) & (outCon10 !=
41) & (outCon10 != 42) , 12, outCon10) )
outCon11.save("Results/Z2_v1/7_Bati.tif")
outCon11 = Raster("Results/Z2_v1/7_Bati.tif")
outShrink = Shrink(outCon11, 5, [77])
outShrink.save("Results/Z2_v1/7_outShrink.tif")
outShrink = Raster("Results/Z2_v1/7_outShrink.tif")
outExpand = Expand(outShrink, 5, [77])
outExpand.save("Results/Z2_v1/7_outExpand.tif")
outExpand = Raster("Results/Z2_v1/7_outExpand.tif")
outCon111 = Con ( (outExpand==77) , 12, outCon11)
outCon111.save("Results/Z2_v1/outCon111.tif")
outCon111 = Raster("Results/Z2_v1/outCon111.tif")

```

```

print("End of step 7")

## 8. Reclass Roads
outCon12 = Con((ROUTE==1)&(outCon111 != 41)&(outCon111 !=
42),11,outCon111)
outCon12.save("Results/Z2_v1/8_Route.tif")
outCon12 = Raster("Results/Z2_v1/8_Route.tif")
print("End of step 8")

## 9. Final reclass
outCon13 = Con((outCon12==99)&(OBIA == 5)&((PIXEL == 31)|(PIXEL ==
32)),5,outCon12)
outCon13.save("Results/Z2_v1/outCon13.tif")
outCon13 = Raster("Results/Z2_v1/outCon13.tif")
Expand2 = Expand(outCon13,50,[2, 5, 6, 11])
Expand2.save("Results/Z2_v1/Expand2.tif")
Expand2 = Raster("Results/Z2_v1/Expand2.tif")
Expand3 = Expand(outCon13,50,[12, 41, 42])
Expand3.save("Results/Z2_v1/Expand3.tif")
Expand3 = Raster("Results/Z2_v1/Expand3.tif")
outCon14 = Con((outCon13==99)&(MNH >=
3)&((Expand3==41)|(Expand3==42)),Expand3,outCon13)
outCon14.save("Results/Z2_v1/outCon14.tif")
outCon15 = Con((outCon14==99)&(MNH<3),Expand2,outCon14)
outCon15.save("Results/Z2_v1/outCon15.tif")
outCon15 = Raster("Results/Z2_v1/outCon15.tif")
outCon16 =
Con(((outCon15==99)|(outCon15==88)|(outCon15==77)),Expand2,Con(IsNull(out
Con15),Expand2,outCon15))
outCon16.save("Results/Z2_v1/outCon16.tif")
outCon16 = Raster("Results/Z2_v1/outCon16.tif")
outCon17 =
Con(((outCon16==99)|(outCon16==88)|(outCon16==77)),Con((PIXEL!=9),PIXEL,L
CSEN),outCon16)
outCon17.save("Results/Z2_v1/outCon17.tif")
outCon17 = Raster("Results/Z2_v1/Rule_Based_Z2_V1.tif")
outCon18 = Con((outCon17==99),5,Con((outCon17==91),
11,Con((outCon17==92),2,Con((outCon17==93),5,Con(((outCon17==31)|(outCon1
7==32)),3,outCon17))))))
outCon18.save("Results/Z2_v1/Rule_Based_Z2_V2.tif")
print("End of process")

```

Annexe 2 – Détails techniques de l’approche de fusion par la méthode Dempster Shafer.

Un des essais de fusion a été réalisé en suivant la théorie de Dempster-Shafer, basée sur la notion de preuves, à savoir que, pour chaque pixel, la classe dont la fonction de confiance est maximale sera sélectionnée. Cette fonction de confiance indique la confiance que chaque classification en entrée de la fusion présente pour chaque valeur de classe. Afin de déterminer cette fonction de confiance, une matrice de confusion est calculée pour chaque classification en entrée de la fusion via l’outil ComputeConfusionMatrix de OTB en utilisant les points photo-interprétés comme vérités de terrain. Des matrices ont été produites en tenant compte de la probabilité d’échantillonnage des points, soit 3 matrices par classification (= zones où il y a un conflit entre plusieurs bases de données (classification pixels et OBIA), zones à proximité des bordures d’une classe et zones où les bases de données sont en accord).

Un prétraitement des classifications dont la résolution spatiale dépasse 1m (SENTINEL et masque forestier) a été opéré avant d’utiliser ces données dans la fusion. Une segmentation des orthos a tout d’abord été implémentée dans E-Cognition et par la suite, les données des classifications SENTINEL et du masque forestier de Gembloux ont été jointes aux segments suivant la classe majoritaire (join attributes by field value – QGIS).

Les fusions ont été opérées par strates, suivant l’échantillonnage stratifié sur base des zones d’incertitude utilisées pour la création des 1500 premiers points. Les 3 fusions ont ensuite été assemblées en un unique résultat en suivant l’emprise de chaque strate.

1. LIGNES DE COMMANDES

Pour les étapes utilisant des fonctions OTB, les paramètres d’écriture de fichier s’inscrivent après l’extension (ils ne sont pas répétés dans nos lignes de commande pour plus de clarté) :

« fichier.tif ?&gdal :co :COMPRESS=LZW&gdal :co :TILED=YES&gdal :co :BIGTIFF=YES »)

1. Compute the confusion matrix for each classification input

```
otbcli_ComputeConfusionMatrix -in input_classif1.tif -out corresponding_confusion_matrix1.csv -  
ref vector -ref.vector.in points_with_class_label_stratified.shp -  
ref.vector.field field_with_class_label -ref.vector.nodata 255
```

2. Fusion of classifications by the Dempster-Shafer method

```
otbcli_FusionOfClassifications -il input_classif1.tif input_classif2.tif input_classifX.tif -  
method dempstershafer -method.dempstershafer.cmfl corresponding_confusion_matrix1.csv  
corresponding_confusion_matrix2.csv corresponding_confusion_matrixX.csv -  
method.dempstershafer.mob precision -nodatalabel 0 -undecidedlabel 100 -out fusion_stratum1.tif
```

3. Merging of the fusion results by stratum in an unique result

```
gdalbuildvrt -separate result_of_fusions_bystratum.vrt strata.tif fusion_stratum1.tif  
fusion_stratum2.tif fusion_stratum3.tif
```

```
otbcli_BandMath -il result_of_fusions_bystratum.vrt -out result_of_fusions_stratum1.tif uint8- exp  
"im1b1==1?im1b2:255"
```

```
otbcli_BandMath -il result_of_fusions_bystratum.vrt -out result_of_fusions_stratum2.tif uint8 - exp  
"im1b1==2?im1b3:255"
```

```
otbcli_BandMath -il result_of_fusions_bystratum.vrt -out result_of_fusions_stratum3.tif uint8 - exp  
"im1b1==3?im1b4:255"
```

```
gdalbuildvrt result_of_fusions.vrt result_of_fusions_stratum*.tif
```

```
gdal_translate result_of_fusions.vrt result_of_fusions.tif -co COMPRESS=LZW -co TILED=YES -co  
BIGTIFF=YES
```

Un plus grand nombre de points pourrait améliorer nettement la qualité d'une fusion Dempster-Shafer. Des points créés au niveau de la strate à proximité des bordures pourraient aider à affiner la géométrie des zones artificialisées et aider à la distinction conifères-feuillus. Créer plus de points spécifiques aux demandes de la fusion limite le côté automatisable de la méthode.