



Gembloux Agro-Bio Tech

Unité de Gestion des Ressources forestières et des Milieux naturels

Faculté des Sciences

Unité de Géomatique

TELEDETECTION SPATIALE DE L'OCCUPATION DU SOL EN WALLONIE

Rapport final

Septembre 2010

Convention financée par le Service Public de Wallonie – DGO3



Gembloux Agro-Bio Tech

Unité de Gestion des Ressources forestières et des Milieux naturels

Faculté des Sciences

Unité de Géomatique

TELEDETECTION SPATIALE DE L'OCCUPATION DU SOL EN WALLONIE

Rapport final

Septembre 2010

Encadrement scientifique : Philippe Lejeune, Professeur
Yves Cornet, Chargé de cours

Personnel sous contrat : Stéphanie Bonnet, Delphine Brogna – Assistantes de recherche
Julie Piron, Nadia Poncelet – Assistantes de recherche

Table des matières

1. INTRODUCTION ET ORGANISATION DE L'ETUDE	3
2. ACTIVITES ET RESULTATS	6
2.1. Analyse de la qualité radiométrique de la couche ORTHOS_DGA (Unité de Géomatique)	6
2.1.1. Rappel sur les seamlines et les zones d'étude	6
2.1.2. Analyse statistique à partir des aires de référence du SIGEC	7
2.1.2.1. Statistiques monovariées par affectation dans chaque image mosaïquée	8
2.1.2.2. Analyse des parcelles de prairie permanente découpées par une seamline	11
2.1.3. Analyse à partir d'aires de référence digitalisées	15
2.1.3.1. Moyennes par parcelle dans les aires de résineux digitalisées	15
2.1.3.2. Analyse statistique par classification supervisée	19
2.1.4. Correction radiométrique de l'ortho-image en utilisant l'image SPOT	23
2.1.5. Recherche de la période idéale de prise de vue	25
2.1.6. Conclusion	26
2.2. Mise au point d'une stratégie d'acquisition d'images HR destinées à être utilisées en complément de la couche ORTHOS_DGA (Unité de Géomatique)	28
2.2.1. Détermination du niveau de prétraitement géométrique idéal	28
2.2.1.1. Méthode automatique pour générer des points de certification	28
2.2.1.2. Comparaison des images SPOT de niveau 3 et niveau 3 + DEM	30
2.2.1.3. Comparaison des images SPOT de niveau 3 et niveau 1B	33
2.2.1.4. Comparaison de l'ortho-image et de l'image SPOT de niveau 3	37
2.2.1.5. Conclusion générale	39
2.2.2. Considérations techniques relatives à l'utilisation d'une information diachronique dans les procédures de classification	40
2.2.2.1. Prétraitements	41
2.2.2.2. Classifications et analyse diachronique	51
2.3. Production d'une couche de différence d'altitude entre le MNS_DGA et un MNT (Unité de Géomatique - GRFMN)	54
2.3.1. Analyse des MNT et production d'un MNH	54
2.3.1.1. Données numériques d'élévation disponibles	54
2.3.1.2. Production de couches d'information sur la disponibilité spatiale en données d'altitude	58
2.3.1.3. Analyse de la qualité des MNT	62
2.3.1.4. Production d'un TIN	76
2.3.1.5. Comparaison des MNH construits à partir du TIN et du MNT ERRUISSOL	77
2.3.2. Evaluation de l'apport des MNH dans les procédures de classification	90
2.3.2.1. Analyse des différences des hauteurs données par les MNH dans des zones de résineux	90
2.3.2.2. Analyse dans des zones agricoles	93
2.3.2.3. Conclusion	94
2.4. Préparation d'une base de données d'objets de référence (GRFMN)	95
2.4.1. Objectif	95
2.4.2. Méthodologie	95
2.4.3. Types d'objets de référence	97
2.4.3.1. Objets « géométriques »	97
2.4.3.2. Objets « spectraux »	105

2.5.	Mise au point des procédures de segmentation d'images (GRFMN)	108
2.5.1.	Introduction	108
2.5.2.	Solutions logicielles pour la segmentation	108
2.5.2.1.	Definiens Developer 8.1 (eCognition)	108
2.5.2.2.	ENVI Feature Extraction (ENVI Fx)	110
2.5.3.	Mise au point des procédures de segmentation	111
2.5.3.1.	1 ^{ère} étape : choix de la ou des image(s) de base	111
2.5.3.2.	2 ^{ème} étape : phase préliminaire et exploratoire de segmentation	112
2.5.3.3.	3 ^{ème} étape : tests de segmentation et évaluation de la qualité	112
2.5.3.4.	Mise en œuvre des tests de segmentation	113
2.5.3.5.	Analyse de la qualité de la segmentation	115
2.5.3.6.	Choix de combinaisons de paramètres de segmentation	120
2.6.	Préparation de la base de données « images » (GRFMN)	134
2.6.1.	Introduction	134
2.6.2.	Type d'attributs	134
2.6.2.1.	Attributs radiométriques	134
2.6.2.2.	Attributs «texturaux »	137
2.6.2.3.	Attributs de forme	138
2.6.2.4.	Autres	139
2.6.3.	Développement d'une application informatique pour la préparation de la base de données « images »	139
2.7.	Analyses thématiques (GRFMN – Unité de Géomatique)	142
2.7.1.	Introduction	142
2.7.2.	Calcul des attributs	143
2.7.2.1.	Terminologie	143
2.7.2.2.	Mise en œuvre	144
2.7.3.	Analyse statistique des attributs des segments (Unité de Géomatique)	145
2.7.3.1.	Introduction	145
2.7.3.2.	Analyse et résultats	145
2.7.4.	Analyse intégrée des attributs et classification (GRFMN)	154
2.7.4.1.	Méthodologie	154
2.7.4.2.	Résultats et discussion	158
2.7.4.3.	Conclusions	182
2.8.	Intégration des données produites dans la base de données géographique de la Région wallonne (Unité de Géomatique - GRFMN)	184
2.9.	Recommandations pour les réalisations des futures couches « image satellitale HR » de la Wallonie (Unité de Géomatique - GRFMN)	185
2.10.	Journée d'étude sur l'occupation du sol en Wallonie	186
3.	CONCLUSIONS GENERALES	187
4.	REFERENCES BIBLIOGRAPHIQUES	192

1. Introduction et organisation de l'étude

La présente convention de recherche a pour objet la mise en œuvre du projet « Télédétection Spatiale de l'Occupation du Sol en Wallonie ». Cette étude s'inscrit dans le cadre du Projet de Cartographie Numérique de l'Occupation du Sol en Wallonie (PCNOSW). La convention était initialement prévue pour une durée de 24 mois à dater du 1er septembre 2008 jusqu'au 31 août 2010. Elle a été prolongée, à budget constant, jusqu'au 30 novembre 2010. Ce prolongement a permis l'organisation d'une journée d'étude sur l'occupation du sol en Wallonie le 20 octobre 2010 à Gembloux (cf. paragraphe 2.10).

Le présent document constitue le rapport final des activités de recherche de cette convention.

L'**objectif général** de cette étude est la valorisation des ortho-images à très haute résolution (50 cm et canaux R, V, B, IR) et d'un Modèle Numérique de Surface (MNS, 5m) produits par l'Administration wallonne pour rencontrer les besoins en matière de cartographie de l'espace rural. Ces deux couches de base sont complétées par plusieurs images satellitaires haute résolution (SPOT, 10m) pour disposer de jeux de données multi-dates utiles à la classification thématique de l'occupation du sol.

Les **acteurs** intervenant dans le cadre de cette convention sont l'Unité de Géomatique et l'Unité de Gestion des Ressources forestières et des Milieux naturels (GRFMN) de l'Université de Liège. Les travaux de recherche sont menés conjointement par ces deux équipes. La recherche est structurée en neuf parties. Les différentes étapes préconisées pour la méthodologie ainsi que leur répartition entre les deux équipes sont reprises dans le tableau 1. Lorsqu'il s'agit d'une tâche commune, le partenaire principal est souligné. Les objectifs et résultats de chaque tâche sont explicités aux paragraphes 2.1 à 2.9. Le calendrier des activités est, quant à lui, présenté au tableau 2.

Afin de faciliter la lecture du document, les abréviations des principales données vectorielles et « raster » utilisées sont présentées dans les tableaux 3 et 4.

Tableau 1. Présentation des axes de recherche et répartition entre les deux équipes.

Tâches	Equipe
1. Analyse de la qualité radiométrique de la couche ORTHOS_DGA	UNITÉ DE GÉOMATIQUE
2. Mise au point d'une stratégie d'acquisition d'images HR destinées à être utilisées en complément de la couche ORTHOS_DGA	UNITÉ DE GÉOMATIQUE
3. Production d'une couche de différence d'altitude entre le MNS_DGA et un MNT	<u>UNITÉ DE GÉOMATIQUE</u> - GRFMN
4. Préparation d'une base de données d'objets de référence	GRFMN
5. Mise au point des procédures de segmentation d'images	GRFMN
6. Préparation de la base de données « images »	GRFMN
7. Analyses thématiques	UNITÉ DE GÉOMATIQUE - <u>GRFMN</u>
8. Intégration des données produites dans la base de données géographique de la Région wallonne	<u>UNITÉ DE GÉOMATIQUE</u> - GRFMN
9. Recommandations pour la réalisation d'une couche image HR de la Wallonie	<u>UNITÉ DE GÉOMATIQUE</u> - GRFMN

Tableau 2. Planning des activités de recherche

Mois	Septembre 08	Octobre 09	Novembre 09	Décembre 09	Janvier 09	Février 09	Mars 09	Avril 09	Mai 09	Juin 09	Juillet 09	Août 09	Septembre 09	Octobre 09	Novembre 09	Décembre 09	Janvier 10	Février 10	Mars 10	Avril 10	Mai 10	Juin 10	Juillet 10	Août 10
	Tâches																							
1. U. Géomatique																								
2. U. Géomatique																								
3. U. Géomatique																								
4. GRFMN																								
5. GRFMN																								
6. GRFMN																								
7. GRFMN U. Géomatique																								
8. U. Géomatique GRFMN																								
9. U. Géomatique GRFMN																								

Tableau 3. Dénomination des principales couches de données « raster »

Couche	Dénomination	Résolution
Images aériennes orthorectifiées à très haute résolution fournies par la DGA (2006 - 2007)	ORTHOS_DGA	0.5 m
Modèle Numérique de Surface fourni par la DGA (2006 - 2007)	MNS_DGA	5 m
Modèle Numérique de Terrain issu du projet ERRUISSOL	MNT_ERRUISSOL	10 m
Images satellitales haute résolution (SPOT-5)	SPOT 5	10 m

Tableau 4. Les principales couches de données vectorielles

Couche	Dénomination
Le Système Intégré de Gestion et de Contrôle : inventaire annuel du parcellaire agricole à partir des déclarations de superficies des agriculteurs. Les données sont disponibles pour les deux années de vols (2006 -2007)	SIGEC_2006 et SIGEC_2007
Parcellaire forestier du Département Nature et Forêts, élaboré au niveau des cantonnements	PARC_DNF_canto
Inventaire Permanent des Ressources Forestières de Wallonie : maille de l'inventaire	IFW_POINTS

2. Activités et résultats

2.1. Analyse de la qualité radiométrique de la couche ORTHOS_DGA (Unité de Géomatique)

Cette partie de la recherche vise à mieux évaluer l'importance de l'hétérogénéité de la qualité radiométrique de la couche ORTHOS_DGA, en relation avec le mosaïquage des images originales orthorectifiées. Cette analyse est importante à considérer dans la perspective d'une exploitation de la couche ORTHOS_DGA dans un processus de classification automatique ou semi-automatique.

Un premier contrôle de la qualité radiométrique de l'ortho-image avait déjà été réalisé par l'ULg (objet du MP2006_CTRL_QUALITE_ORTHOPHOTO/MNS). Il s'agissait d'une analyse superficielle des signatures spectrales de quelques parcelles de part et d'autre de limites d'images mosaïquées. Nous réalisons à présent une étude statistique plus complète.

2.1.1. Rappel sur les seamlines et les zones d'étude

Les *seamlines* sont les limites des zones couvertes par chaque image aérienne mosaïquée. La fourniture de ce produit n'était pas spécifiée au cahier des charges de la production de l'ortho-image par la société WALPHOT, mais celle-ci a été contactée et a pu nous fournir ces seamlines pour le bloc 6 uniquement (zone Hesbaye-Condroz-Famenne-Ardenne).

La zone « Sud de l'Ardenne et Lorraine belge » avait été survolée en 2006, et, à cette époque, la société WALPHOT avait utilisé un logiciel différent de celui utilisé en 2007 (pour le bloc 6), moins performant, pour réaliser la mosaïque des ORTHOS_DGA. Le mosaïquage étant de moins bonne qualité radiométrique pour cette zone sud, il a été possible de retracer manuellement les seamlines.

Nous disposons donc des seamlines pour le bloc 6 (Hesbaye-Condroz-Famenne-Ardenne) et pour la zone sud (Sud de l'Ardenne et Lorraine belge), nos deux zones-test (Figure 1). Les seamlines délimitent les images mosaïquées (images aériennes de départ après orthorectification) qui ont chacune un identifiant (numéro de prise de vue) (Figure 2).

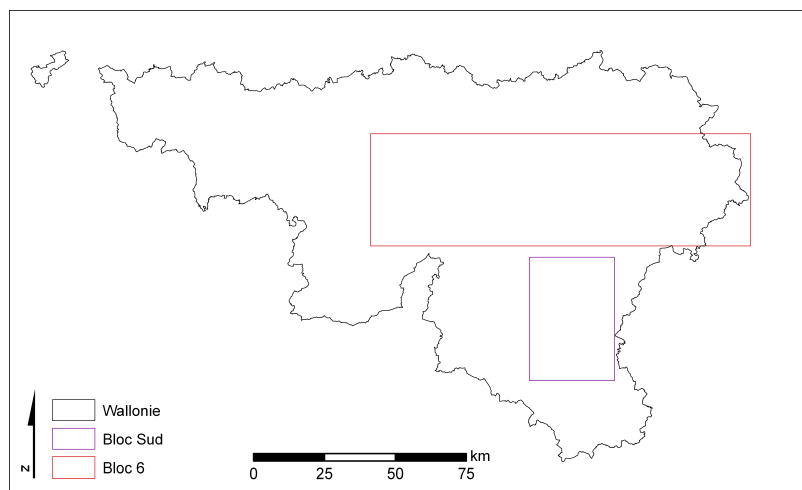


Figure 1. Localisation du bloc 6 et du bloc Sud.

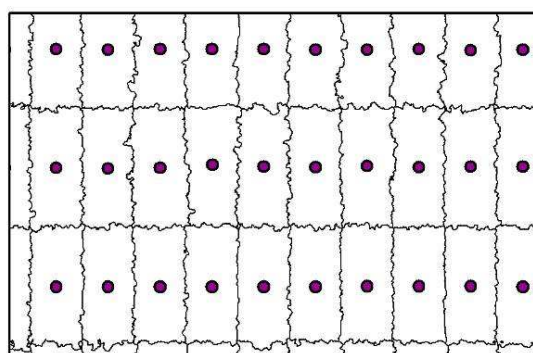


Figure 2. Extrait des seamlines du bloc 6 (2007). Les points mauves sont les centres des images mosaïquées, auxquelles sont associés les identifiants de prise de vue au cours du vol.

2.1.2. Analyse statistique à partir des aires de référence du SIGEC

Le SIGEC (Système Intégré de Gestion Et de Contrôle) est une couche vectorielle constituée de polygones reprenant toutes les parcelles agricoles déclarées annuellement par les exploitants agricoles dans le cadre de l'obtention de primes européennes par la PAC. Il couvre approximativement 95 % de la SAU. Un *code culture* a été attribué à chaque polygone, caractérisant le type d'affectation (prairie permanente, maïs ensilage, froment d'hiver, etc.). A partir de ce parcellaire, nous avons développé plusieurs stratégies pour mettre en évidence et quantifier les éventuelles différences de signature spectrale entre les images mosaïquées.

A cause des imprécisions du levé cartographique du parcellaire SIGEC, nous avons préalablement érodé les polygones de 15 m, afin de s'assurer du fait que le polygone rétréci ne contienne que des pixels de l'affectation voulue.

2.1.2.1. Statistiques monovariées par affectation dans chaque image mosaïquée

a) Bloc 6

Pour une affectation SIGEC donnée, nous avons calculé les statistiques des signatures spectrales dans chacune des 4 bandes spectrales (statistiques monovariées), dans chaque image mosaïquée (délimitée par les seamlines). Nous avons fait ce calcul dans tout le bloc 6, pour les affectations suivantes : prairies permanentes, maïs ensilage, betterave sucrière et épeautre. Les résultats obtenus ne nous ont pas permis de détecter de variation spatiale de la radiométrie due aux seamlines tant la signature spectrale d'une affectation au sein de chaque ortho-image, et même au sein des parcelles, est hétérogène, en raison du caractère précoce des prises de vue en 2007 par rapport à l'émergence de la végétation.

Nous en avons conclu que l'utilisation directe des affectations du SIGEC comme aires d'entraînement ne nous convenait pas pour l'analyse dans cette zone, à cause de la trop grande hétérogénéité spectrale régionale, inter-parcelles et intra-parcelle pour une même affectation, et qu'il fallait utiliser des aires d'entraînement plus homogènes, comme celles que nous avons digitalisées.

b) Bloc Sud

Nous avons également calculé les statistiques par image mosaïquée dans la zone du bloc sud (Ardenne-Lorraine belge), à partir des seamlines que nous avons digitalisées en accentuant le contraste. Parmi les affectations sélectionnées, nous n'avons retenu que les prairies permanentes et le maïs, car les autres affectations ont une trop petite surface et ne sont pas spatialement représentatives de tout le bloc Sud. Les résultats montrent des variations spatiales importantes (Figures 3 et 4), qui sont en partie expliquées par le mosaïquage.

La figure 5 montre un exemple où la différence radiométrique due au mosaïquage est bien visible dans plusieurs affectations. Rappelons que les seamlines pour le bloc sud sont des lignes droites horizontales et verticales, contrairement au bloc 6. On remarque une meilleure continuité spatiale dans la direction est-ouest que dans la direction nord-sud. Ce phénomène est probablement dû au plan de vol de l'avion, qui balaie le territoire selon une direction est-ouest, et peut-être aussi au fait que les images mosaïquées sont des rectangles, et qu'il y a donc plus de problèmes aux endroits les plus éloignés du centre de l'image.

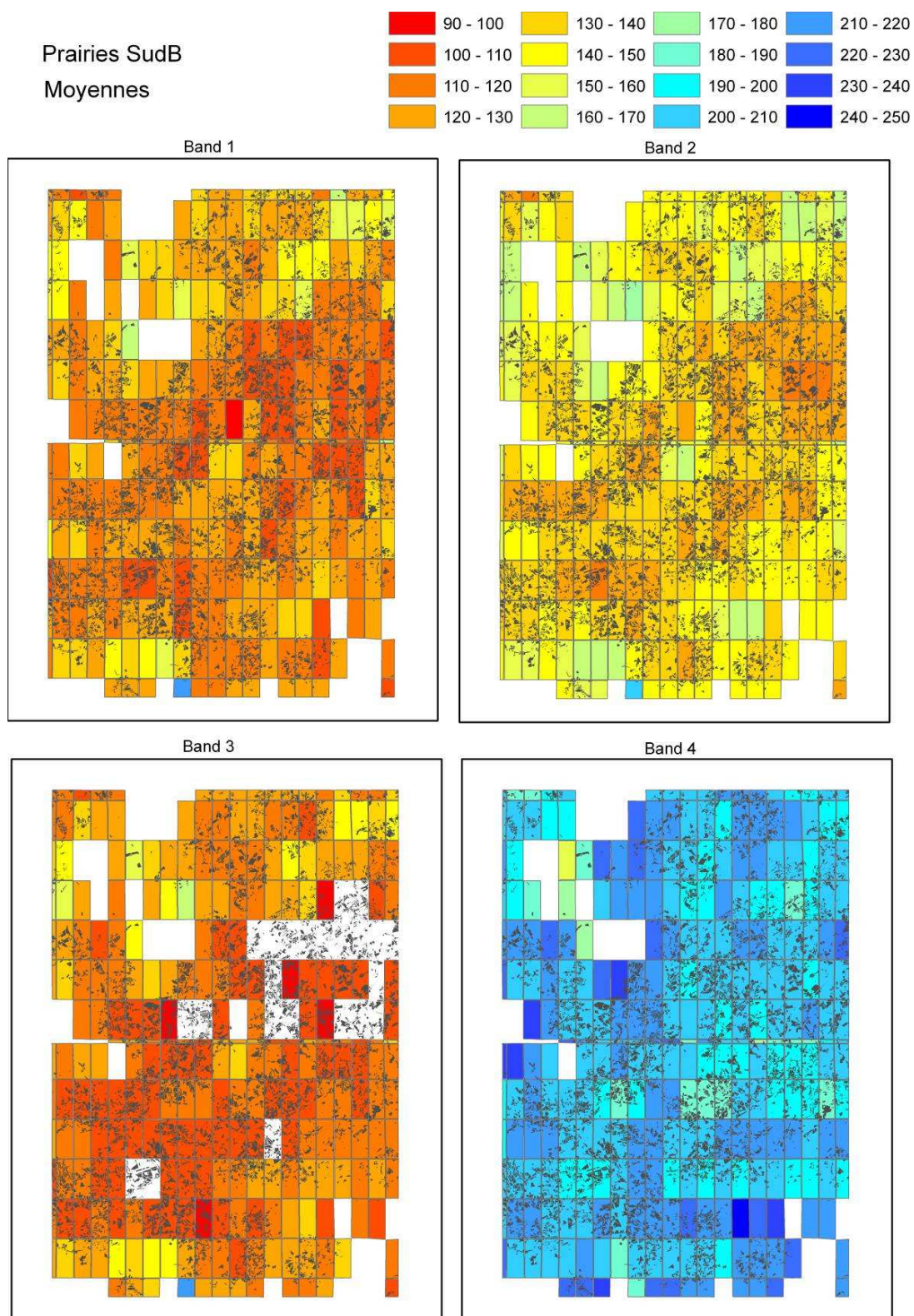


Figure 3. Moyennes des signatures spectrales par image mosaiquée pour les affectations « prairies permanentes » dans le bloc sud.

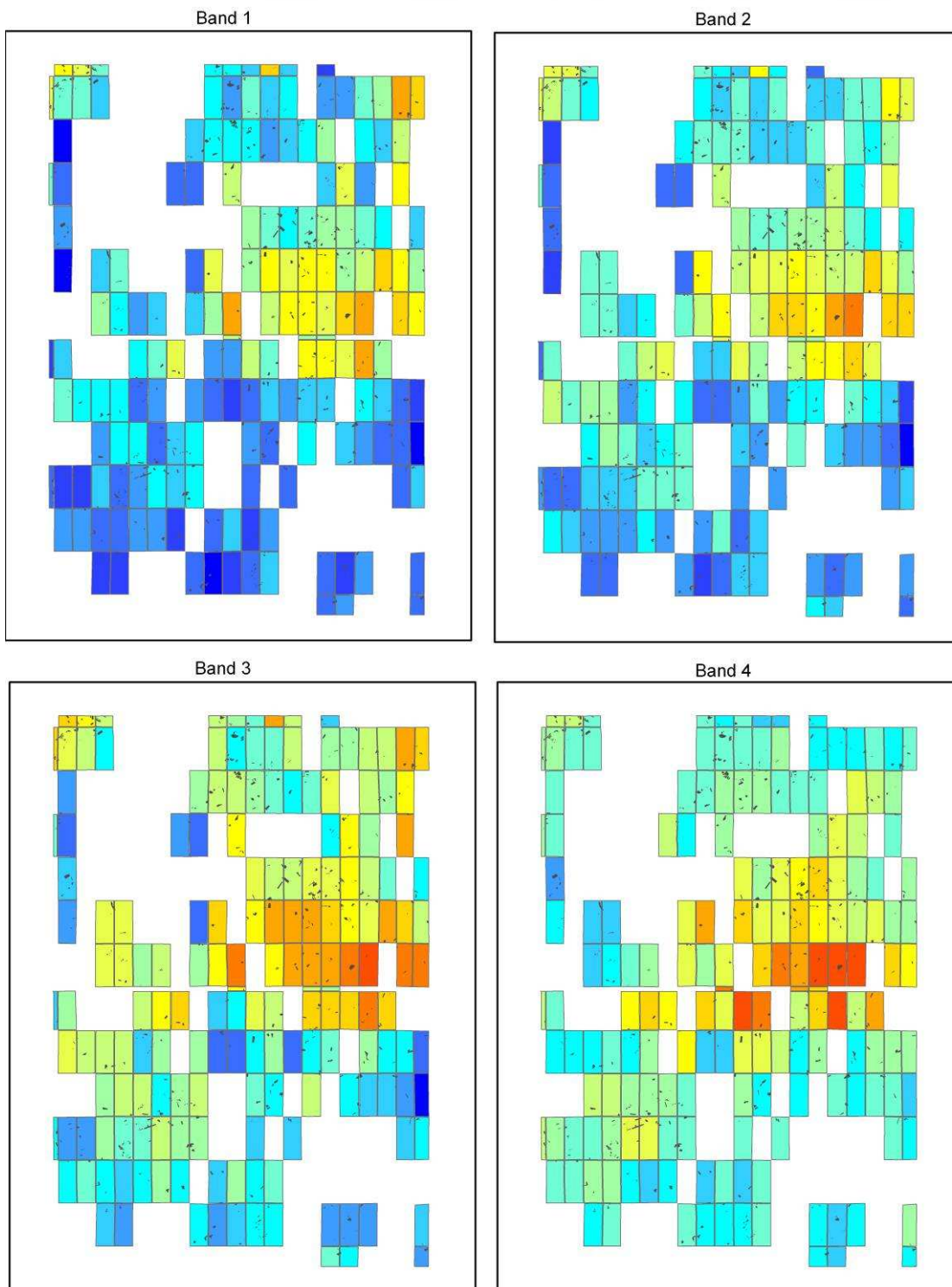


Figure 4. Moyennes des signatures spectrales par image mosaïquée pour les affectations « maïs » dans le bloc sud.

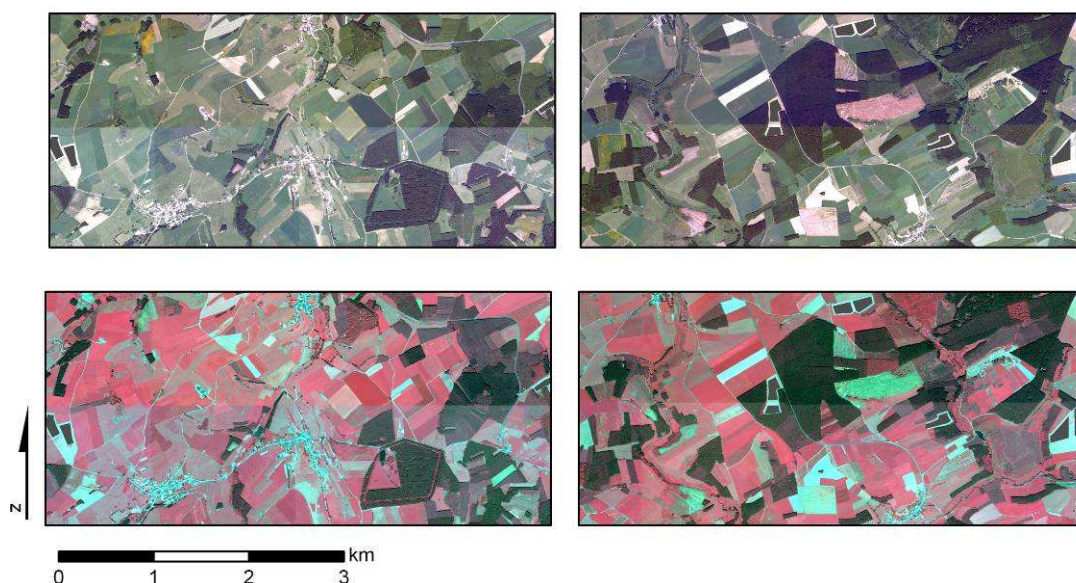


Figure 5. Différences radiométriques dues au mosaïquage, bien visibles dans plusieurs affectations en vraie couleur et en fausse couleur infrarouge, dans le bloc sud.

2.1.2.2. Analyse des parcelles de prairie permanente découpées par une seamline

Une autre stratégie a consisté en l'analyse des statistiques des parcelles traversées par une seamline. Nous avons choisi d'utiliser les parcelles de prairies permanentes uniquement pour cette analyse, car c'est l'affectation la plus représentée dans le bloc 6, elle a donc une bonne représentativité spatiale. Cette analyse a été réalisée dans le bloc 6 uniquement. Nous avons d'abord calculé les statistiques zonales (moyenne et écart-type) dans chaque parcelle traversée par une seamline. Ensuite, nous avons découpé chacune de ces parcelles en deux ou plusieurs parties, délimitées par les seamlines (Figure 6). Les statistiques zonales ont alors été calculées dans ces parties de parcelles.

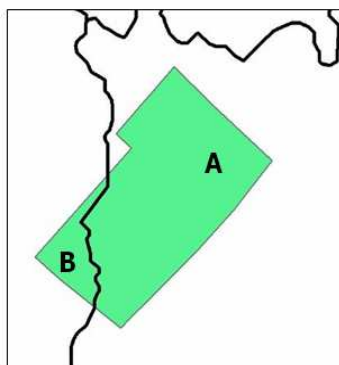


Figure 6. Exemple d'une parcelle découpée par une seamline en 2 parties A et B. Les statistiques zonales sont calculées dans chacune des parties et comparées ensuite entre elles.

Afin de comparer les statistiques calculées dans les différentes parties des parcelles, nous calculons simplement la différence des moyennes :

$$\Delta m = m_A - m_B$$

Avec Δm , la différence des moyennes,
 m_A , la moyenne dans la partie A de la parcelle,
 m_B , la moyenne dans la partie B de la parcelle.

En effet, les échantillons correspondant à chaque partie contiennent un très grand nombre d'individus (pixels), donc, l'erreur sur la moyenne est négligeable. La différence des moyennes est donc représentative pour la comparaison des différentes parties des parcelles. Nous réalisons ce calcul pour toutes les parcelles découpées en 2 parties, à l'aide d'un code programmé sous MATLAB. Les parcelles découpées en 3 parties ou plus ne seront pas considérées dans l'analyse.

Les histogrammes des Δm dans chaque bande sont repris à la figure 7. Les distributions sont normales dans les 4 bandes, mais on remarque quelques outliers. Après avoir éliminé les outliers, le tableau 5 montre que les moyennes des Δm sont proches de zéro, et s'en rapprochent encore plus pour les bandes 2 et 3. C'est dans la bande 4 que la moyenne des Δm est la plus éloignée de zéro. Les écarts-types sont un peu plus importants dans les bandes 1 et 4 mais les dispersions restent relativement faibles dans tous les cas. Les outliers sont des parcelles très hétérogènes (Figure 8) et/ou dont un des deux morceaux est très petit. Dans ce dernier cas, le calcul des différences des moyennes sans prendre en compte l'erreur ne se justifie plus car l'un des échantillons est petit et l'erreur sur la moyenne n'est plus négligeable.

Comme illustré à la figure 9 sur un extrait du bloc 6 pour la bande 1, après élimination des outliers, les différentes valeurs de Δm se répartissent de manière aléatoire dans l'espace. On ne détecte donc pas d'influence significative du mosaïquage sur les différences radiométriques entre les parties de parcelle situées de part et d'autres des seamlines, pour les prairies permanentes. La figure 10 confirme la répartition aléatoire en montrant, par la construction de semi-variogrammes, qu'il n'y a pas de continuité spatiale des Δm .

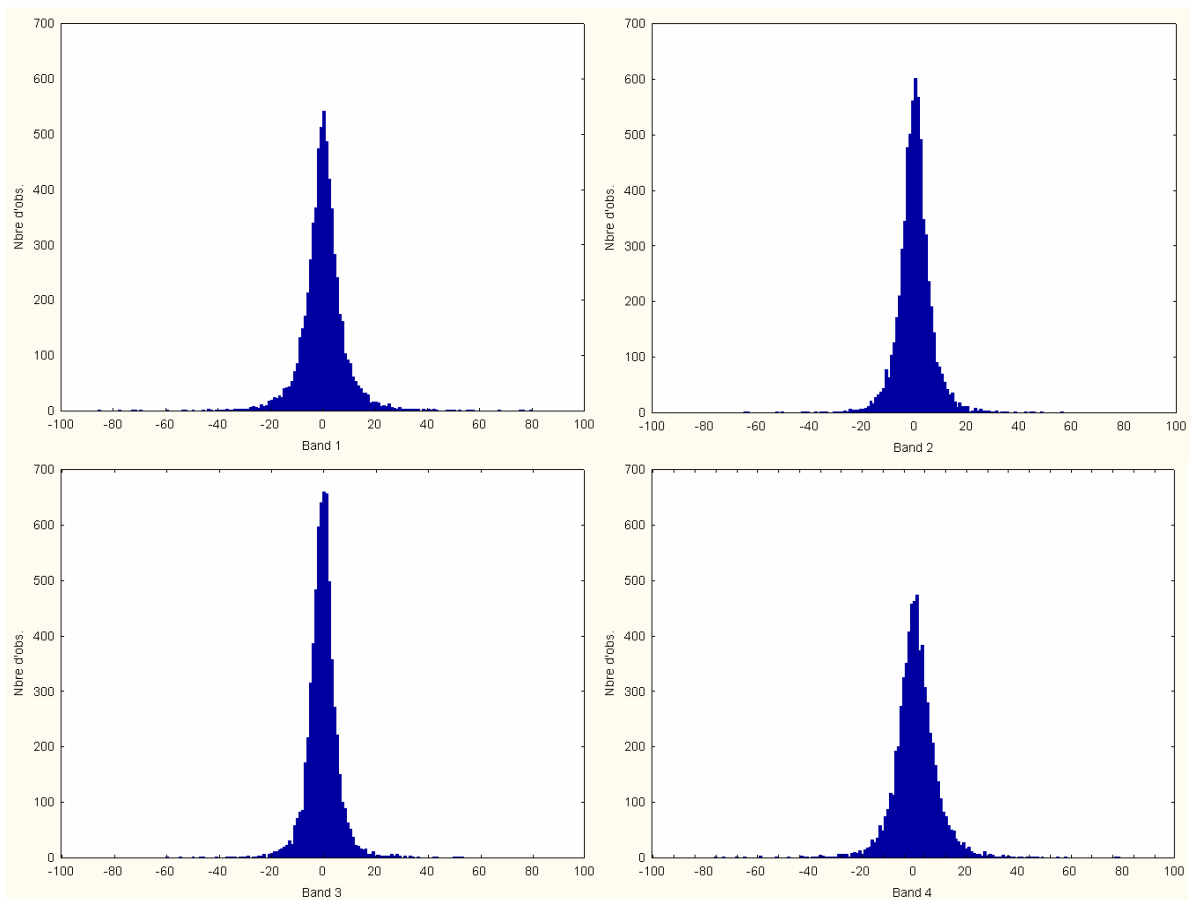


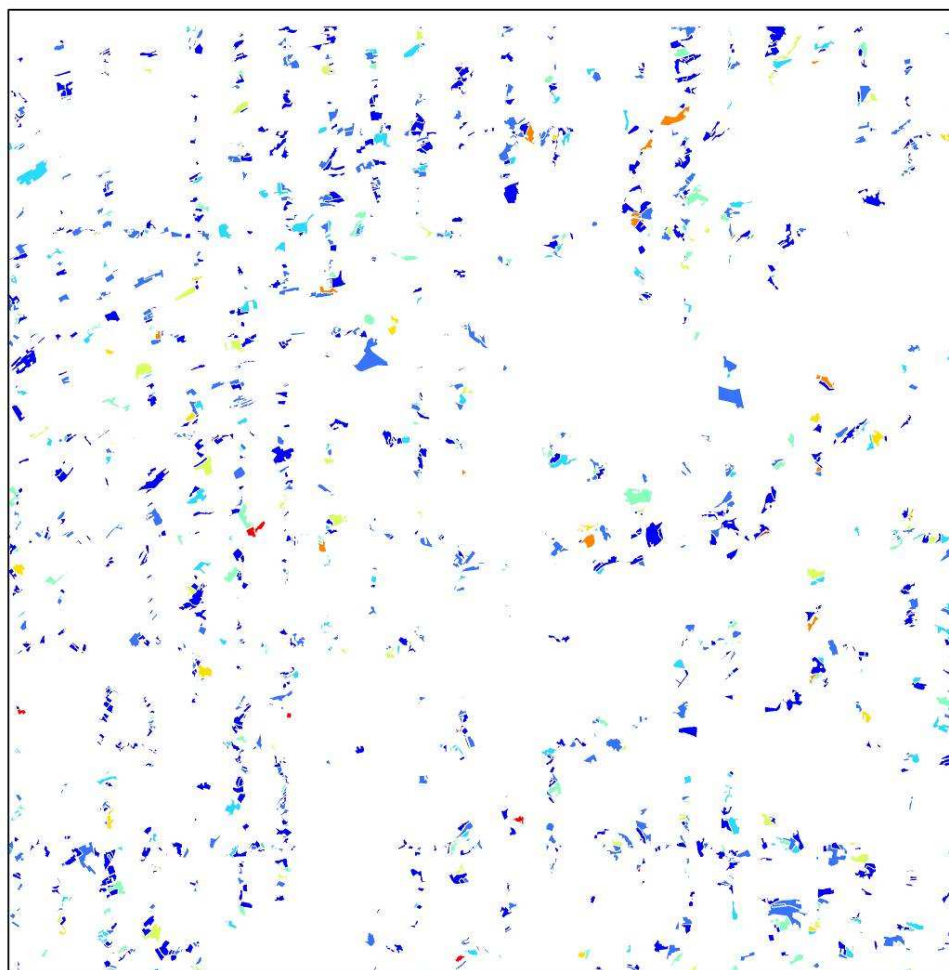
Figure 7. Histogrammes des Δm dans chaque bande.



Figure 8. Exemples d'outliers : parcelles très hétérogènes qui induisent des Δm très importants.

Tableau 5. Moyennes et écarts-types des Δm dans chaque bande

	Band 1	Band 2	Band 3	Band 4
Moyenne	0.3	0.2	0.0	0.9
Ecart-type	9.1	6.9	6.4	8.7



Delta m (valeur absolue) - band 1



Figure 9. Répartition spatiale des valeurs absolues des Δm de la bande 1 dans une partie du bloc 6.

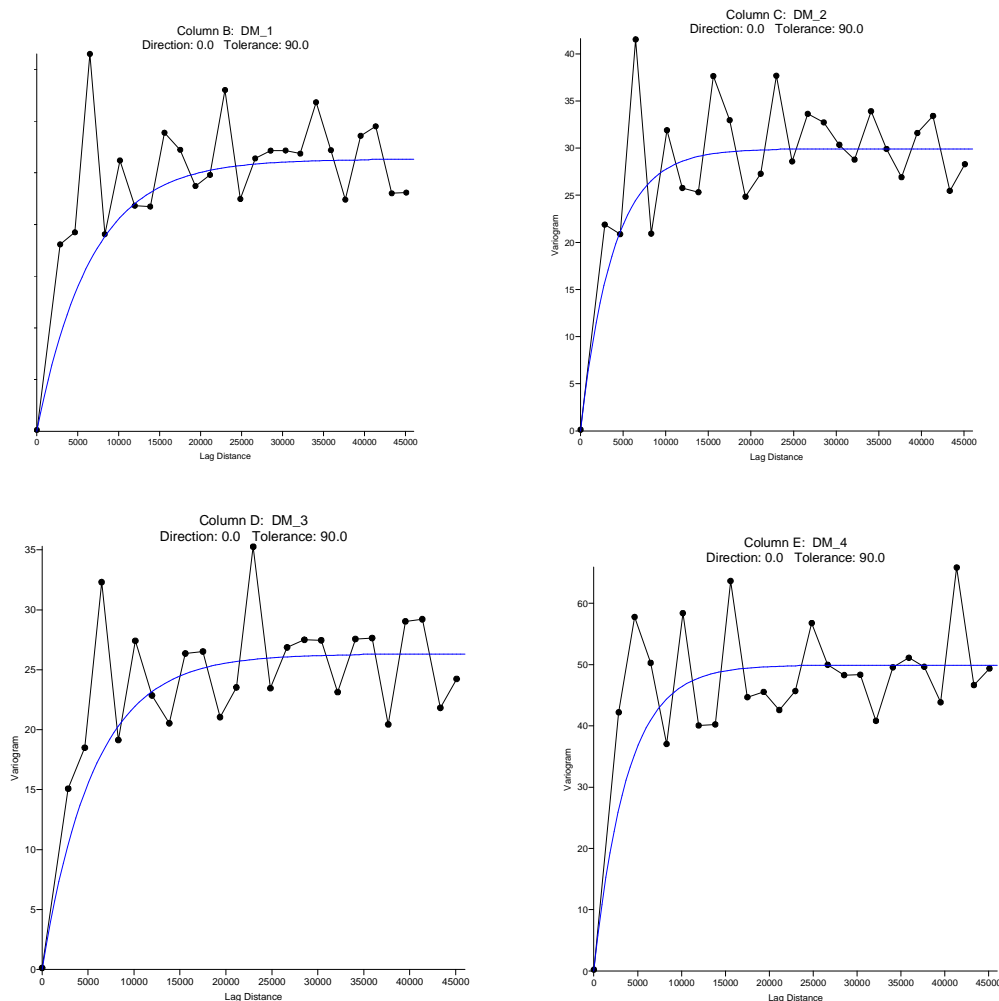


Figure 10. Semi-variogrammes des Δm dans les quatre bandes.

2.1.3. Analyse à partir d'aires de référence digitalisées

Pour réaliser la validation radiométrique de la couche ORTHOS_DGA, le SIGEC n'est pas suffisant. En effet, celui-ci ne reprend que la SAU. Les zones forestières n'y sont donc pas répertoriées. De plus, les affectations du SIGEC utilisées telles quelles ne sont pas assez homogènes du point de vue radiométrique pour pouvoir détecter une influence du mosaïquage sur la radiométrie. Nous avons donc utilisé des aires de référence digitalisées.

2.1.3.1. Moyennes par parcelle dans les aires de résineux digitalisées

Des aires de référence ont été digitalisées dans les zones forestières. Nous avons utilisé 550 parcelles de 'Résineux divers', qui couvrent une superficie totale d'environ 860 000 km². Ces aires de référence sont moins nombreuses et moins étendues que les parcelles du SIGEC, mais beaucoup plus homogènes et cohérentes pour être utilisées dans une analyse statistique.

Nous avons subdivisé les résineux en sous-classes selon leur nature : pessière, pinnaie, mélèzière, douglasière. Seules les pessières ont une représentativité spatiale suffisante permettant une analyse spatiale.

Les statistiques ont été calculées dans les aires de pessière, pour chaque bande, dans le bloc 6 (Figure 11). Remarquons que la variabilité spatiale n'a une signification que dans la partie est du bloc 6, là où la représentativité spatiale est suffisante. La variabilité spatiale peut être liée à divers facteurs, comme l'âge et la taille des épicéas, la densité d'arbres, l'exposition, etc. Cependant, dans la bande 4 (proche infra-rouge), on voit clairement qu'une zone se démarque dans le sud-est du bloc 6, avec des signatures spectrales plus élevées.

Nous avons analysé l'image à l'endroit des 2 seamlines horizontales suspectes qui délimitent cette zone qui s'individualise, que nous appellerons seamlines « nord » et « sud » (Figure 12). En effet, pour la seamline « nord », une différence de couleur est visible sur les compositions colorées en vraie couleur et en fausse couleur infrarouge, principalement dans les affectations forestières et sur les plans d'eau (Figures 13 et 14). Ces affectations couvrent des superficies relativement faibles dans le bloc 6, et il est probable que leur signature spectrale marginale n'ait pas pu être homogénéisée par la procédure radiométrique utilisée lors du mosaïquage, par rapport à d'autres affectations de superficie plus importante, comme les affectations agricoles.

Nous avons également détecté une erreur relative à la seamline « sud ». La figure 15, en effet, montre que la ligne que la société WALPHOT nous a renseignée comme seamline (en rouge), n'est pas la seamline qui a été utilisée pour réaliser le mosaïquage. Nous avons pu retracer la véritable seamline à certains endroits, dans les affectations forestières (en jaune), en observant la différence radiométrique dans les compositions colorées en fausse couleur infrarouge. Cependant, cette nuance n'est pas visible dans les autres affectations.

L'erreur détectée relative à la seamline « sud » s'étend probablement dans toute la partie sud du bloc 6. Nous avons également remarqué des différences radiométriques le long de lignes horizontales que nous avons éliminées des seamlines en pensant qu'il s'agissait de repères géographiques. Il est donc clair que nos analyses précédentes sont remises en cause par ces constatations. En effet, la production des seamlines n'a pas été spécifiée au cahier des charges de WALPHOT, et les seamlines récupérées ne sont probablement pas toutes celles qui ont été réellement exploitées dans la procédure de mosaïquage. Il convient donc, dans le futur, de spécifier cette production de seamlines au cahier des charges dès le début de la production de l'ortho-image.

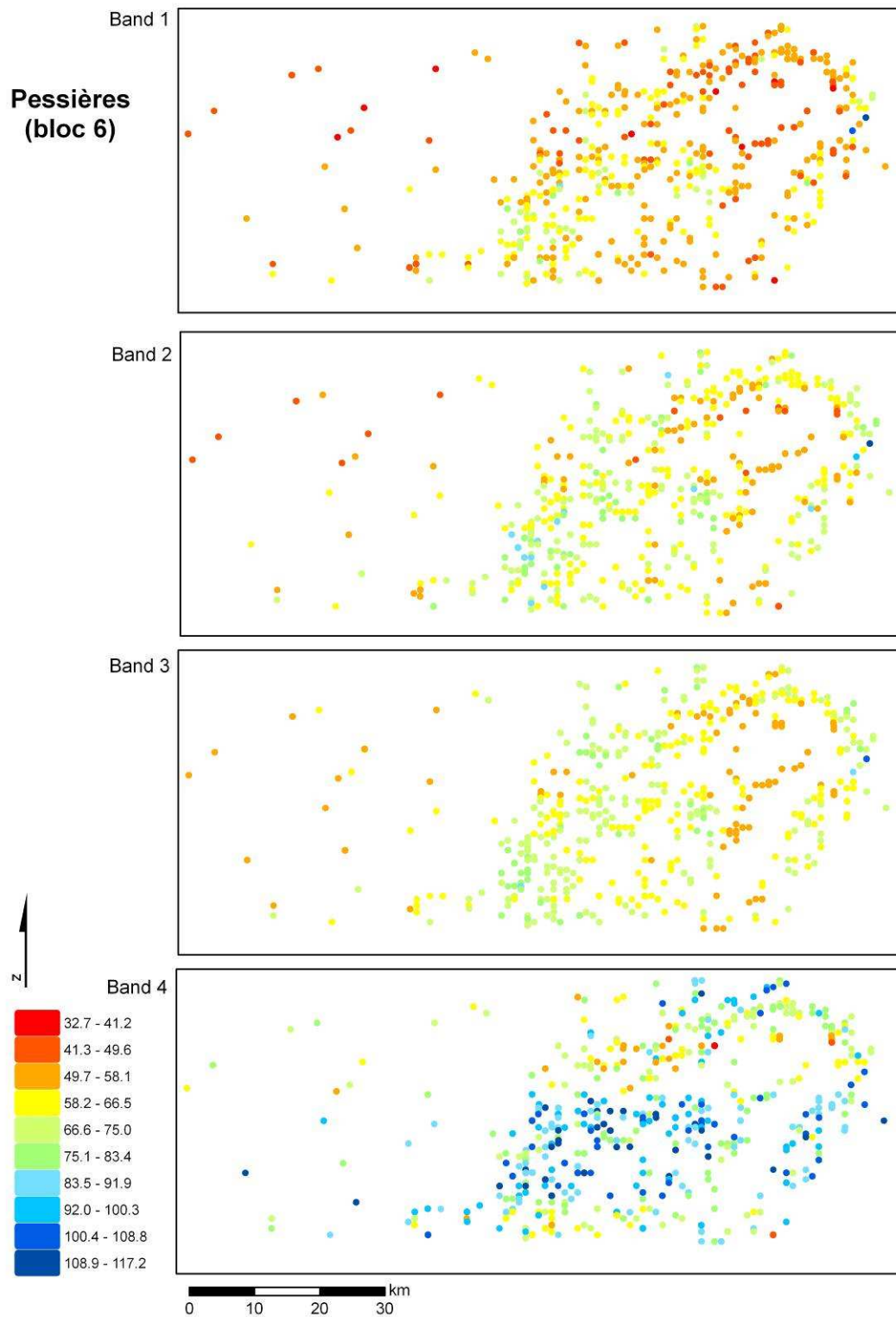


Figure 11. Moyennes des signatures spectrales dans les aires de pessières, dans le bloc 6.

**Pessières
(bloc 6)**

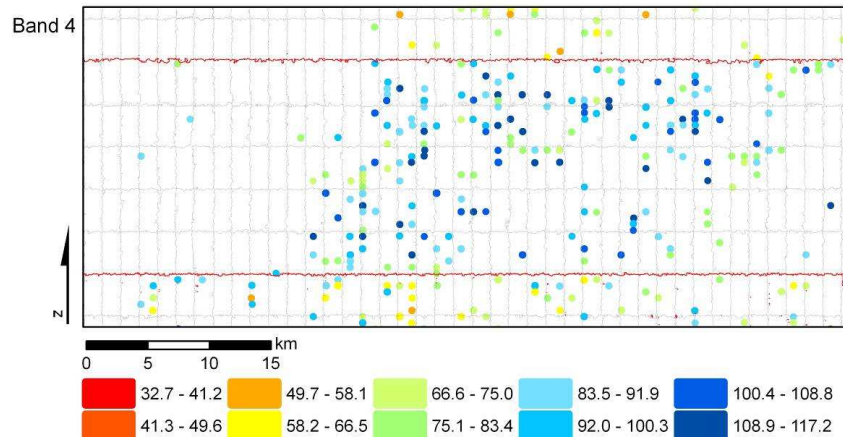


Figure 12. Localisation des seamlines suspectes « nord » et « sud » (en rouge), qui ont été examinées plus attentivement.

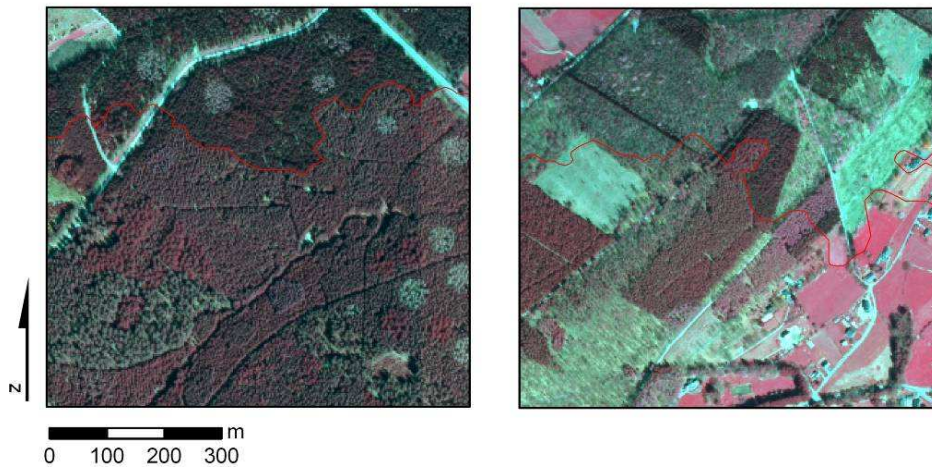


Figure 13. Différence radiométrique due à la seamline « nord », visible en fausse couleur infrarouge dans les affectations forestières.

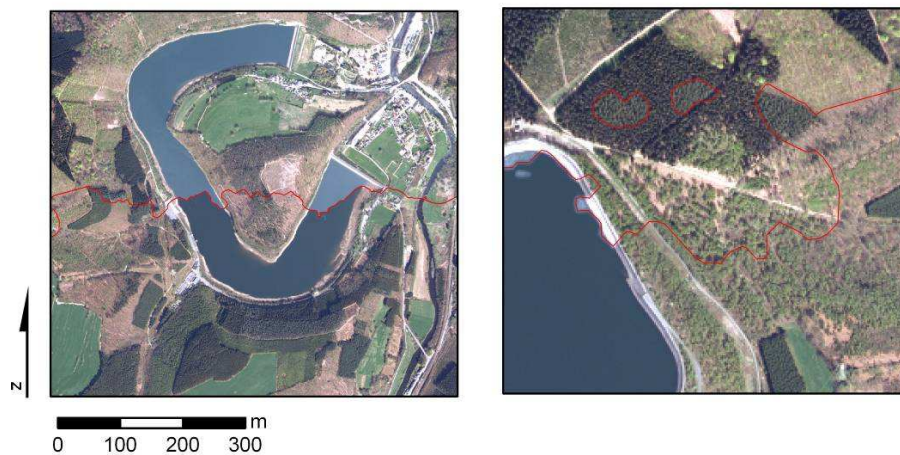


Figure 14. Différence radiométrique due à la seamline « nord », visible en vraie couleur dans les affectations forestières et sur les plans d'eau.

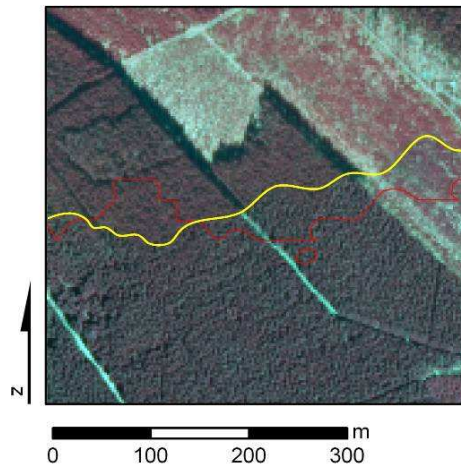


Figure 15. Preuve d'une erreur dans la seamline « sud », et différence radiométrique due à la seamline « sud », en fausse couleur infrarouge dans les affectations forestières. La seamline erronée en rouge est celle qui nous a été fournie par WALPHOT. La ligne jaune représente approximativement la seamline « sud » réelle, d'après les différences radiométriques observées dans les affectations forestières.

Les statistiques par parcelles dans les pessières et l'observation détaillée des seamlines suspectes ont permis non seulement de détecter une erreur dans les seamlines de la partie sud du bloc 6, mais aussi d'arriver à la conclusion suivante en ce qui concerne la validation radiométrique : l'ORTHOS_DGA présente une différence de radiométrie liée au mosaïquage non négligeable pour les affectations forestières et l'eau. En ce qui concerne l'eau, cela avait d'ailleurs déjà été remarqué dans le premier contrôle de la qualité radiométrique de l'ortho-image (objet du MP2006_CTRL_QUALITE_ORTHOPHOTO/MNS). L'exploitation des valeurs radiométriques est donc à proscrire pour la réalisation de la classification des zones forestières.

2.1.3.2. Analyse statistique par classification supervisée

Pour atténuer les variations radiométriques dues à la très haute résolution, l'ortho-image (0,5 m de résolution) a été dégradée par agrégation spatiale (moyenne) de 25 pixels, pour obtenir une image à 2,5 m de résolution. Nous avons également produit l'image dégradée des variances, qui nous donne une indication sur l'hétérogénéité des signatures spectrales au sein de chaque pixel de 2,5 m.

Nous avons effectué une analyse de la qualité radiométrique basée sur des méthodes de classification supervisée par pixel, appliquée à l'image dégradée.

Nous avons d'abord digitalisé un certain nombre d'aires d'entraînement dans la partie nord-ouest du bloc 6, à l'intérieur des affectations SIGEC. La méthode utilisée est la suivante. Pour chaque affectation SIGEC, une ou deux parcelles bien représentatives de l'affectation sont choisies. Au sein de ces parcelles, la délimitation de la zone échantillonnée est effectuée en appliquant le principe du '*Region Growing*', par propagation sur base d'un critère de distance spectrale qui a été optimisé pour chaque sélection (Figure 16). Cela nous permet de ne pas tenir compte d'objets qui 'pollueraient' les signatures extraites des aires d'entraînement (arbres, silos, ombres, etc.).

La délimitation de ces aires d'entraînement est cruciale pour le reste de l'analyse, car la qualité de la classification dépend de leur qualité. Nous ne prenons pas d'aire d'entraînement lorsqu'il s'agit de sols nus, ni pour les affectations trop marginales (moins de 10000 pixels dans le bloc 6). Seuls les pixels appartenant aux hyperellipsoïdes (4D) de probabilité à 95 % sont classés. Ceux qui se trouvent en dehors de ces hyperellipsoïdes restent non classés.

Nous cherchons à savoir si certaines classes peuvent se regrouper, en déterminant la « séparabilité » entre les classes, en calculant la distance euclidienne entre les moyennes des signatures spectrales de chacune des classes 2 à 2, dans l'espace spectral à 4 bandes. Nous analysons les classes qui ont une faible « séparabilité » (moins de 10). Les classes qui ont des signatures spectrales difficilement différenciables sont regroupées dans le tableau 6.

Certaines affectations du SIGEC sont trop hétérogènes pour pouvoir donner lieu à des classes. Ce sont des affectations qui regroupent plusieurs types de cultures différentes, mais que nous n'avons pas éliminé de la classification car elles couvrent une partie non négligeable du territoire : couvert naturel, couvert de légumineuses, couvert de légumineuses et graminées, couvert de graminées, jachères, jachères à base d'un mélange certifié de semences, autres fourrages, ... D'autres affectations ne sont pas différenciables et doivent être regroupées selon la logique : les prairies et les tournières enherbées, les conifères (pins noirs, douglas, épicéas). Les affectations non différenciables mais qu'il n'est pas logique de regrouper doivent être considérées avec prudence, sachant que des erreurs peuvent exister (colza et navette d'hiver, orge d'hiver, couvert de légumineuses, froment d'hiver, cultures maraîchères, ...).

Nous réalisons une classification supervisée à partir des aires d'entraînement délimitées. Les classes que nous avons définies pour la classification ne correspondent pas toujours au SIGEC, car nous avons parfois subdivisé les affectations SIGEC en plusieurs catégories selon leurs caractéristiques radiométriques. Par exemple, l'affectation 'prairie permanente' du SIGEC peut correspondre à 'prairie très claire', 'prairie claire', 'prairie moyenne' ou 'prairie foncée' dans la classification.

Nous analysons chaque affectation SIGEC en fonction de la façon dont elle se répartit dans les classes de la classification. Pour cela, nous comptons le nombre de pixels de chaque affectation SIGEC qui ont été classés dans chacune des classes de la classification. Pour chaque affectation SIGEC, nous connaissons alors le pourcentage de pixels classés dans chacune des classes de la classification.

Le tableau 7 reprend les pourcentages de pixels bien classés pour chaque affectation. Par exemple, pour l'affectation SIGEC 'Triticale', 19 % des pixels ont été effectivement classés dans la classe 'Triticale' que nous avons définie. Pour certaines affectations SIGEC, comme les prairies permanentes, nous calculons la somme des pourcentages des pixels classés dans les différentes classes que nous avons définies, ici, 'prairie très claire', 'prairie claire', 'prairie moyenne' et 'prairie foncée'.

Tableau 6. Groupes de classes qui ont une séparabilité plus petite que 10.

prairies claires colza et navette d'hiver verts tournières enherbées cultures maraîchères vertes couvert de légumineuses A	prairies foncées orge d'hiver moyen	froment d'hiver clair couvert de graminées D
froment d'hiver moyen triticale couvert naturel B couvert de graminées A couvert de graminées B jachère à base d'un mélange certifié de semences (851) A	orge d'hiver fonce orge d'hiver moyen couvert de graminées	seigle d'hiver couvert de graminées et légumineuses A
épeautre autres fourrages	couvert naturel C couvert de graminées D	couvert de graminées B jachère B jachère à base d'un mélange certifié de semences (851) A
cultures maraîchères vertes autres fourrages clairs	épicéas pins noirs douglas	



Figure 16. Délimitation d'une aire d'entraînement par 'Region Growing'. Seuls les pixels les plus proches de la valeur spectrale d'un pixel représentatif choisi sont pris en compte.

Tableau 7. Part des pixels bien classés pour chaque affectation SIGEC (%).

Code culture		Pourcentage de pixels bien classés
	0	
35	Triticale	19.0
36	Epeautre	1.9
61	Prairie permanente	28.4
72	Trèfle	23.7
73	Luzerne	10.2
81	Couvert naturel	22.4
82	Couvert de graminées	8.4
83	Couvert de légumineuses	10.9
84	Couvert graminées + légumineuses	4.5
85	Autres couvertures avec droits de jachère	0.0
86	Autres couverts avec droits ordinaires	48.5
311	Froment d'hiver	6.9
312	Froment de printemps ou froment alternatif	12.7
321	Orge d'hiver	6.2
331	Seigle d'hiver	18.9
411	Colza et navette d'hiver	10.1
412	Colza et navette de Printemps	38.5
613	Pâturage à statut particulier	79.4
743	Autres fourrages	1.8
751	Tournière enherbée	12.7
851	Couvert à base d'un mélange certifié de semences	5.1
883	Cultures forestières à rotation courte	44.3
951	Cultures maraichères	2.8
971	Cultures fruitières pluri-annuelles	34.9
7431	Ortie	6.1

On remarque que la part des pixels bien classés est relativement faible en général. On a trois cas : soit, l'affectation est complètement hétérogène et il est logique que la plupart des pixels soient classés ailleurs ('mélange de semences', jachère, 'cultures maraîchères', 'autres fourrages'), soit l'affectation comprend une bonne partie de sols nus, soit il existe des confusions entre plusieurs classes qui doivent être regroupées. La haute proportion de pixels bien classés pour les 'pâturages à statut particulier' s'explique par le fait que quelques parcelles ont une superficie beaucoup plus grande que les autres, et que l'aire d'entraînement a été prise dans l'une de ces parcelles. Ces grandes parcelles ont donc été bien classées et cela représente un pourcentage important de pixels pour l'affectation.

2.1.4. Correction radiométrique de l'ortho-image en utilisant l'image SPOT

Nous avons réalisé un test de correction de la radiométrie de l'ortho-image à l'aide de l'image SPOT. Pour chacune des affectations du SIGEC, le principe de correction repose sur la comparaison des statistiques par parcelle calculées dans l'image SPOT avec celles calculées dans l'ortho-image, et d'en tirer une relation permettant ensuite de corriger l'image.

Nous avons donc calculé les moyennes par parcelle, pour toutes les affectations du SIGEC, dans les images SPOT n° 01 et 02 séparément. Ces images couvrent, à elles deux, une bonne partie du bloc 6 (Figure 17). Nous les avons alors comparées avec les moyennes par parcelles calculées de la même façon dans l'ortho-image. Le tableau 8 indique les correspondances entre les bandes de l'ortho-image et des images SPOT.

Tableau 8. Correspondance spectrale des bandes de l'ortho-image et des images SPOT.

	Ortho_DGA	SPOT
Bleu	Band 3	
Vert	Band 2	Band 3
Rouge	Band 1	Band 2
Proche infrarouge	Band 4	Band 1
Infrarouge moyen		Band 4

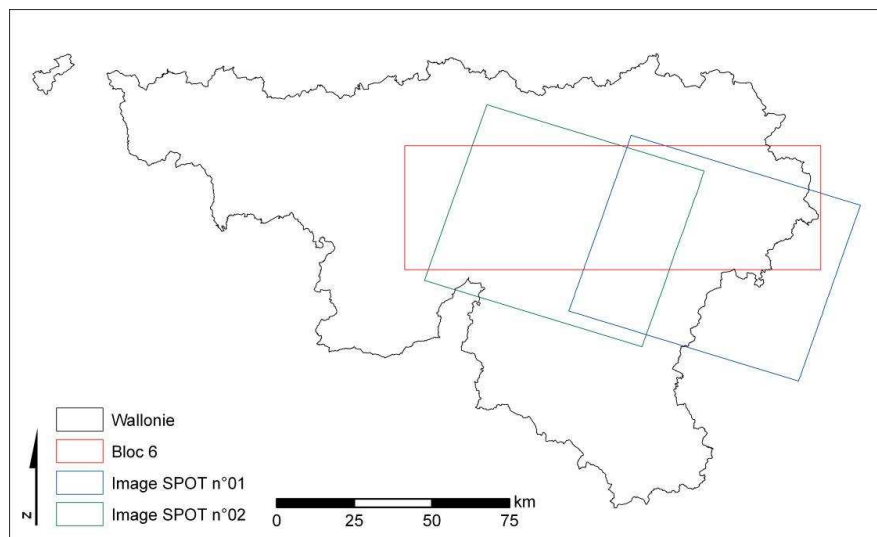
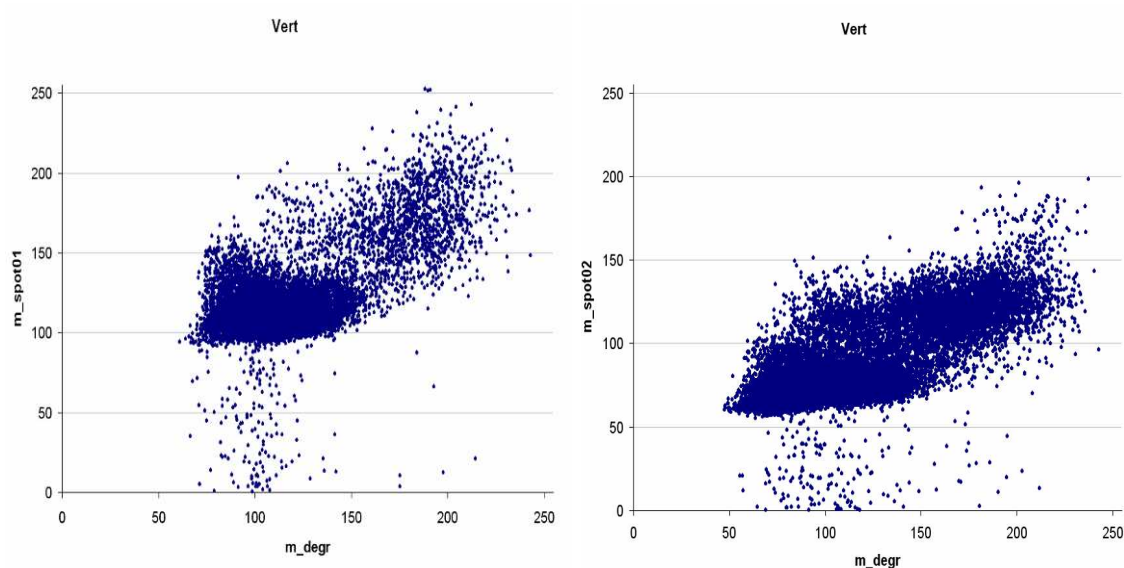


Figure 17. Localisation des images SPOT 01 et 02.

La figure 18 montre la relation qu'il existe entre les moyennes par parcelle de l'ortho-image (en x) et des images SPOT (en y). On compare l'ortho-image à l'image SPOT 01 sur les graphiques de gauche, et à l'image SPOT 02 sur les graphiques de droite, dans le vert, le rouge et le proche infrarouge, pour lesquelles l'ortho-image et l'image SPOT ont un enregistrement spectral.

Ces graphiques montrent que l'on est loin d'obtenir une relation mathématique évidente, même si on observe une légère tendance. Rappelons que les moyennes ont été calculées dans des zones agricoles. Il serait intéressant de réaliser cette même analyse dans des zones de forêts lorsque nous disposerons d'une base de données suffisamment représentative spatialement pour les affectations forestières.



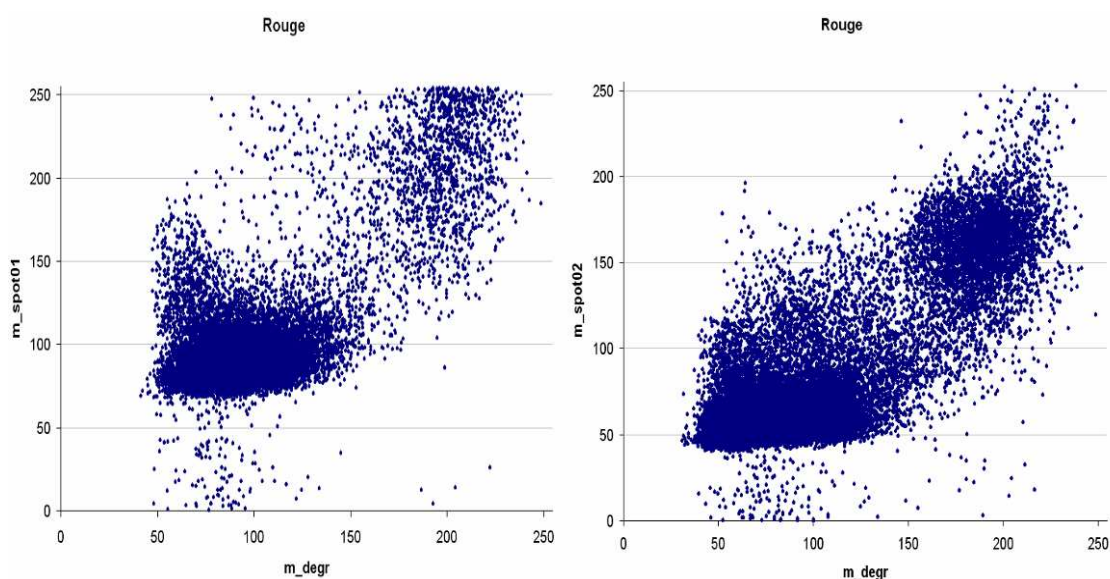


Figure 18. Comparaison des moyennes par parcelle de l'ortho-image (en x) par rapport à celles des images SPOT (en y), l'image 01 à gauche et l'image 02 à droite, dans le vert, le rouge et le proche infrarouge.

2.1.5. Recherche de la période idéale de prise de vue

Afin d'argumenter la décision quant à la meilleure période pour réaliser les vols de l'ortho-image, nous avons effectué une recherche bibliographique dont nous rapportons ci-dessous une synthèse.

Lorsque des images à des périodes différentes de l'année sont utilisées, la carte de l'occupation du sol peut être affinée. En effet, certaines affectations ne sont distinguables qu'à des périodes de l'année bien précises.

Par exemple, au Pays de Galles (Lucas et al., 2007), les images prises au début du printemps, avant l'apparition des feuilles des arbres, permettent de distinguer les fougères, les feuillus et d'autres types de couvertures. Sur celles prises à la fin du printemps, pendant l'émergence des feuilles, on reconnaît les feuillus, les conifères et les forêts mixtes. En été, la végétation est au maximum de productivité et on peut différencier presque toutes les espèces végétales. A la mi-automne, on voit les herbages et les landes, et en hiver, les cultures hivernales.

Dans la vallée du Mississippi où les pratiques agricoles ne sont malheureusement pas tout à fait comparables à celles de nos régions, Gorham (1998) distingue sur des images Landsat TM : au printemps, le blé et l'avoine, les sols nus, les herbages, les fourrages, les légumes et les herbages d'hiver ; en été, le riz, les jachères, le soja, le coton, le maïs et le sorgho ; en automne, le coton et le soja.

D'après les constatations du *California Department of Fish and Game* (site 3) à propos de la cartographie de la végétation, en Californie, toutes les affectations agricoles ont des signatures spectrales proches au printemps, et sont donc difficilement distinguables. Au milieu ou à la fin de l'été, la distinction devient possible pour la plupart des affectations.

La meilleure période de prise de vue pour distinguer un maximum d'occupations du sol différentes semble être l'été. Certaines affectations spécifiques peuvent cependant être mieux cartographiées à d'autres périodes de l'année. Mais réaliser une carte de l'occupation du sol complète et correcte, il est préférable d'utiliser plusieurs images à des périodes différentes de l'année.

2.1.6. Conclusion

L'analyse de la qualité radiométrique n'est pas facile à réaliser car la variation spatiale de la radiométrie est influencée par de nombreux facteurs, comme les variations régionales, le stade de développement de la végétation, l'humidité, les différents types de végétation, ... Il est donc difficile d'isoler l'effet du mosaïquage de ces autres facteurs. Il est préférable d'utiliser des aires de référence digitalisées, plutôt que le SIGEC, où les parcelles sont trop hétérogènes. Il est nécessaire d'utiliser des zones très homogènes si l'on veut mettre en évidence de faibles effets du mosaïquage.

Dans le bloc sud (ortho-image de 2006), les effets du mosaïquage sont très visibles. Par contre, dans le bloc 6 (ortho-image de 2007), la radiométrie a été homogénéisée de manière beaucoup plus précise lors du mosaïquage et les effets sont très peu visibles. Cependant, les affectations les moins représentées comme les affectations forestières ou l'eau sont pénalisées dans les corrections radiométriques effectuées pour le mosaïquage quand elles sont peu couvrantes. Dans ces affectations, on remarque une différence radiométrique le long de certaines seamlines horizontales. Nous n'avons pas détecté d'effets des seamlines sur la radiométrie dans les affectations agricoles.

La production de la couche des seamlines n'avait pas été spécifiée au cahier des charges de WALPHOT, et une partie des seamlines récupérées ne sont pas celles qui ont été réellement exploitées dans la procédure de mosaïquage. Nous avons pu prouver cela en observant les affectations de résineux et les plans d'eau le long d'une seamline dans le sud du bloc 6. Il convient donc, dans le futur, de spécifier au cahier des charges de la production de l'ortho-image l'obligation de fournir la couche des seamlines.

Une première classification par pixel a été réalisée à partir d'aires d'entraînement digitalisées. Le but de cette classification était de caractériser les affectations SIGEC par rapport à des classes plus homogènes générées par cette classification supervisée. On remarque que la plupart des classes SIGEC correspondent très mal aux classes de la classification auxquelles elles devraient théoriquement se rapporter. De plus, une bonne partie des classes sont difficilement distinguables et il paraît nécessaire de les regrouper, dans un premier temps, pour les pré-classifications.

Les résultats de cette analyse ne permettent pas de tirer une conclusion sur la qualité radiométrique de l'ortho-image, mais pourront peut-être être utiles dans les procédures de classification à venir.

La correction radiométrique de l'ortho-image à partir de l'image SPOT avait été envisagée. Dans le bloc 6, cependant, il n'y a pas de relation claire entre les signatures spectrales des images SPOT et de l'ortho-image.

La période de prise de vue est importante pour pouvoir distinguer un maximum d'occupations du sol différentes. Après une recherche bibliographique exploratoire à ce sujet, nous concluons que la meilleure période de prise de vue est l'été, lorsque la végétation est à son maximum de productivité. Si on veut se focaliser sur des affectations particulières, il faudra envisager éventuellement des prises de vue à des périodes spécifiques pour ces affectations. Mais pour réaliser une carte complète de l'occupation du sol, il paraît indispensable d'utiliser conjointement des images à des dates différentes.

2.2. Mise au point d'une stratégie d'acquisition d'images HR destinées à être utilisées en complément de la couche ORTHOS_DGA (Unité de Géomatique)

Cette partie vise à étudier les possibilités d'utiliser des images satellitaires en combinaison avec la couche ORTHOS_DGA. Elle a un double objectif. Il s'agit, d'une part, d'évaluer l'intérêt de l'information diachronique de cette imagerie satellitaire dans la procédure de classification et, d'autre part, de déterminer le niveau de pré-traitement géométrique idéal dans le contexte wallon pour exploiter cette imagerie en combinaison avec la couche ORTHOS_DGA.

2.2.1. Détermination du niveau de prétraitement géométrique idéal

Afin de déterminer le niveau de pré-traitement géométrique idéal, nous disposons de 3 mêmes images SPOT à différents niveaux de pré-traitement : niveau 1B, niveau 3 et niveau 3 + DEM. Le niveau 1B correspond à l'image brute, non rectifiée. Le niveau 3 correspond à une image orthorectifiée standard (Basic Ortho Standard). L'image niveau 3 + DEM est l'image niveau 3 qui a été orthorectifiée par SPOT Image à partir d'un modèle numérique de terrain et de points de contrôle que nous lui avons fournis.

Nous avons comparé les images en calculant les vecteurs-erreurs à partir de points de certification. Ces points doivent être repérables et localisables dans les deux images sans ambiguïté (par exemple au croisement de routes, sur un pont, au coin d'un bâtiment, etc.). Une fois ces points localisés dans les deux images, on calcule leur différence de position dans les deux images, ce qui nous donne les composantes du vecteur-erreur pour chaque point. Les images SPOT ont été comparées entre elles par une méthode automatique, et l'ortho-image a été comparée à l'image SPOT par une méthode manuelle.

2.2.1.1. Méthode automatique pour générer des points de certification

Nous avons développé une méthode qui génère des points de certification automatiquement. Au sein de zone dépassant un certain seuil de variance, le logiciel Erdas Imagine génère 2000 points aléatoires dans l'une des images. Il est important d'établir un seuil de variance afin de générer des points repérables et localisables sans ambiguïté dans les deux images. Ces points doivent donc être situés sur des discontinuités, à forte variance focale.

Nous faisons ensuite correspondre manuellement 6 points. A partir de ces 6 premiers points, le logiciel calcule un modèle polynomial du deuxième degré qui prédit la position de tous les autres points. Finalement, le logiciel réalise un matching spectral : il fait correspondre le point dans la deuxième image en se basant sur les points situés autour du point en question, et en maximisant le coefficient de corrélation calculé sur les couples de valeurs spectrales du voisinage au sein des 2 images. La fonction de voisinage est mobile autour de la position prédite dans la seconde image.

Nous avons contrôlé visuellement la précision de cette procédure, et elle s'est avérée très fiable. Elle permet de générer un très grand échantillon de points rapidement. Elle ne fonctionne que pour des images semblables (ce qui est le cas ici) ou très similaires. Nous n'avons donc pas pu l'utiliser pour comparer l'ortho-image à l'image SPOT.

En fonction des coordonnées obtenues pour ces points de certification dans les deux images, nous calculons les composantes du vecteur-erreur en chaque point :

$$e_x = X_{input} - X_{ref}$$

$$e_y = Y_{input} - Y_{ref}$$

La RMSE a également été calculée.

La RMSE (Root Mean Square Error) se calcule de la façon suivante :

$$RMS_Error = \sqrt{\frac{\sum_{i=1}^N Error_i^2}{N}}$$

Elle combine l'inexactitude et l'imprécision puisqu'elle calcule une dispersion de l'erreur par rapport au point (0,0). Si l'inexactitude est négligeable, la RMSE est égale à l'écart-type de l'erreur. Etant donné qu'il n'y a pas indépendance entre e_x et e_y , nous avons calculé séparément la RMSE selon x et selon y.

Le tableau 9 synthétise l'ensemble des erreurs calculées.

Tableau 9. Synthèse des erreurs calculées.

Images comparées	SPOT Niveau 3 et SPOT Niveau 3 + DEM		SPOT Niveau 3 et SPOT Niveau 1B (*)		SPOT Niveau 3 et Orthos_DGA (Zone 2 Ourthes)		SPOT Niveau 3 et Orthos_DGA (Zone Martelange)		SPOT Niveau 3 et Orthos_DGA (Zone Semois)		SPOT Niveau 3 et Orthos_DGA (3 zones)	
	e_x	e_y	e_x	e_y	e_x	e_y	e_x	e_y	e_x	e_y	e_x	e_y
Nombre de valeurs	2000	2000	2006	2006	100	100	101	101	100	100	301	301
Moyenne	-2.71	0.71	-48.03	406.94	1.50	-1.47	-1.89	0.46	-1.43	5.78	-0.61	1.59
Ecart-type	4.90	3.29	36.46	12.17	3.73	4.55	4.62	4.68	4.36	4.45	4.50	5.49
RMS Error	5.60	3.36	60.29	407.12	4.00	4.76	4.97	4.68	4.57	7.28	4.53	5.70

(*) : L'image de niveau 1B a d'abord été ré-échantillonnée en coordonnées Lambert, puis a subi une translation avant de générer les points de certification (cf. paragraphe 2.2.1.3). Les statistiques de l'erreur reprises dans ce tableau concernent les vecteurs-erreurs réels, en tenant compte de cette translation.

2.2.1.2. Comparaison des images SPOT de niveau 3 et niveau 3 + DEM

Les composantes e_x et e_y des vecteurs-erreurs ont été calculées à partir de 2000 points de certification générés par la méthode automatique expliquée au paragraphe 2.2.1.1.

a) Analyse des paramètres statistiques généraux de l'erreur

La moyenne et la variance de l'erreur nous permettent de caractériser celle-ci. Pour cela, nous réalisons un graphique mettant en relation e_x et e_y (Figure 19). Le centre de gravité du nuage de points nous indique l'erreur moyenne (biais ou inexactitude), et nous traçons une ellipse de dispersion en fonction des variances et de la covariance, après avoir pris soin d'éliminer les points qui ont un vecteur-erreur anormalement grand, qui ont été mal positionnés par la méthode automatique.

L'inexactitude (biais) est le décalage moyen selon e_x et e_y par rapport au point (0,0) et l'imprécision est une mesure de la dispersion, qui dépend des variances et covariance des erreurs. Pour la comparaison des images niveau 3 et niveau 3 + DEM, on voit que l'imprécision est non significative en y (moins de 10 m, donc moins d'un pixel) et est faible en x (moins de 2 pixels). L'inexactitude est négligeable selon x et y. Le diagramme de dispersion montre par ailleurs qu'il n'y a pas indépendance des 2 composantes de l'erreur.

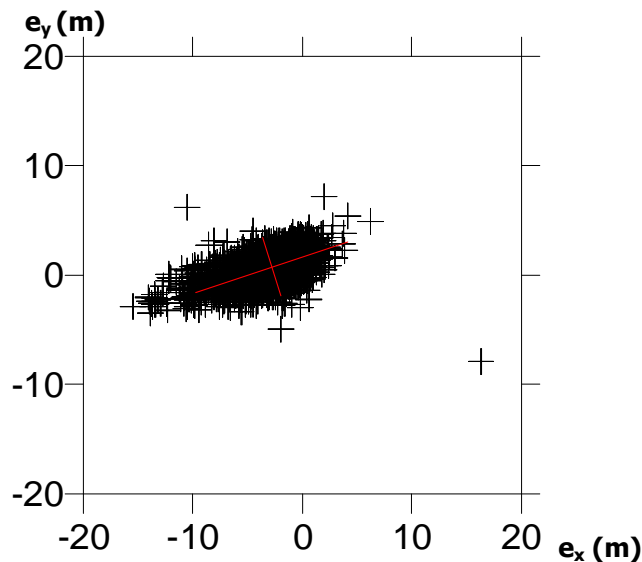


Figure 19. Graphique de e_x par rapport à e_y pour la comparaison des images niveau 3 et niveau 3 + DEM. Les axes de l'ellipse de dispersion sont tracés en rouge et le centre de gravité est l'intersection de ces axes.

Nous considérons donc que la différence géométrique entre ces deux images est négligeable vis-à-vis de leur résolution.

b) Analyse de la variation spatiale de l'erreur

Le but de cette analyse est de comprendre comment l'erreur varie dans l'espace. A partir des 2000 points de certification générés par la méthode automatique expliquée au point 2.2.1.1, nous obtenons 2000 vecteurs-erreurs. Nous calculons alors, à partir de ces 2000 vecteurs-erreurs, la valeur de l'erreur en chaque point, par une méthode d'interpolation par krigeage.

Nous avons étudié individuellement la continuité spatiale des 2 composantes du vecteur-erreur, par la construction de semi-variogrammes (Figure 20). Il ressort de cette analyse que les composantes de l'erreur varient de façon continue. Nous pouvons donc calculer les semi-variogrammes modélisés. Ces modèles théoriques nous informent sur le poids que l'on donnera aux points au voisinage du point considéré dans l'interpolation par krigeage. Nous réalisons alors l'interpolation par krigeage.

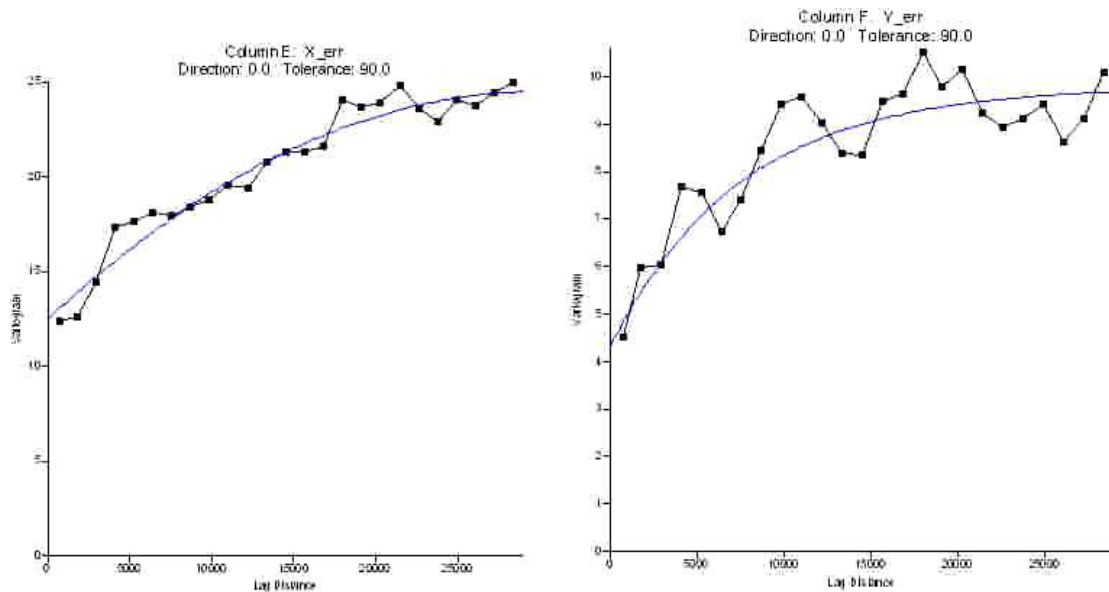


Figure 20. Semi-variogrammes expérimentaux (courbe en noir) et semi-variogrammes modélisés leur correspondant (courbe en bleu), pour les composantes e_x (à gauche) et e_y (à droite) des vecteurs-erreurs entre les images de niveau 3 et niveau 3 + DEM.

En observant les résultats de l'interpolation obtenus à la figure 21, nous remarquons, qu'à part le long de la frontière, où l'interpolation est faussée par le manque de données au voisinage des points de certification, que l'erreur est toujours plus petite que 20 m, c'est-à-dire 2 pixels. On peut donc considérer que les images niveau 3 et niveau 3 + DEM sont semblables du point de vue de la géométrie. La direction de l'erreur est toujours dirigée vers l'ouest, mais cette direction est peu significative lorsque l'intensité des vecteurs est faible.

c) Conclusion

Nous estimons donc que l'image niveau 3 + DEM n'apporte rien de plus que l'image niveau 3. Nous prendrons donc dans la suite de cette analyse l'image niveau 3 comme image de référence pour les comparaisons.

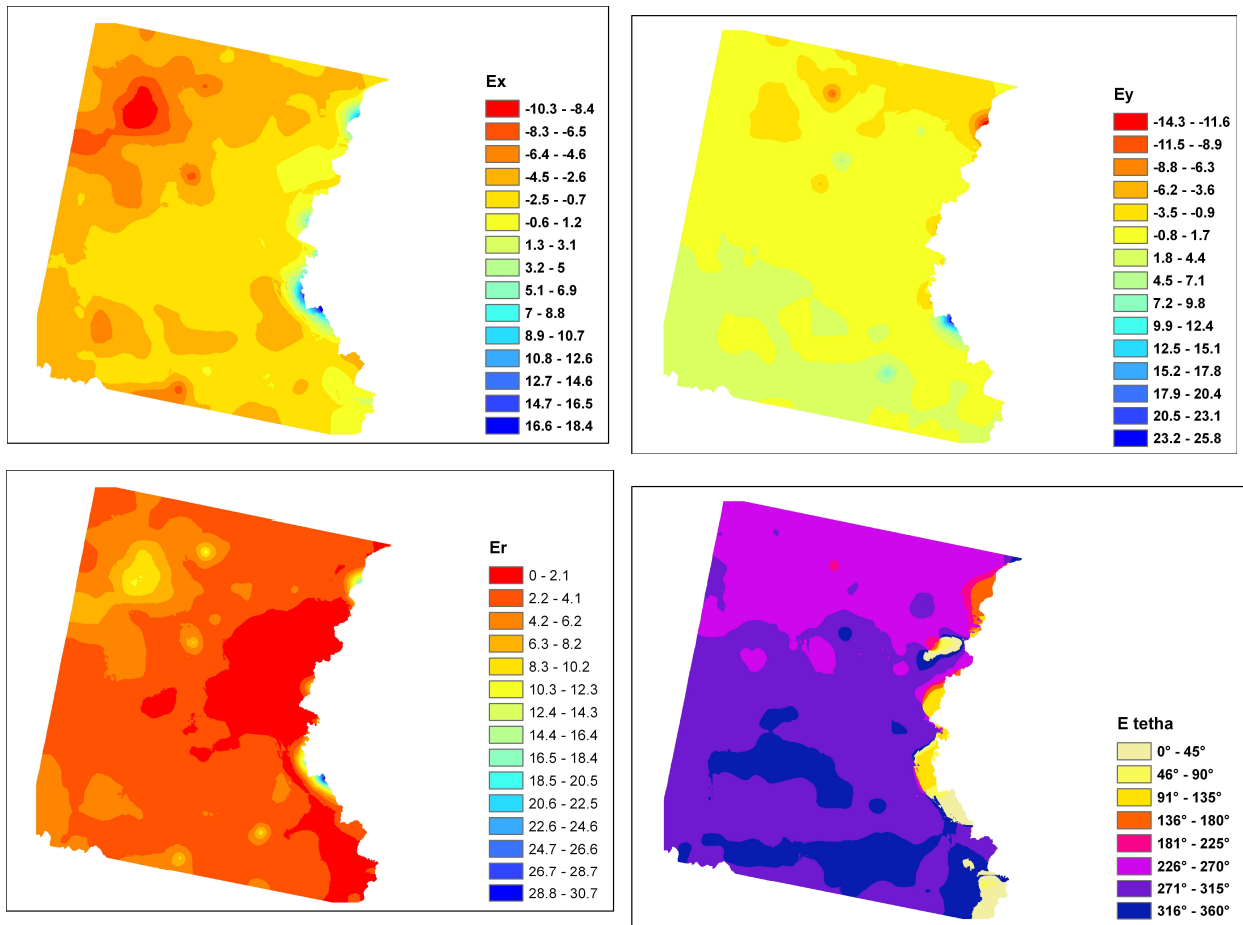


Figure 21. Résultat de l'interpolation des vecteurs-erreurs à partir des 2000 points de certification pour la comparaison des images de niveau 3 et niveau 3 + DEM. Composantes e_x et e_y des vecteurs-erreurs (en haut), et composantes e_r et e_θ , en coordonnées polaires de ces mêmes vecteurs (en bas).

2.2.1.3. Comparaison des images SPOT de niveau 3 et niveau 1B

Nous comparons à présent deux images semblables mais qui ont des aspects très différents puisque l'image niveau 1B n'a subi aucune correction géométrique par le producteur. L'image de niveau 1B a d'abord du être ré-échantillonnée en coordonnées Lambert, notre système de référence, en exploitant les coordonnées géographiques des 4 coins et du centre de l'image fournis par SPOT Image. Nous avons ensuite appliqué à l'image une translation, calculée à partir de 3 points de contrôle définis manuellement. Une fois cette translation réalisée, nous avons pu utiliser la méthode automatique expliquée au point 2.2.1.1 et générer ainsi 2000 points de certification. Les composantes des vecteurs-erreurs e_x et e_y ont ensuite été calculées à partir de ces points de certification et de la translation réalisée.

a) Analyse des paramètres statistiques généraux de l'erreur

De la même façon qu'au point 2.2.1.2 a), nous avons représenté e_x par rapport à e_y (Figure 22).

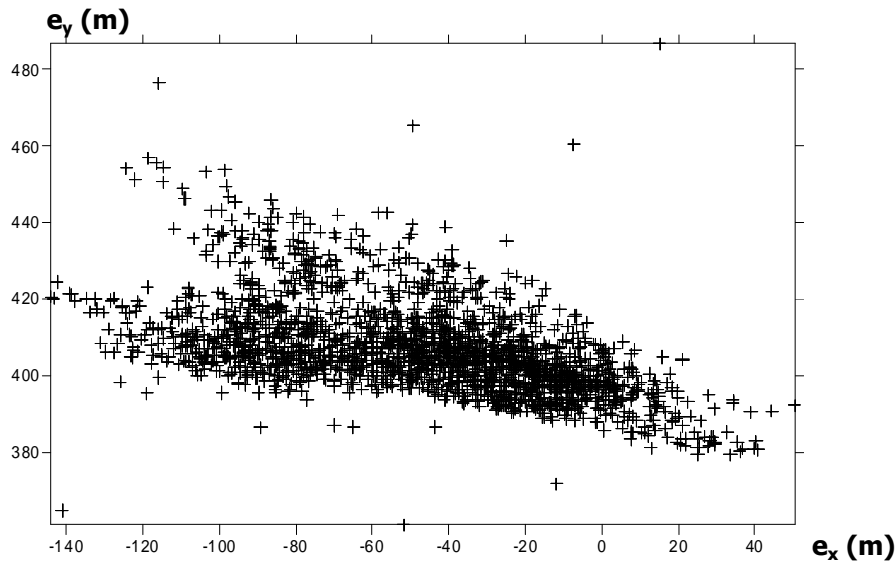


Figure 22. Graphique d' e_x par rapport à e_y pour la comparaison des images de niveau 1B et niveau 3.

L'imprécision est assez forte en y et très forte en x. L'inexactitude est très forte en y et assez forte en x. Cette fois, il est impossible de tracer une ellipse de dispersion car il y a plusieurs populations de points. Ces populations sont difficiles à identifier, mais, simplement en en définissant grossièrement quatre, on voit qu'elles sont regroupées dans l'espace (Figure 23). L'erreur dépend donc de la position des points.

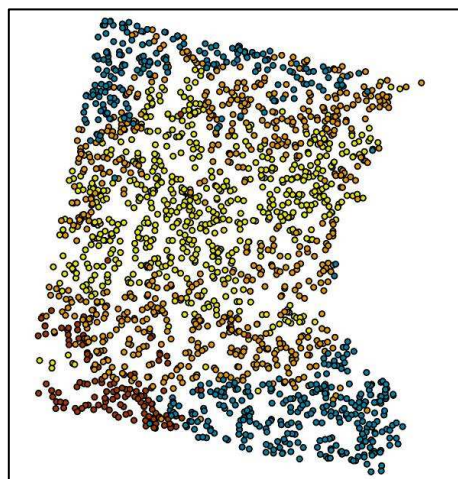
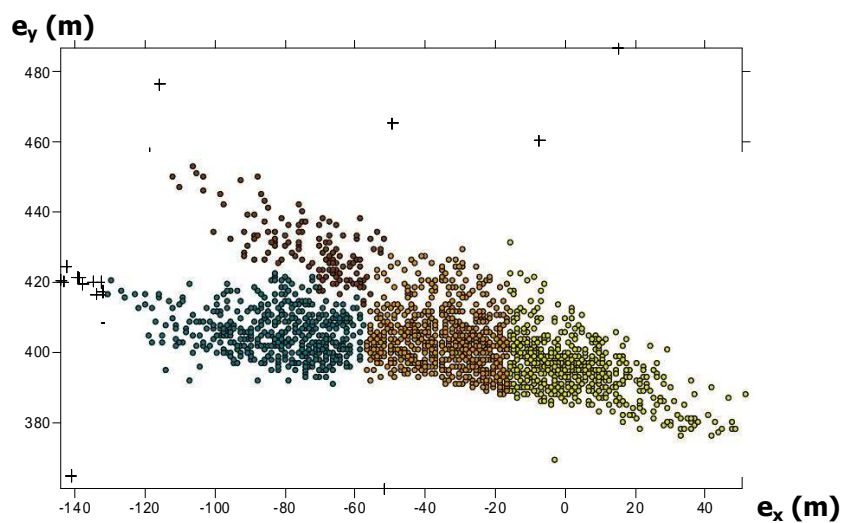


Figure 23. Définition grossière de 4 populations (en différentes couleurs) dans le nuage de points e_x par rapport à e_y (en haut), et visualisation de la répartition de ces points dans l'espace (en bas). On voit que les vecteurs-erreurs ont des statistiques différentes selon la position des points dans l'espace.

b) Analyse de la variation spatiale de l'erreur

Pour cette analyse, nous avons à nouveau utilisé la méthode d'interpolation par krigeage (cf. point 2.2.1.2 b)). Les figures 24 et 25 reprennent les résultats des semi-variogrammes et de l'interpolation par krigeage.

L'erreur est importante et varie dans l'espace. Elle est toujours dirigée vers le nord.

Conclusion

L'image de niveau 1B ne convient pas directement pour l'utilisation en complément de l'ortho-image. L'orthorectification standard proposée par SPOT Image est donc indispensable.

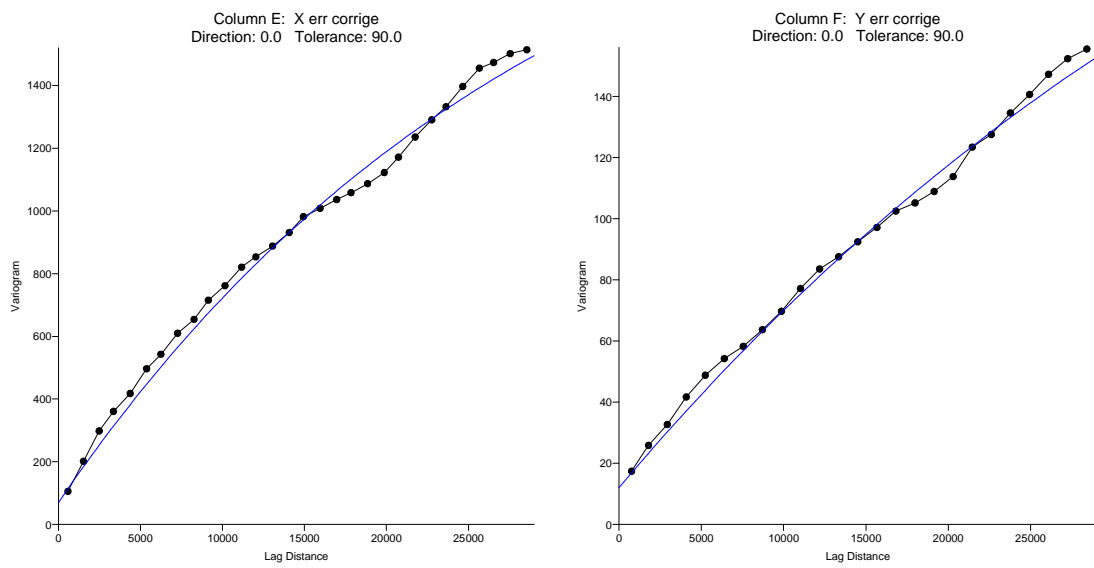


Figure 24. Semi-variogrammes expérimentaux (courbe en noir) et semi-variogrammes modèles leur correspondant (courbe en bleu), pour les composantes e_x (à gauche) et e_y (à droite) des vecteurs-erreurs entre les images de niveau 1B et niveau 3.

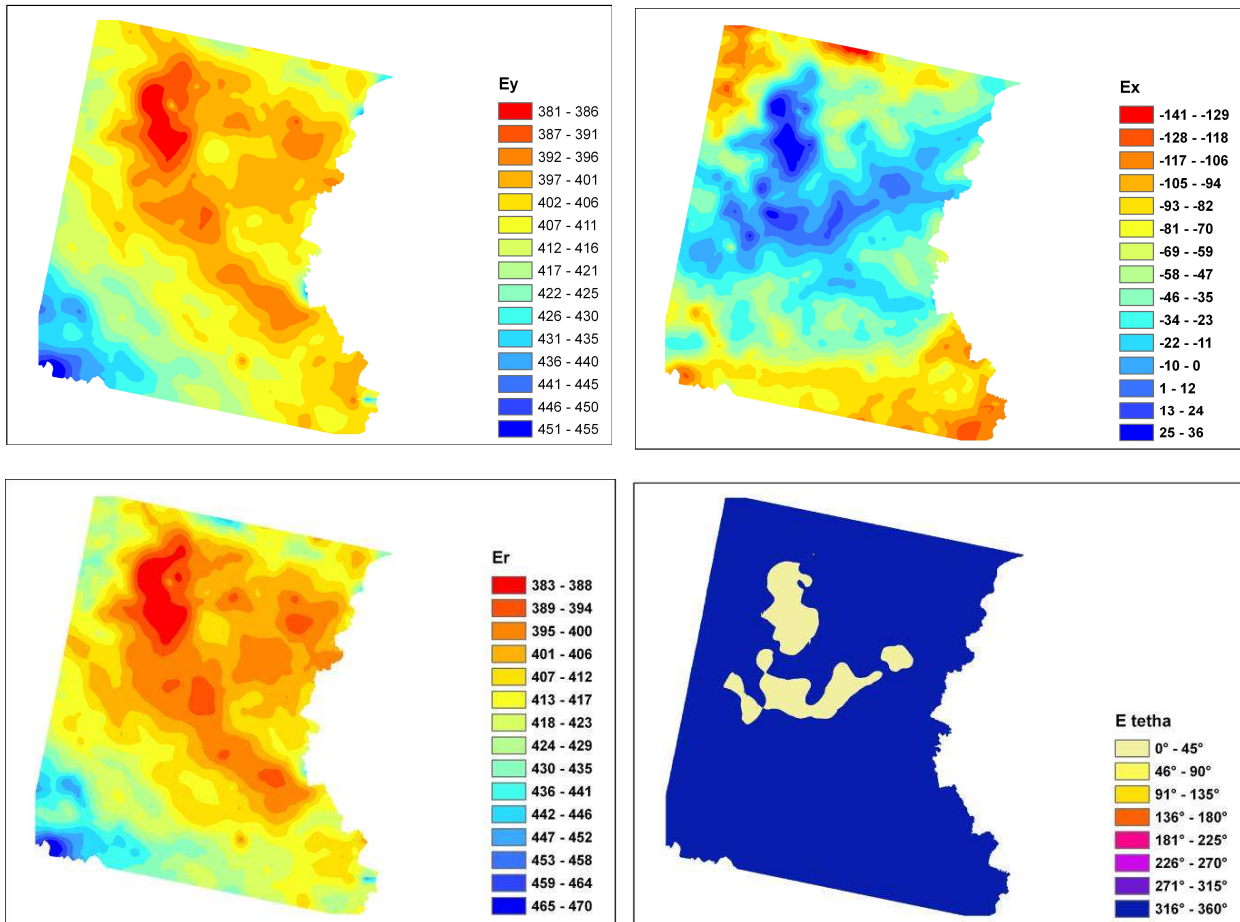


Figure 25. Résultat de l'interpolation des vecteurs-erreurs à partir des 2000 points de certification pour la comparaison des images de niveau 1B et niveau 3. Composantes e_x et e_y des vecteurs-erreurs (en haut), et composantes e_r et e_θ , en coordonnées polaires de ces mêmes vecteurs (en bas).

2.2.1.4. Comparaison de l'ortho-image et de l'image SPOT de niveau 3

a) Points de certification

L'image SPOT et l'ortho-image sont très différentes par leurs résolutions respectives de 10 m et 0,5 m. Elles sont donc difficilement comparables. La méthode automatique est inapplicable. Nous avons donc sélectionné 300 points de certification manuellement dans trois zones : la vallée de la Semois, la vallée des deux Ourthes et une zone de plateau près de Martelange (Figure 26). Ces trois zones ont été choisies car elles sont bien réparties dans la zone couverte par l'image SPOT, et aussi, pour les 2 zones de vallées, car la grande énergie du relief peut être un facteur influençant la géométrie de l'image. Dans chacune de ces zones, 100 points de certification ont été digitalisés, et ont servi au calcul des vecteurs-erreurs.

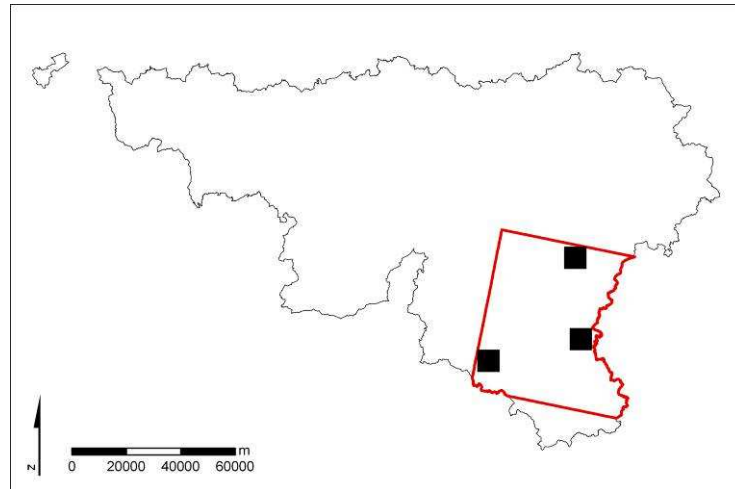


Figure 26. Localisation des 3 zones dans chacune desquelles 100 points de certification ont été digitalisés.

b) Analyse des paramètres statistiques généraux de l'erreur

La figure 27 est le graphique de dispersion des 2 composantes de l'erreur, e_x et e_y . De manière générale, l'imprécision est très faible (moins de 2 pixels SPOT). Il faut remarquer que nous sommes incapables en digitalisant des points de certification, de travailler à une résolution plus petite que le pixel de l'image SPOT (10 m). Vu la grande différence de résolution entre les deux images, il est d'ailleurs difficile de repérer un même point dans les deux images, et des erreurs d'un pixel peuvent être dues à l'imprécision de la digitalisation des points de certification. L'inexactitude est faible également, mais on voit tout de même qu'elle varie en fonction des 3 zones.

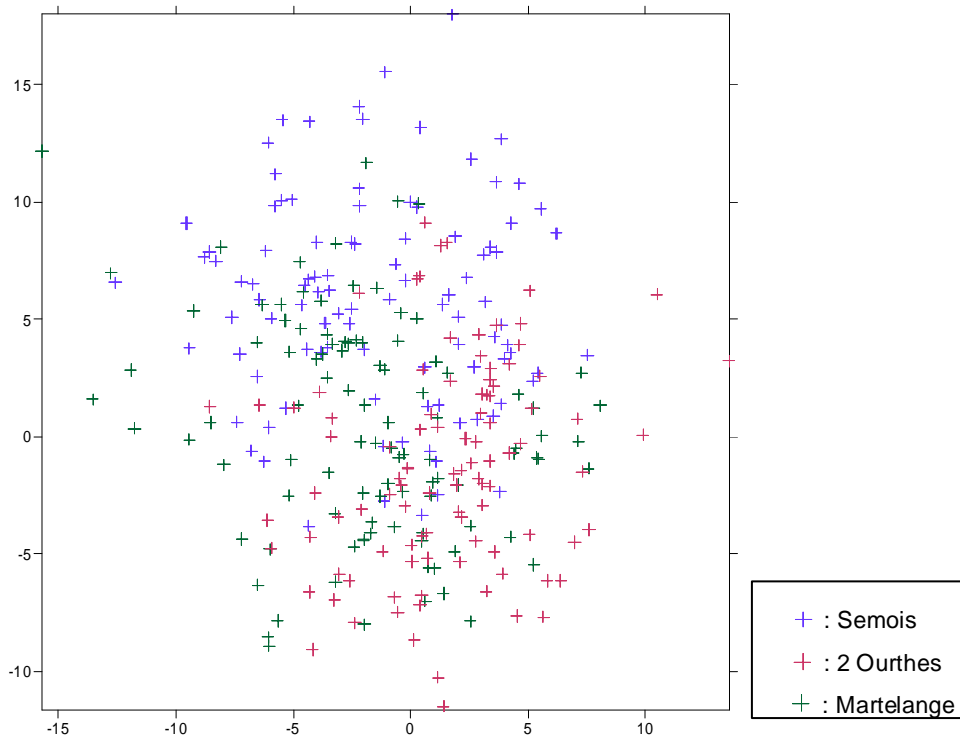


Figure 27. Graphique d' e_x par rapport à e_y pour la comparaison de l'ortho-image et de l'image SPOT de niveau 3. La couleur indique de quelle zone provient le point de certification.

c) Conclusion

A 10 m près, l'ortho-image et l'image SPOT de niveau 3 sont similaires du point de vue de la géométrie. Les méthodes utilisées ne nous permettent pas de valider cette cohérence géométrique à une résolution plus fine que 10 m.

2.2.1.5. Conclusion générale

L'image SPOT orthorectifiée de niveau 3 est celle qui convient le mieux pour être exploitée en combinaison avec la couche ORTHOS_DGA. Il est inutile de fournir des points de contrôle supplémentaires ni un modèle numérique de terrain pour cette orthorectification. La possibilité de sous-traiter l'orthorectification avait été évoquée, mais, vu la faible différence de prix entre les images, cette solution ne nous semble pas justifiée.

2.2.2. Considérations techniques relatives à l'utilisation d'une information diachronique dans les procédures de classification

Le présent paragraphe aborde différents aspects techniques liés à l'utilisation d'images diachroniques dans les procédures de classification.

L'apport effectif des données diachroniques disponibles dans le cadre de cette étude sera évalué lors des essais de classification proprement dits (2.7).

Pour cette tâche, nous sélectionnons dans le set d'images SPOT disponibles deux images couvrant la même zone, à des saisons différentes. La zone de recouvrement de ces **deux images**, enregistrées respectivement le **12/10/2005** et le **19/07/2006**, est représentée par le « Bloc Spot » de la figure 28. La seconde de ces images est quasiment synchrone des vols réalisés pour produire l'ORTHOS_DGA de la région (Blocs 1 et 3). Dans la suite du document ces images sont respectivement nommées 03 et 05 car la première a été enregistrée par le capteur HGR2 de SPOT 5 et la seconde par le capteur HGR 1 de SPOT 5.

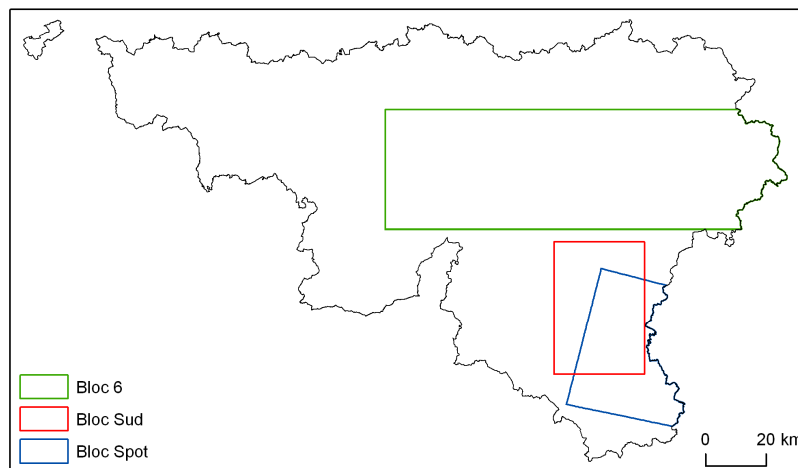


Figure 28. Localisation du bloc correspondant au recouvrement des deux images SPOT utilisées
Les caractéristiques géométriques et moments de prise de vue des deux images sont fournis dans le tableau 10.

Tableau 10. Paramètres géométriques et moments de prise de vue des images SPOT5 traitées

	SPOT 03	SPOT 05
	19/07/2006 à 10:42:15	12/10/2005 à 10:26:47
Jour julien	200	285
d² (UA²)	0.968273315	1.004171
Capteur	HGR 2	HGR 1
Résolution (m)	10	10
Incident angle (DEG)	-4.188155	-24.725603
Viewing angle (DEG)	-3.714752	-21.727524
Sun azimuth angle (DEG)	151.862484	163.362258
Sun elevation (DEG)	59.034767	31.533647

2.2.2.1. Prétraitements

a) Calibration

Afin d'extraire de ces deux images une information diachronique valable, il peut être utile de procéder à leur **calibration** préalable puis à leur **normalisation** par calcul de **réflectance**. La comparaison des deux images ne peut se faire simplement en terme de luminance mesurée par deux capteurs différents à des moments différents, car celle-ci dépend des conditions d'éclairement au sol, c'est-à-dire des conditions atmosphériques, de la position du soleil, de la position du capteur, de la topographie locale, de l'occupation du sol (BRDF si l'hypothèse lambertienne n'est pas vérifiée) ... Selon le niveau de précision désiré, les modèles à mettre en œuvre prendront en considération l'ensemble de ces facteurs ou une partie seulement de façon simple ou plus ou moins complexe.

La première étape est la **calibration** ou calcul des valeurs de **luminance** à partir des codes numériques (**DN**) de l'image, sur base des métadonnées radiométriques de *gain* (a) et *d'offset* (b) fournies par SPOT Image (cf. tableau 11), sur base de la formule :

$$L = DN/a + b$$

L : luminance enregistrée par le capteur ($W.m^{-2}.sr^{-1}.\mu m^{-1}$)

DN : code numérique de l'image

a : gain ($W^{-1}.m^2.sr.\mu m$)

b : offset ($W.m^{-2}.sr^{-1}.\mu m^{-1}$)

Tableau 11. Paramètres utilisés pour la calibration et le calcul de la réflectance des images SPOT 5

		DN haze	gain	off set	Angle de vue (°)	Es (W.m ⁻² .µm ⁻¹)	d ²	ζ (°)
SPOT 12/10/2005	XS3 PIR	24	2,4098	0	-21,73	1047	1,004171	58,47
	XS2 R	42	3,8067	0		1575		
	XS1 G	69	3,1438	0		1858		
	XS4 MIR	17	10,8896	0		234		
SPOT 19/07/2006	XS3 PIR	17	1,0795	0	-3,71	1043	0,968273	30,97
	XS2 R	32	2,2278	0		1573		
	XS1 G	48	1,6577	0		1858		
	XS4 MIR	28	10,6395	0		236		

b) Réflectance

Le calcul de la **réflectance** est inspiré de Chavez (1996) et exploite la formule suivante basée sur l'hypothèse lambertienne :

$$REF = \frac{[\pi * (L_{sat} - L_{haze})]}{[TAU_v * (E_0 * \cos(\zeta) * TAU_z + E_{down})]}$$

REF : réflectance spectrale de la surface

L_{sat} : luminance exo-atmosphérique enregistrée par le capteur (W.m⁻².sr⁻¹.µm⁻¹)

L_{haze} : contribution de l'atmosphère à la luminance enregistrée par le capteur (W.m⁻².sr⁻¹.µm⁻¹)

TAU_v : transmittance de l'atmosphère, du sol vers le capteur

E₀ : irradiance solaire exo-atmosphérique sur une surface perpendiculaire au flux lumineux (W.m⁻².µm⁻¹)

ζ : angle d'incidence du flux solaire direct sur la surface de la terre (angle zénithal solaire)

TAU_z : transmittance atmosphérique le long du trajet du rayonnement solaire de l'entrée dans l'atmosphère à la surface terrestre

E_{down} : éclairage indirect de la surface de la terre due à la diffusion du flux solaire par l'atmosphère (W.m⁻².µm⁻¹)

Nous avons d'abord estimé la luminance de l'atmosphère, **L_{haze}(DN_{haze})**, en utilisant la méthode de l'objet sombre (*Dark Object Subtraction model*) et en isolant quelques zones relativement profondes du lac de la Haute Sûre, présent dans la zone de recouvrement des images. Via les statistiques descriptives de ces surfaces, une valeur minimum a été extraite dans chaque bande. Cependant, pour les bandes verte et rouge de l'une des deux images Spot, nous avons constaté qu'un certain nombre de

pixels dans le reste de l'image (dans les bordures ombragées de certaines parcelles forestières, d'autres plus petits plans d'eau, ...) portaient des valeurs de luminance inférieures à la valeur du **Lhaze** estimée sur ce du lac. Ce nombre pouvait atteindre une proportion de l'histogramme allant jusqu'à 6 ou 7%. Les valeurs minima du lac de la Haute Sûre ne pouvaient donc constituer une bonne estimation de la contribution de l'atmosphère. Nous avons alors estimé le **Lhaze** en travaillant dans différents autres échantillons, et identifié des valeurs de luminance plus proches du minimum de l'image. Ces valeurs correspondent à des pixels compris dans divers plans d'eau ou autres zones très sombres, où l'on considère que le capteur ne perçoit que la luminance atmosphérique. Les valeurs du **DNhaze** finalement retenues pour calculer les **Lhaze** par application de la formule de calibration sont reprises dans le tableau 11.

Au dénominateur de la formule, la transmittance **TAU_v** est représentée par le cosinus de l'angle de vue du capteur fournis dans le tableau 10.

Spot Image fournit les valeurs **Es** de « luminances solaires équivalentes » pour les quatre bandes (Tableau 12). **Es** dépend de la gamme de longueurs d'onde de la bande spectrale traitée et de la courbe de sensibilité spectrale du capteur dans cette bande. Les valeurs de **Es** fournies dans le tableau 13 sont des valeurs moyennes pendant un cycle annuel obtenues sur base du spectre d'énergie solaire ASTM G173-03 (Reference Spectra Derived from SMARTS v. 2.9.2.) fourni à la figure 29. Elles doivent être divisées par le carré de la distance Terre-Soleil exprimée en Unité Astronomique et estimée par la formule proposée par Methner (1999) le jour julien de la prise de vue et fournie dans le tableau 10 pour d'obtenir une valeur de **E₀**.

Les valeurs de ζ est l'angle complémentaire de l'élévation solaire, également fournie dans les métadonnées des images Spot. **TAU_z** est défini égal au $\cos \zeta$ (*COST model*) considéré comme constant sur toute la scène traitée par rapport à une surface horizontale.

Les valeurs d'**Edown** ont été estimées sur base de données sur le rayonnement diffus et direct mesuré par l'IRM à Uccle entre 310 et 2800 nm toutes les 30 minutes (Figure 30). L'opération consiste à repartir de la valeur de rayonnement direct + diffus donnée par l'IRM dans la partie du spectre dans laquelle le capteur est sensible en fonction de la bande spectrale considérée, en pondérant la contribution de chaque intervalle de longueur d'onde à la fois par la valeur de la courbe du spectre solaire et par la valeur de la courbe de sensibilité spectrale de chaque bande Spot. Les valeurs sont obtenues d'une part avec la courbe bleue global de la figure 29 et d'autre part avec la courbe rouge. Les valeurs fournies par la seconde de ces courbes ont été utilisées pour le calcul des images de réflectance. Elles sont reprises dans le tableau 12.

Tableau 12. Luminances solaires équivalentes des différentes bandes spectrales des 2 capteurs de SPOT5

Luminances équivalentes solaires		
Bandes spectrales	SPOT 5	
	HRG 1	HRG 2
HMA/HMB	1762	1773
B1	1858	1858
B2	1573	1575
B3	1043	1047
MIR	236	234

ASTM G173-03 Reference Spectra

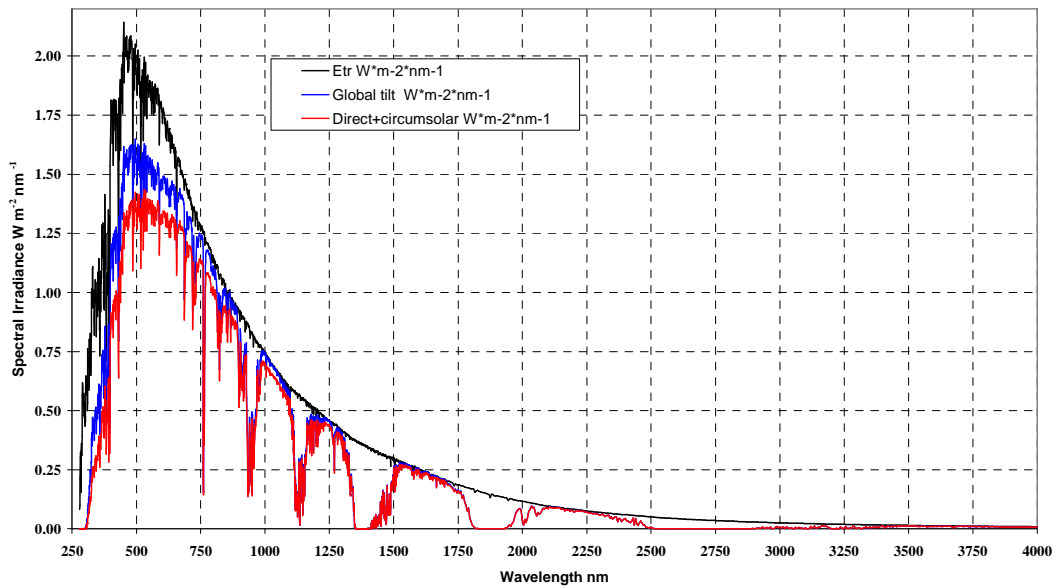


Figure 29. Irradiance solaire

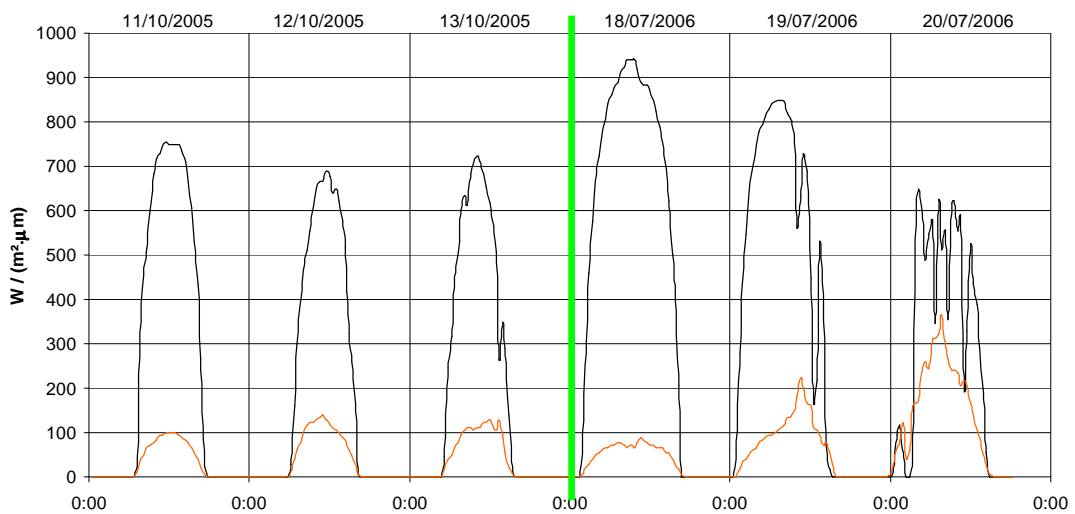


Figure 30. Rayonnement direct et diffus mesuré à Uccle du 11 au 13/10/2005 et du 18 au 20/7/2006

Tableau 13. Valeurs des rayonnements direct et diffus calculées pour les deux dates considérées

	SPOT 19/07/2006	SPOT 12/10/2005
Rayonnement direct Uccle (W/m²)	844	556
Rayonnement diffus Uccle (W/m²)	100	122

c) Calcul de réflectance sous Idrisi Kilimanjaro

Plusieurs tests de calcul de réflectance ont également été réalisés sous Idrisi Kilimanjaro à l'aide du module de correction radiométrique (AtmosC). Un modèle complet de correction (*full model*) a été exploité. Il permet notamment d'estimer les transmittances, TAUz et TAUv, en fonction de l'épaisseur optique de l'atmosphère. Cette dernière a été estimée à partir du produit *Aerosol Thickness* (Diffusion de Mie indépendante de la longueur d'onde) de MODIS (<http://neo.sci.gsfc.nasa.gov/Search.html>). Les valeurs maximales, moyennes et minimales ont été calculées dans une fenêtre carrée de 7*7 (résolution 10^{-1°}) autour du centre approximatif du bloc spot le jour de la prise de vue (Tableau 14). Ce voisinage a été utilisé pour prendre en compte la variabilité spatio-temporelle du paramètre extrait.

Tableau 14. Valeurs de l'épaisseur optique due aux aérosols de l'atmosphère extraites des données MODIS de la NASA

	min	max	moy
SPOT 19/07/2006	0.043	0.201	0.108
SPOT 12/10/2005	0.039	0.079	0.060

Remarquons cependant que l'épaisseur optique de l'atmosphère liée à la présence d'aérosol fournie par les produits MODIS ne constitue qu'une partie de l'épaisseur optique de l'atmosphère comme cela est rappelé ci-dessous

(http://modis-atmos.gsfc.nasa.gov/docs/ATBD_MOD04_C005_rev2.pdf):

*“Optical thickness is the negative natural logarithm of the proportional transmittance of the atmosphere. Thus a perfectly transparent atmosphere, with a transmittance of 1.00 has an optical thickness of 0.00. Optical thickness varies as primarily a function of Rayleigh scattering, aerosols, moisture and ozone content. **Total optical thickness can be reasonably estimated as the sum of these major elements.** Forster (1984) provides perhaps the best guidance on estimating these values. However, briefly, **optical thickness due to Rayleigh scattering** ranges from about 0.22 in the blue wavelengths, to 0.09 in the green, 0.05 in the red and 0.01 in the near-infrared. These values apply to a standard pressure of 1013 mbar at sea level. To adjust for standard pressure at a higher altitude, multiply the optical thickness by the ratio of the pressure at the new altitude by that at sea level. Aerosols are the primary determinants of haze. **Aerosol optical thickness** can be estimated from visibility estimates (see Turner and Spencer, 1972), with values ranging from (very approximately) 0.02 at 110 km to 0.2 at 30 km to 1.0 at 3 km. However, aerosol optical thickness declines exponentially with altitude. To adjust, multiply that determined from visibility by $\exp(-h/1.2)$ where the height, h, is expressed in kilometers. The **optical thickness of water vapor** can be ignored in the visible range. However, for longer wavelengths it should be accounted for, and can be estimated using a measure of relative humidity. Forster (1984) gives a useful table. However, at 20 degrees centigrade, optical thickness of water vapor very approximately ranges from 0.04 at 20% RH (very dry) to 0.13 at 95% RH (very humid) in the near-infrared. **Optical thickness of Ozone** is greatest between 0.5-0.7 μm where a value of 0.03 is typical.”*

Les valeurs de l'épaisseur optique des aérosols a donc été corrigée en lui ajoutant la valeur d'un terme correctif du à la diffusion de Rayleigh qui dépendant de la longueur d'onde (Tableau 15).

Tableau 15. Valeurs de la contribution de la diffusion de Rayleigh sur l'épaisseur optique de l'atmosphère (Idrisi Kilimanjaro help) extraites des données MODIS de la NASA

Bleu	0.22
Vert	0.09
Rouge	0.05
Proche IR	0.01

Nous avons planifié un protocole de tests de calcul de réflectance en utilisant les valeurs maximales et minimales de l'épaisseur optique et en faisant varier la valeur du **DNhaze** et du rayonnement diffus qui constituent deux autres facteurs incertains de la procédure. Nous n'avons réalisé les calculs que pour la bande 3 de l'image du 19/07/2006 sans faire varier le **DNhaze** et les rayonnements diffus et directs. On constate que dans ce cas de figure l'effet de l'épaisseur optique sur la réflectance est très marqué.

La grande variabilité des résultats pour un canal normalement relativement peu affecté par les effets atmosphériques rend difficile la comparaison des signatures spectrales d'une date à l'autre ce qui justifie, dans l'état actuel de l'analyse, l'adoption d'autres approches préalables aux classifications telles que la **normalisation radiométrique relative** sur base de *Pseudo-Invariant Features (PIF)*, par exemple, qui est présentée par la suite.

Tableau 16. Tests de calcul de réflectance en utilisant les valeurs maximales et minimales de l'épaisseur optiques pour la bande 3 de l'image du 19/07/2006

ATMOSC	FULL Model	FULL Model
DN haze	48	48
OT	0.13	0.29
R direct (donné)	169.77	169.77
R diffus	Premier centile	Premier centile
REF. max	0.351	0.495
REF. mode 1	0.018	0.026
REF. mode 2	0.049	0.069
REF. moy	0.04	0.056
REF. SD	0.022	0.031

d) Normalisation topographique

Nous avons aussi évalué la nécessité de corriger de **l'effet topographique**, étant donné que les signatures spectrales sont influencées par l'exposition aux rayons du soleil, qui dépend de la date et l'heure et varie en fonction de la topographie. La normalisation topographique exploite la méthode du *hillshading* (estompage) et nécessite un MNS/MNT. Le MNT utilisé est le MNT IGN rééchantillonné à 10 m. L'application du *hillshading* permet le calcul de cosinus de l'angle entre le vecteur normal à la surface topographique et le vecteur pointant vers le soleil. Ce cosinus est ensuite mis en relation avec la réflectance ($DN_{haze}=17$; Epaisseur optique = min de l'épaisseur optique des aérosols; E_{down} = premier centile de l'image) par régression entre images (Figure 31).

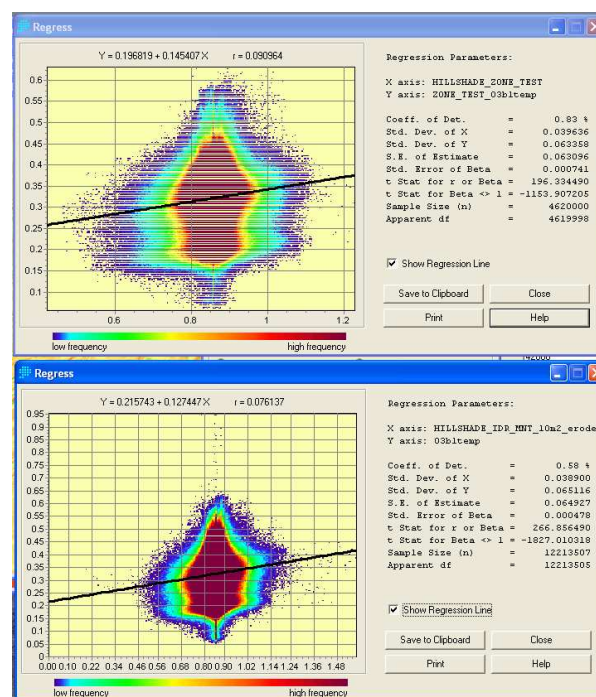


Figure 31. Relation entre réflectance de la bande 1 (XS3 - PIR) du 12/10/2005 sur toute l'image (en bas) et sur la zone commune entre les deux images (en haut).

Pour la bande spectrale PIR enregistrée le 12/10/2005, la plus influencée par le relief et sur la zone commune aux deux images (bloc spot), la régression n'est pas significative. Les valeurs de cosinus calculé par *hillshading* sont presque exclusivement comprises entre 0.6 et 1. Pour cette image d'automne, le relief n'est pas suffisamment marqué dans la zone du bloc SPOT pour qu'une normalisation topographique se justifie réellement. L'effet de la variation de réflectance lié à la variation de la couverture du sol est bien plus important que celui du à la variation de l'angle d'incidence des rayons solaires sur la surface topographique. Les procédures de classification doivent cependant prendre en considération les possibles effets de site sur la définition des classes et sous-classes d'occupation du sol.

e) Normalisation radiométrique relative

Le but de la normalisation radiométrique relative est de rendre comparables les réflectances spectrales d'images saisies à des dates différentes. Cette méthode repose sur le concept des *Pseudo-Invariant Features*. Il s'agit d'objets au sol dont la dimension est suffisamment grande (\gg pixel) et dont la réflectance au sol est homogène dans l'espace et n'a pas varié dans le temps sur les images comparées.

Des zones homogènes présentant une réflectance stable ont été sélectionnées de façon à obtenir un échantillon de valeurs de réflectance représentative de toute la gamme d'intensités de l'image. La sélection a été réalisée visuellement en analysant les images et les diagrammes de dispersion des pixels de chaque image dans l'espace multispectral des 4 canaux.

Les PIF sélectionnés sont des polygones correspondant généralement à des toits, parkings, routes, points d'eau ... Douze polygones ont été digitalisés mais cinq d'entre eux présentaient une trop grande hétérogénéité radiométrique surtout dans le canal PIR. Ils ont été supprimés.

Les relations entre les valeurs des deux dates dans chaque canal ont été formulées sur base des médianes extraites dans chaque zone (Figure 32). Les droites de régression obtenues pour chaque canal permettent de convertir les réflectances d'une image par rapport à une autre choisie comme référence et qui est généralement celle présentant la plus grande dynamique.

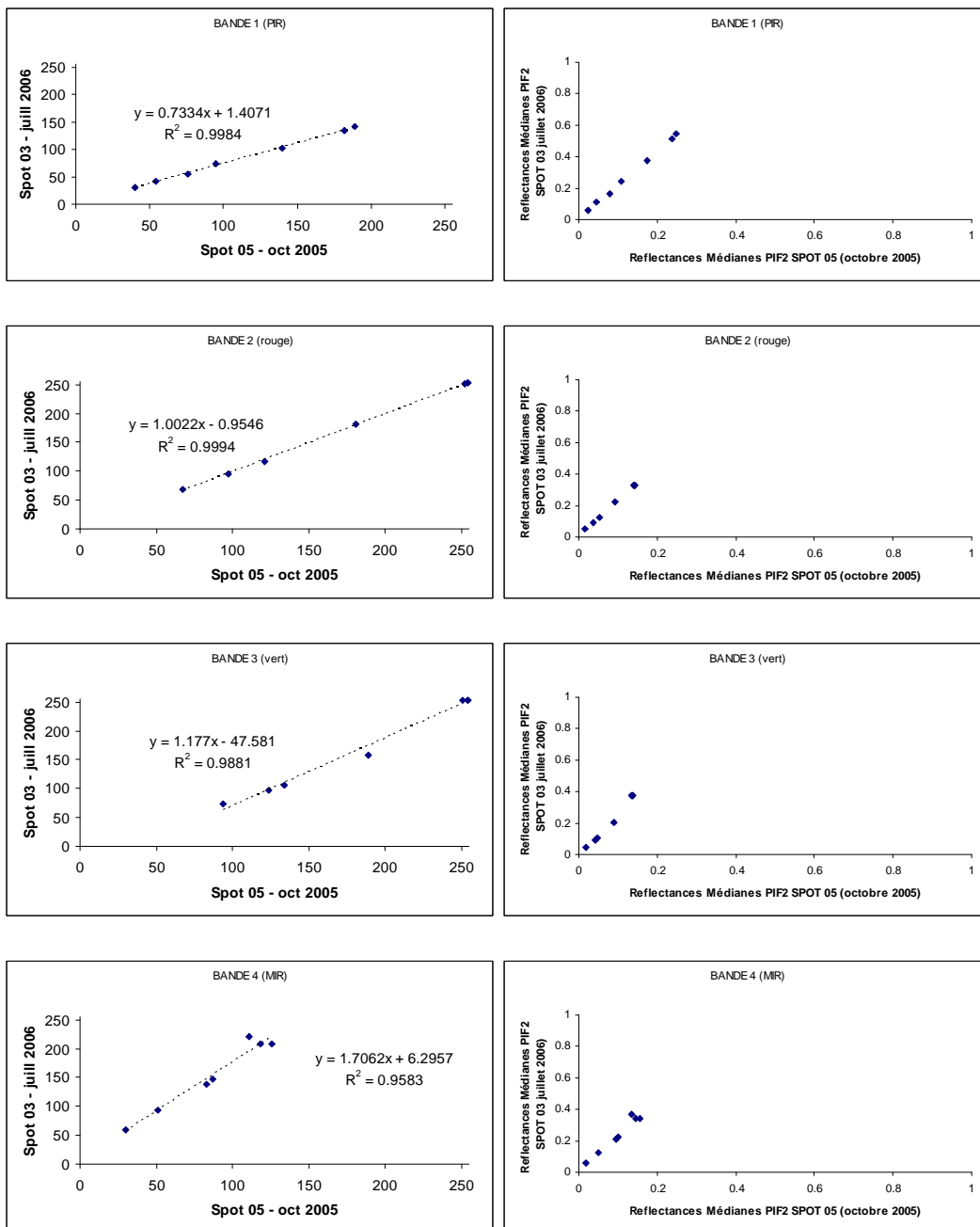


Figure 32. Droites de régression entre les valeurs de réflectance médianes dans chaque bande spectrale des PIF sélectionnés sur les deux images à normaliser. Les réflectances des graphiques de droite ont été calculées à partir des DN correspondant des graphiques de gauche en utilisant les paramètres du Tableau 17.

Tableau 17. Paramètres de calcul de la réflectance utilisés pour la normalisation radiométrique relative

	SPOT	SPOT	BANDE 2	SPOT	SPOT
BANDE 1 (PIR)	12/10/2005	19/07/2006	(rouge)	12/10/2005	19/07/2006
Gain	2.410	1.080	Gain	3.807	2.228
DN haze	24	17	DN haze	42	32
Edown	928.8	664.0	Edown	1323.8	948.6
TAUv	0.379	0.065	TAUv	0.379	0.065
	SPOT	SPOT	BANDE 4	SPOT	SPOT
BANDE 3 (vert)	12/10/2005	19/07/2006	(MIR)	12/10/2005	19/07/2006
Gain	3.144	1.658	Gain	10.890	10.640
DN haze	69	48	DN haze	17	28
Edown	1441.7	1035.5	Edown	215.9	155.9
TAUv	0.379	0.065	TAUv	0.379	0.065

La droite de régression calculée a ensuite été utilisée pour essayer de retrouver d'autres PIF en masquant dans l'image tous les pixels qui sortaient de l'intervalle symétrique suivant :

$$[a * X + b + \frac{1}{2} \text{intervalle} ; a * X + b - \frac{1}{2} \text{intervalle}]$$

avec

$Y = a X + b$: l'équation de la droite donnant la valeur d'un PIF dans 03 (Y) en fonction de sa valeur dans l'image 05 (X).

Cependant, les masques binaires obtenus ne sont pas vraiment exploitables car les pixels sont éparpillés dans différentes affectations, routes, prairies, forêts ...

Sur les graphiques de la figure 32, on constate que, dans chaque bande, les points sont alignés sur des droites, mais les réflectances médianes des PIF dans l'image 03 sont pratiquement le double de celles de l'image 05. Cette constatation remet en question la procédure de calcul de réflectance utilisée pour construire ces graphiques. Il s'agit d'un modèle simple. Il conviendrait de tenter l'expérience avec les autres valeurs de paramètres critiques (Edown, Epaisseur optique, DNhaze ...) proposés dans le test suggéré plus haut. La raison de cette différence peut résider dans une sous-estimation des valeurs de l'éclairement direct et/ou diffus en raison d'une surestimation de l'épaisseur optique ou d'une sous-estimation des valeurs de Es et Edown.

2.2.2.2. Classifications et analyse diachronique

a) Analyse en composantes principales

Une analyse en composantes principales (ACP) a été appliquée de façon indépendante à chacune des deux images 05 et 03 sur leur zone commune. L'aspect diachronique n'a donc pas été pris en compte à ce niveau de traitement. Vu que les transformations appliquées aux données sont des transformations linéaires et que les images sont traitées séparément, on peut traiter les données de départ, les DN codés sur 1 octet. Il n'est pas nécessaire d'utiliser les images de luminance ou de réflectance. L'ACP a été réalisée sur les 4 canaux (PIR, R, V, MIR) sous IMAGINE en utilisant une AOI (*Area Of Interest*) permettant d'éliminer l'effet des valeurs nulles représentant près de 50% de l'image traitée. Le résultat a été soumis à une procédure de classification basée sur l'histogramme trivarié (composante 1, composante 2, composante 3). Ces 3 composantes expliquent 99,7% et 99,8 % de la variance totale respectivement de l'image de 2005 et de 2006.

b) RGB clustering

L'objectif de cette procédure est d'analyser le pouvoir discriminant dans ce néo-espace spectral après avoir éliminé les zones de valeur nulle sur les images de départ. Sur base de plusieurs tests de sensibilité de la méthode à la segmentation de chaque composante en un certain nombre de niveaux lors de la construction de l'histogramme trivarié (Figure 33), il semble que ce nombre doive être proportionnel à la racine de sa valeur propre pour accorder un poids majeur aux composantes présentant la plus grande dynamique dans la procédure de classification.

Ces tests font également ressortir que les combinaisons 32/13/7 (CP1/CP2/CP3) pour l'image de 2006 et 32/17/7 pour celle de 2005 semblent judicieuses lors de l'utilisation de la fonction avancée de clustering RVB. Mais, même en retenant un nombre de classes importants (73 ou 256), le traitement de l'image de 2005 qui présente la plus faible dynamique, donne une image peu contrastée avec une grande perte d'information pour les classes les plus marginales (dans le jaune, le rouge et certains verts de la composition en fausse couleur de synthèse R:CP1, V:CP2, B:CP3) alors que la composition colorée de départ était très bien contrastée. Les tests effectués à l'aide de la fonction simple de clustering RVB qui ne permet pas d'imposer un nombre minimum de pixels par cluster produit cependant des résultats plus satisfaisants.

La procédure de classification basée sur la technique du clustering RVB n'a cependant pas été poursuivie (interprétation a posteriori par comparaison avec le SIGEC, regroupement de classes, post-classification et validation date par date, comparaison des classifications aux deux dates et analyse de la cohérence des trajectoires temporelles ...).

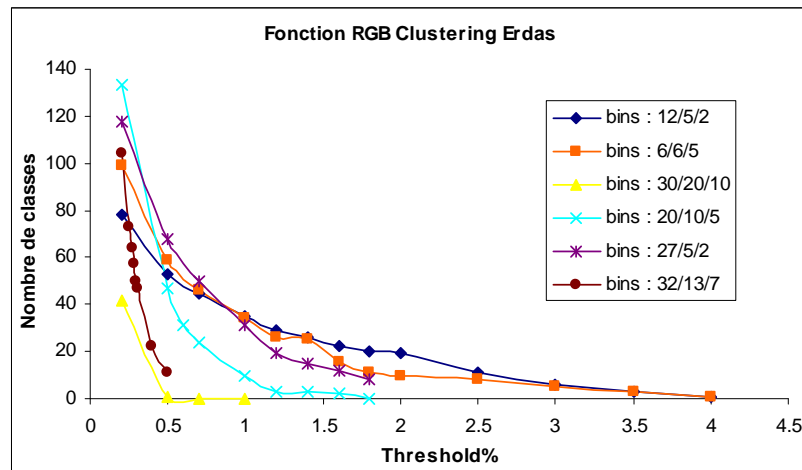


Figure 33. Sensibilité à la façon de construire l'histogramme multivarié de la méthode du clustering RVB basée ce dernier image de 2006 – le *threshold* correspond au pourcentage minimum de pixels que doit compter un cluster (signification d'un mode de histogramme trivarié) par rapport au nombre de pixels de l'image.

c) Stratégie de classification par nuée dynamique (Isodata – k-means)

La stratégie de classification non-supervisée par nuée dynamique en 74 classes dans l'espace des 3 premières composantes principales extraites, suivie d'une interprétation et d'un regroupement des classes, d'une extraction de signature et d'une classification par maximum de vraisemblance a été préférée à la classification par clustering RVB basé sur l'histogramme multivarié dans l'espace des 3 premières composantes de l'ACP. L'interprétation et le regroupement des classes avant d'extraire les signatures ont été effectués à l'aide de la composition en fausse couleur infra-rouge de la date considérée, de l'ortho-image aérienne de 2006 et de la consultation du SIGEC 2006. Une analyse de la matrice de séparabilité (divergence*, page suivante) entre classes calculée à l'aide des signatures extraites a aussi été prise en compte.

* La divergence est une mesure de la distance statistique standardisée au centre des classes (vecteur des moyennes) à l'aide de la matrice variance-covariance. La formule de calcul de la divergence entre deux classes i et j est donnée par :

$$D_{ij} = \frac{1}{2} \text{tr}((C_i - C_j)(C_i^{-1} - C_j^{-1})) + \frac{1}{2} \text{tr}((C_i^{-1} - C_j^{-1})(\mu_i - \mu_j)(\mu_i - \mu_j)^T)$$

avec

i et j = les signatures comparées

C_i et C_j = les matrices des covariances des i et j

μ_i et μ_j = les vecteurs des moyennes des signatures i et j

tr = la fonction « trace de la matrice »

T = la fonction « transposée de la matrice »

La dernière étape de classification par maximum de vraisemblance sur base des signatures extraites de la classification obtenue n'a malheureusement pas été réalisée par manque de temps.

2.3. Production d'une couche de différence d'altitude entre le MNS_DGA et un MNT (Unité de Géomatique - GRFMN)

L'objectif de cette partie de l'étude est d'évaluer l'apport d'une couche de différence d'altitude entre le MNS_DGA et un MNT dans les procédures de classification. Cette couche raster fournira, en chaque pixel, une valeur correspondant à la hauteur du « sur-sol » : bâtiments, ouvrages d'art, peuplements forestiers et éléments arborés.

2.3.1. Analyse des MNT et production d'un MNH

Le MNH (modèle numérique de hauteur) est la couche de différence d'altitude entre le MNS_DGA et un MNT. Il fournira une information sur la hauteur des éléments apparaissant dans le paysage, dont les éléments végétaux. L'apport de cette information sera évalué dans les opérations de classification menées dans les analyses thématiques.

Les analyses des MNT et la production de la couche de différence d'altitude ont été réalisées pour le territoire du **bloc 6**.

2.3.1.1. Données numériques d'élévation disponibles

Les différentes données numériques d'élévation dont nous disposons sont les suivantes :

- le MNT/MNS laser sur les vallées des cours d'eau de première catégorie,
- les données altimétriques du PICC,
- le MNT IGN à 1:10 000e – Format image,
- les données altimétriques de l'IGN à 1:10 000e – Format vectoriel,
- le MNT ERRUISSOL,
- le MNS_DGA.

Ces différentes données d'élévation ont des **avantages** et des **inconvénients** (Tableau 18). Le MNT IGN couvre tout le territoire wallon, tandis que le PICC et le MNT/MNS laser n'en couvrent qu'une partie. Le MNT/MNS laser, généré par le 2^e écho de l'altimètre laser, est très précis mais n'existe que dans les fonds de certaines vallées.

Il représente l'altitude sans la végétation, mais ne donne cependant pas l'altitude du sol au sens strict, puisque les bâtiments et constructions y sont toujours présents. En effet, l'écho est arrêté par les surfaces qui lui sont opaques. Les points du PICC sont relativement précis mais sont quasi inexistantes dans les zones forestières. Leur densité est plus faible dans les zones rurales qu'en milieu urbain. De plus, le PICC n'a pas encore été produit pour l'entièreté du territoire.

Tableau 18. Caractéristiques des différents MNT analysés et du MNS

Caractéristiques	MNT IGN	Données altimétriques du PICC	MNT/MNS laser	MNS DGA
Précision en altitude	0,5 à 2 m	25 cm	11 cm (moyenne)	Max 2 m
Couverture du territoire wallon	100 %	65 % au 16/03/2007 (production en cours)	Très localisé, seulement les lits majeurs des cours d'eau*	100 %
Résolution spatiale	20 m	Données vectorielles à densité de données variable	1 m	5 m

* Cours d'eau navigables et non navigables de première catégorie

Il faut remarquer que le **MNT IGN** en format image a été conçu à partir des données vectorielles de l'IGN. Il s'agit de données de qualité variable sur le territoire wallon. Il s'agit soit de données de type « points et lignes structurelles », qui sont des données restituées par photogrammétrie, soit de courbes de niveau digitalisées, provenant des cartes IGN dessinées (Figure 34).

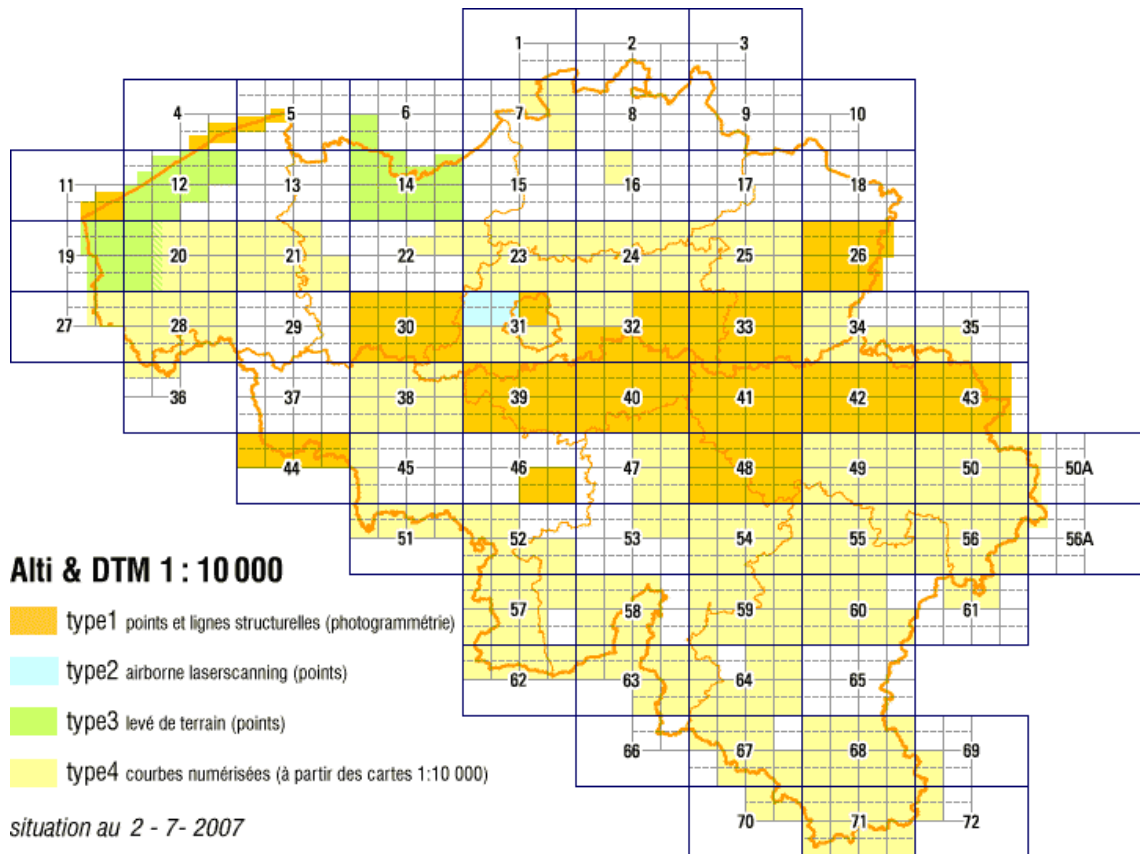


Figure 34. Types de données numériques d'altitude IGN à 1 :10 000e disponibles en Belgique

Le **MNT_ERRUISSOL** a été réalisé en combinant les données suivantes. Compte tenu de sa plus grande précision, c'est le **MNT laser** qui a été pris en compte là où il existait, dans une version filtrée où les bâtiments ont été éliminés. Pour les zones où il n'existait pas, les données du **PICC** ont été choisies, et les données **IGN** ont rempli les « trous » du **PICC** et les zones où ni **PICC** ni le **MNT laser** n'existaient.

Pour générer le **MNT_ERRUISSOL**, seuls certains types de points **PICC** ont été considérés (Tableau 19). Par contre, dans notre analyse de la qualité des **MNT** et pour la production d'un **TIN** (cf. paragraphe 2.3.1.3 et 2.3.1.4), nous avons considéré tous les points au sol du **PICC**, y compris les points du réseau de voirie (Tableau 20).

Tableau 19. Types de points PICC utilisés pour la génération du MNT Erruisol

Type STAR	Description
112	niveau haut des berges (rivière, fleuve, canal, ruisseau)
180	bas talus
182	haut talus
185	point de niveau de terrain
186	falaise pied
187	falaise haut
751	limites parcellaires

Tableau 20. Types de points PICC utilisés pour produire le TIN puis le MNT dans le cadre de ce projet.

Type STAR	Description
112	niveau haut des berges (rivière, fleuve, canal, ruisseau)
121	ruisseau (axe)
141	fossé (axe)
163	bord (lac, étang, piscine)
177	bord supérieur (carrière, sablière, excavations diverses)
178	bord inférieur (terril et dépôts permanents)
180	bas talus
182	haut talus
185	point de niveau de terrain
186	falaise pied
187	falaise haut
233	écluse, barrage, mur de quai (voie d'eau)
313	bord route, autoroute, rue
332	axe chemin, sentier
333	bord chemin, sentier
353	piste cyclable (bord)
360	trottoir structuré
383	filet d'eau, caniveau isolé
400	bord route-trottoir
401	bord route-piste cyclable
402	bord trottoir-piste cyclable
403	haut talus-clôture fil, palissade
404	haut talus-haie
405	haut talus-mur gauche
751	limites parcellaires
860	rail droit (restitué)-chemin de fer
865	rail gauche (calculé)-chemin de fer
870	rail droit -tram, métro-(restitué)
875	rail gauche -tram, métro-(calculé)
995	eau (ponctuel)

2.3.1.2. Production de couches d'information sur la disponibilité spatiale en données d'altitude

Nous avons construit deux couches d'information sur les données d'altitude existantes : une image de densité de points PICC, et une couche d'information sur les MNT existants.

L'image de **densité de points PICC** (Figure 35) a été créée à partir de tous les points PICC au sol, donc de tous les points des types répertoriés dans le tableau 20. La couche raster de densité a été calculée à une résolution de 40 m, en comptabilisant les points se trouvant dans un rayon de 120 m.

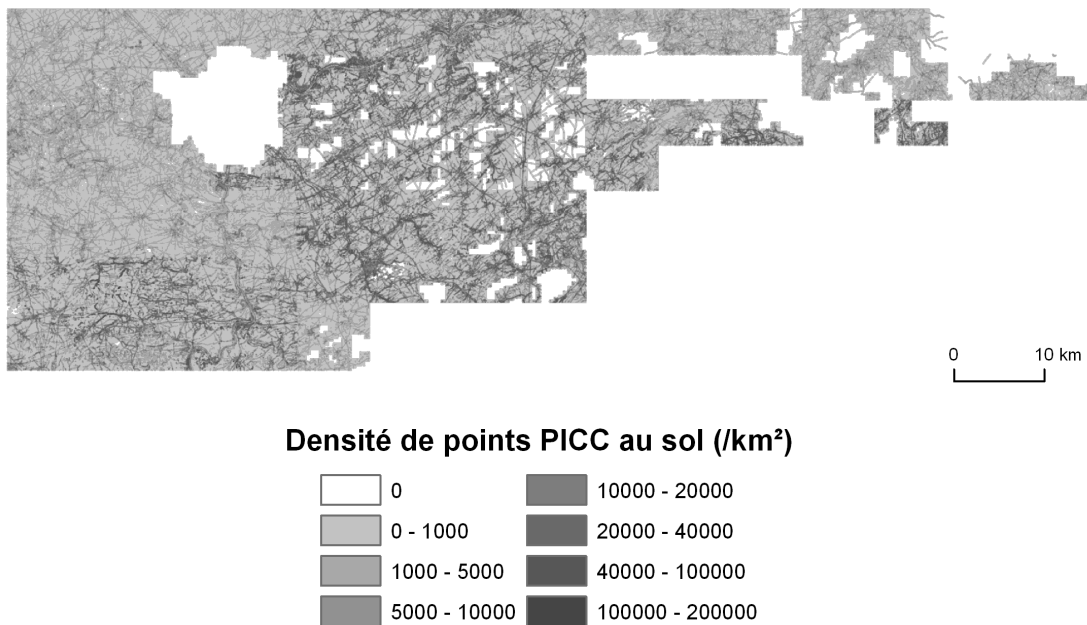


Figure 35. Image des densités de points PICC au sol dans le bloc 6

Une **couche** vectorielle **d'information** sur les données existantes a été créée. Elle permet d'identifier dans quelles zones les différentes données existent. De façon générale, il nous paraît indispensable de produire une couche spécifiant quelles données sont utilisées pour produire un MNT, précisément lorsqu'il s'agit de données de qualité aussi variable que dans le cas présent.

Le territoire du bloc 6 a été divisé en **zones** de **type** 1 à 8. Chaque type précise les différentes données disponibles (Tableau 21). La couche vectorielle d'information créée est illustrée à la figure 36.

Tableau 21. Définition des types de zones selon l'existence des différentes données

Type de zone	Existence des données (X = existant)			
	IGN Courbes de niveau (IGNcn)	IGN Points et lignes de structures (IGNpl)	PICC	MNT laser
1	X		X	X
2	X			X
3	X		X	
4	X			
5		X	X	X
6		X		X
7		X	X	
8		X		

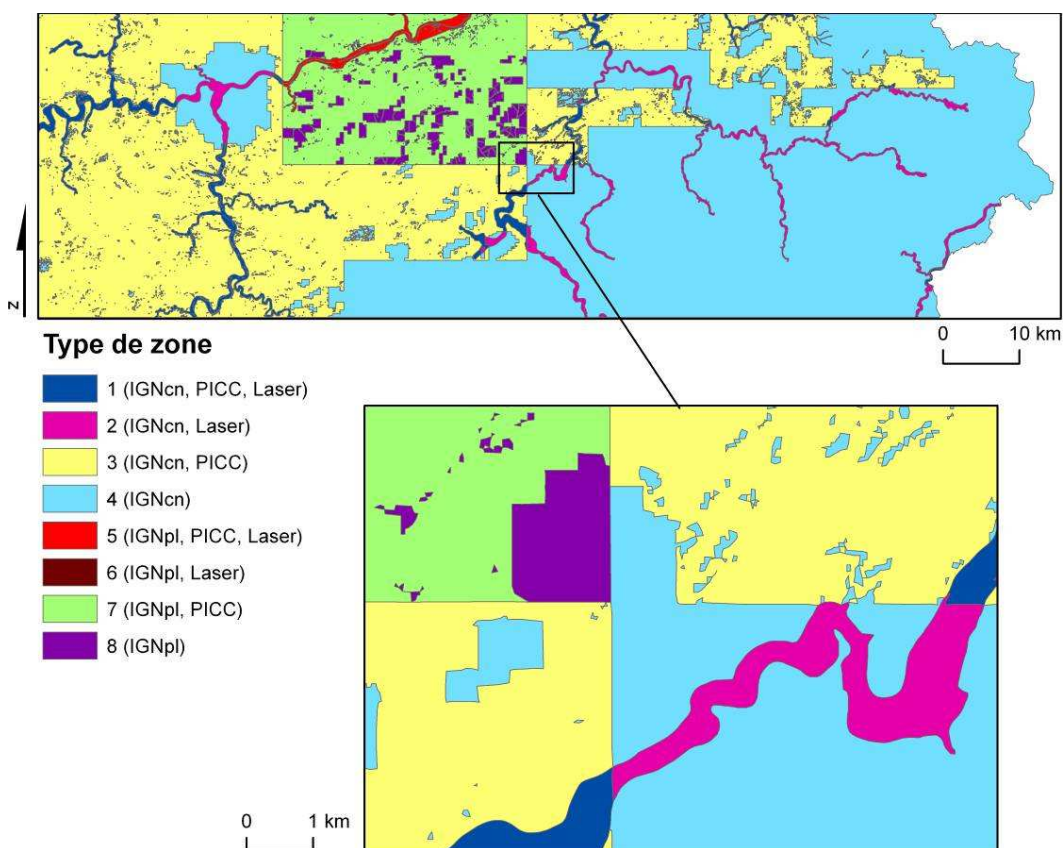
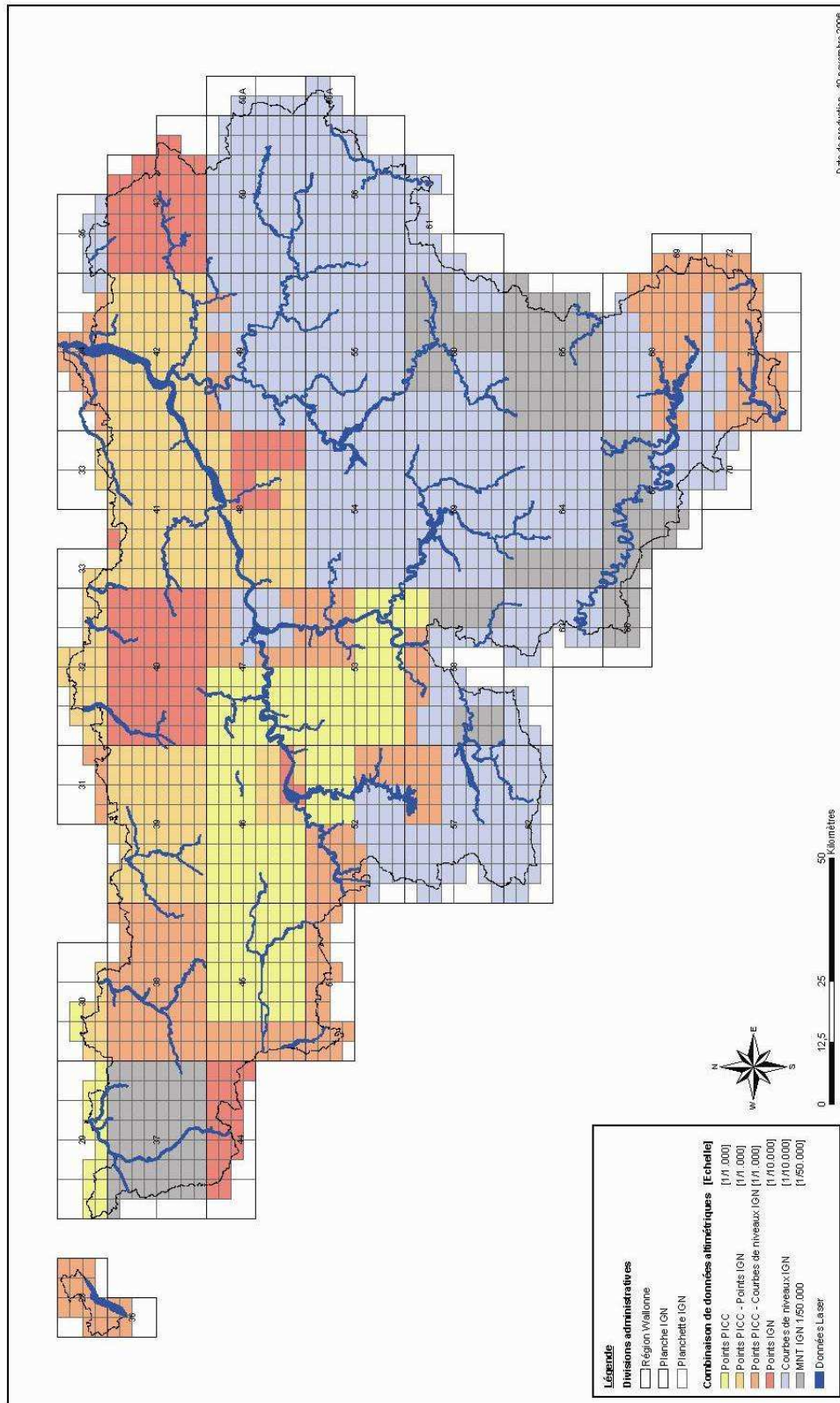


Figure 36. Couche vectorielle d'information sur les données d'altitude existantes (cn : courbes de niveau ; pl : points et lignes structurales)

Remarquons que le PICC a été **actualisé** depuis que le MNT ERRSUISSOL a été créé. Le MNT ERRUISSOL a été créé avec les anciennes données du PICC, alors que la couche vectorielle que nous avons conçue (Figure 36) reprend les données actuelles. Cette couche ne donne donc pas d'information sur les données utilisées pour produire le MNT ERRUISSOL.

La carte des données altimétriques utilisées pour la cartographie des pentes du MNT ERRUISSOL (Figure 37) donne pour chaque planche IGN au 1:10 000 une information sur la qualité locale du MNT ERRUISSOL, élaboré à partir de données de qualité variable (MNT laser, données du PICC, points et lignes structurales IGN, courbes de niveau IGN).



Prof. M^{me} S. Dautrebande
 Ir. F. Colard, Ir. F. Collette, Ir. P. Demarcin, Ir. A. Smoos
 Faculté Universitaire des Sciences Agronomiques (FUSAGx)
 Hydrologie & Hydraulique agricole - Génie rural & environnemental

Passage des Déportés, 2 - B-5030 Gembloux (Belgique)
 Tél : +32(0) 81 62 21 67 - Fax : +32(0) 81 62 21 81
 Email : dautrebande.s@gembloux.ac.be ; smoos.a@gembloux.ac.be



Figure 37. Données altimétriques utilisées pour la conception du MNT ERUISSOL, provenant de Demarcin P., Degré A., Smoos A., Dautrebande S. (2009). Projet ERUISSOL. Cartographie numérique des zones à risque de ruissellement et d'érosion des sols en Région wallonne. Rapport final de convention DGO3-FUSAGx. Unité d'hydrologie et hydraulique agricole. Faculté universitaire des Sciences agronomiques de Gembloux. 55 p+annexes.

2.3.1.3. Analyse de la qualité des MNT

a) Données utilisées

Les données (cf. ci-dessus) nous ont été fournies en différents formats, et la première partie du travail a consisté à les préparer afin qu'elles soient utilisables dans les logiciels utilisés. Pour l'analyse de la qualité des MNT, nous avons uniquement travaillé dans le bloc 6. Les MNT IGN, laser et ERRUISSOL sont sous forme d'image, mais les données PICC sont des données vectorielles (points, lignes et polygones). Nous avons décidé de ne pas réaliser d'interpolation sur ces données, afin de ne pas introduire d'erreur supplémentaire due à l'interpolation dans la comparaison des MNT.

b) Différence d'altitude entre le MNT IGN et le PICC

En chaque point PICC 'pnt' de type 185 (points de niveau de terrain), nous avons calculé la différence entre l'altitude donnée par le MNT IGN et celle donnée par le PICC. Nous avons également calculé une image de la pente du MNT IGN. La précision des pentes calculées de cette manière est évidemment dépendante de la propagation de l'erreur sur les données du MNT IGN.

La différence d'altitude entre le PICC et l'IGN est faible, de 0 à 3 m, sur les pentes faibles, mais plus importante, jusqu'à une quinzaine de mètres dans les zones de fortes pentes. Pour certains points, la différence d'altitude atteint une vingtaine de mètres ou plus.

Ces dernières valeurs résultent d'une différence de niveau de terrain entre les levés du PICC et de l'IGN qui n'ont pas été réalisés aux mêmes dates. Il s'agit de cas particuliers où le niveau du terrain a été modifié, dans les carrières, ou à la suite de travaux de génie civil. Cependant, hormis ces cas particuliers, la différence d'altitude entre les MNT est clairement liée à la pente (Figure 38).

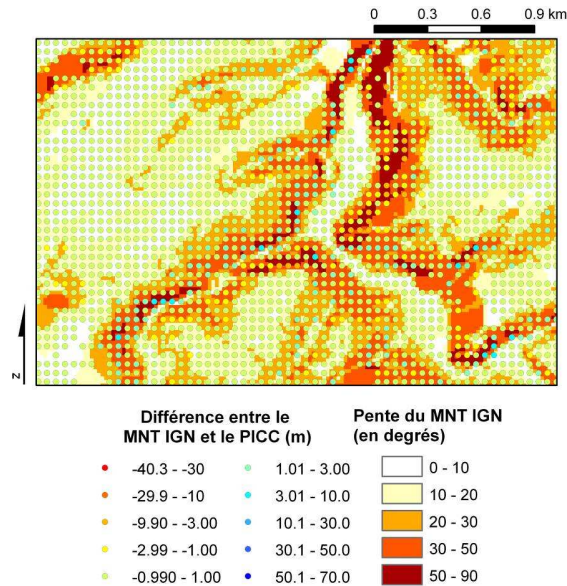


Figure 38. Exemples de différences entre le MNT IGN et le PICC plus importantes sur les fortes pentes.

Le graphique de la figure 39 met en relation la pente du MNT IGN avec la différence d'altitude entre le MNT IGN et le PICC, à partir d'un échantillon de points. La différence d'altitude entre le MNT IGN et le PICC est souvent plus importante en valeur absolue lorsque la pente est forte, mais pas toujours. Par contre, pour des pentes faibles, cette différence est toujours assez faible, mis à part quelques outliers qui sont liés à des phénomènes locaux tels qu'une modification du niveau du terrain entre le levé de l'IGN et celui du PICC. Le PICC apporte donc une information supplémentaire au MNT IGN surtout dans les zones de fortes pentes.

Dans les régions à forte pente, les hauteurs calculées dans les zones où la densité de points PICC est suffisante sera fiable mais probablement pas là où seules les données de l'IGN sont disponibles.

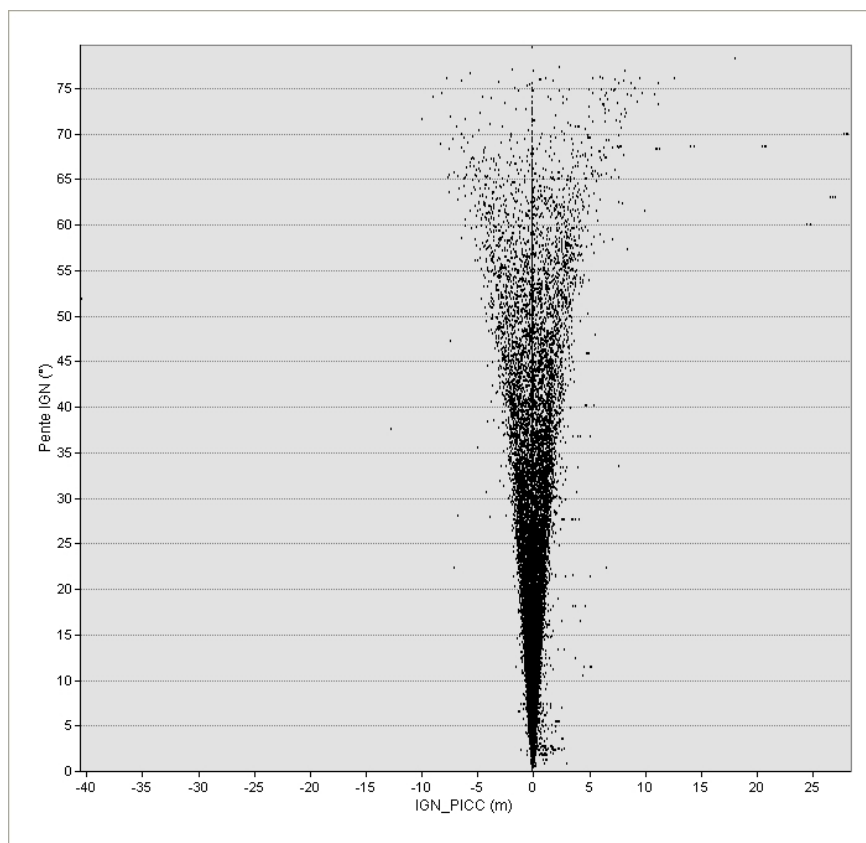


Figure 39. Relation entre la pente du MNT IGN et la différence d'altitude entre le MNT IGN et le PICC.

c) Comparaison des MNT dans des zones test

Afin d'analyser la qualité du MNT IGN et du PICC, nous les avons comparés dans des zones test. Cinq zones tests ont été choisies à des endroits-clés, là où les données du PICC sont disponibles, afin de pouvoir comparer les MNT dans des paysages différents : avec ou sans relief, sur des pentes fortes ou faibles, en zone forestière, cultivée ou en zone urbaine (Figure 40 et tableau 22).

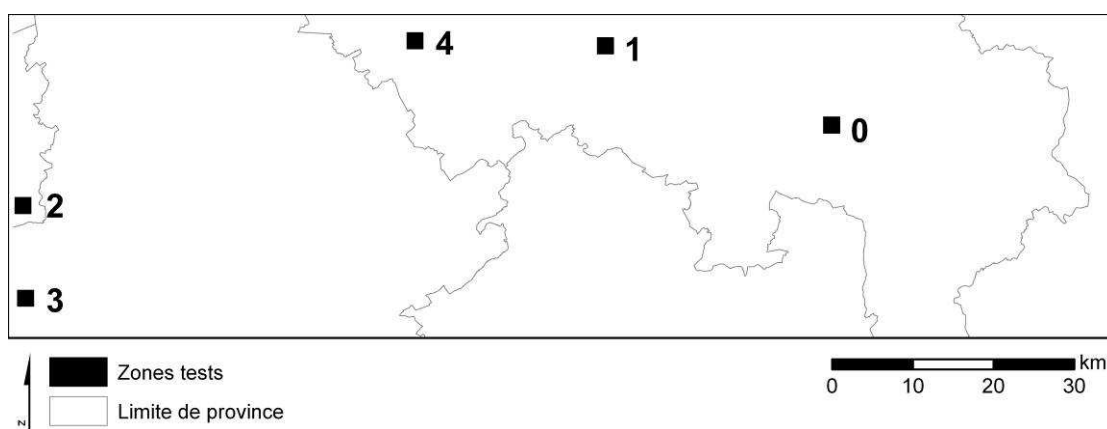


Figure 40. Localisation des zones test et leur identifiant dans le bloc 6.

Tableau 22. Caractéristiques des zones tests choisies.

Identifiant polygone	Couvert	Pente	Lieu
0	Foret + Champs	forte + moy/longue	Stavelot
1	Bois + Champs	forte/vallee	Poulseur
2	Bois + Champs	faible	Villers Poterie
3	Champs	faible + vallee	Saint Aubin
4	Urbain + Bois	forte/vallee	Huy

En calculant les statistiques de ces différences d'altitude par zone test (Tableau 23), on voit que les valeurs absolues des moyennes sont inférieures à 0,1 m dans les zones à faible relief (zones 0, 2 et 3). Pour les zones où l'on a de fortes pentes, elles sont plus élevées et les écarts-types aussi. Ce sont dans les zones test de Huy et de Poulseur que l'on retrouve les plus fortes différences. Par contre, les moyennes et écarts-types des différences d'altitude entre le MNT laser et le PICC restent très faibles.

Tableau 23. Moyenne et écart-type des différences d'altitude entre, d'une part, le MNT IGN et le PICC, et d'autre part, le MNT laser et le PICC, dans chaque zone test.

Identifiant polygone	Lieu	MNT IGN - PICC		MNT laser - PICC	
		Moyenne (m)	Ecart-type (m)	Moyenne (m)	Ecart-type (m)
0	Stavelot	0,03	1,71		
1	Poulseur	-0,43	4,86	-0,10	0,35
2	Villers Poterie	-0,02	0,47		
3	Saint Aubin	-0,04	0,70		
4	Huy	-0,58	4,42	0,23	1,06

Les figures 41 à 45 montrent les différences d'altitude entre le MNT IGN et le PICC, et entre le MNT laser et le PICC (là où le MNT laser existe), dans les 5 zones tests. Dans les zones 0, 2 et 3, on a globalement des différences de -5 à 5 m, réparties dans toute la zone test. Parmi ces 3 zones, la zone 0 comprend les plus grandes différences. Les zones 1 et 4 montrent des différences beaucoup plus importantes, de l'ordre de -30 à 30 m. Ces maxima se situent dans les zones de fortes pentes et de falaises.

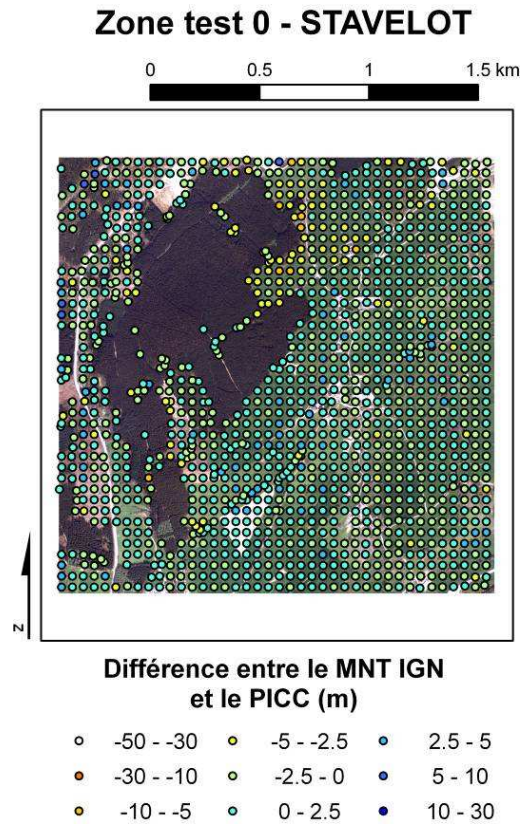


Figure 41. Différence entre le MNT IGN et le PICC dans la zone test 0.

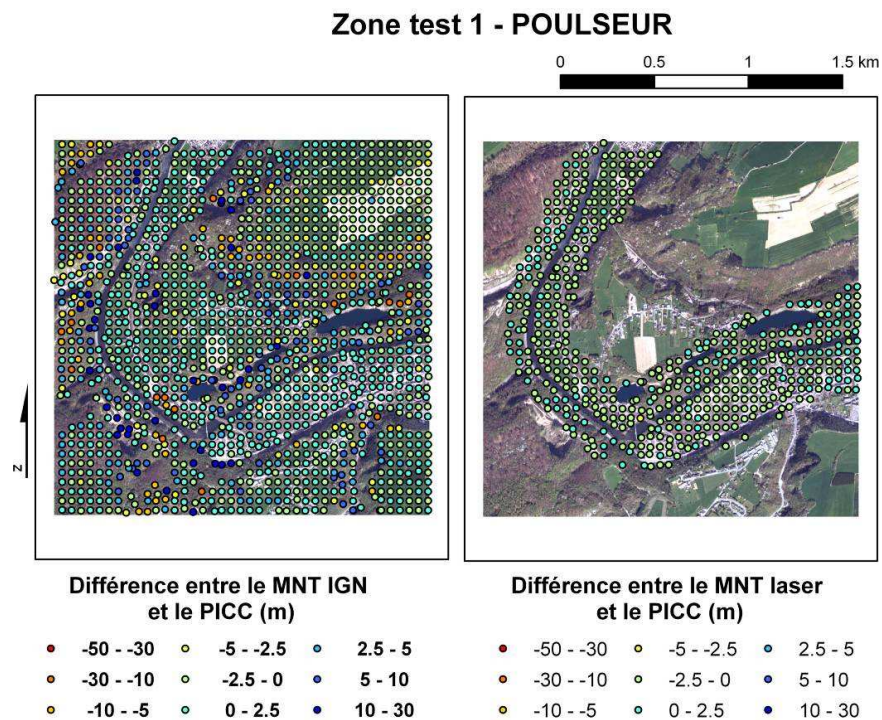


Figure 42. Différences entre le MNT IGN et le PICC et entre le MNT laser et le PICC dans la zone test 1.

Zone test 2 - VILLERS POTERIE

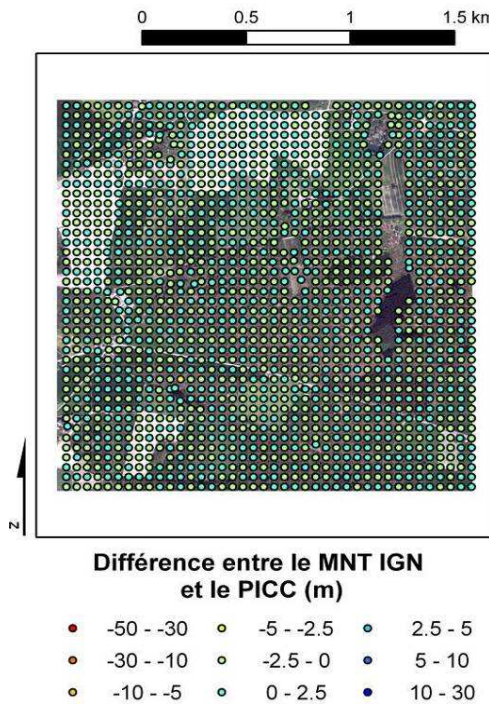


Figure 43 (à gauche). Différence entre le MNT IGN et le PICC dans la zone test 2.

Zone test 3 - SAINT AUBIN

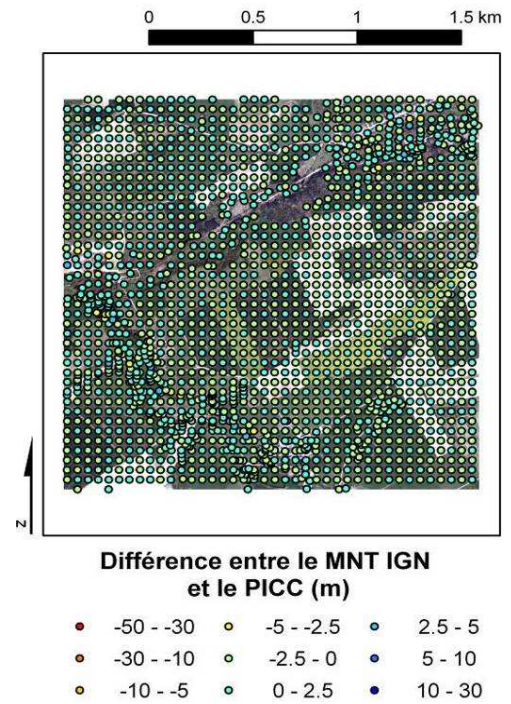


Figure 44 (à droite). Différence entre le MNT IGN et le PICC dans la zone test 3.

Zone test 4 - HUY

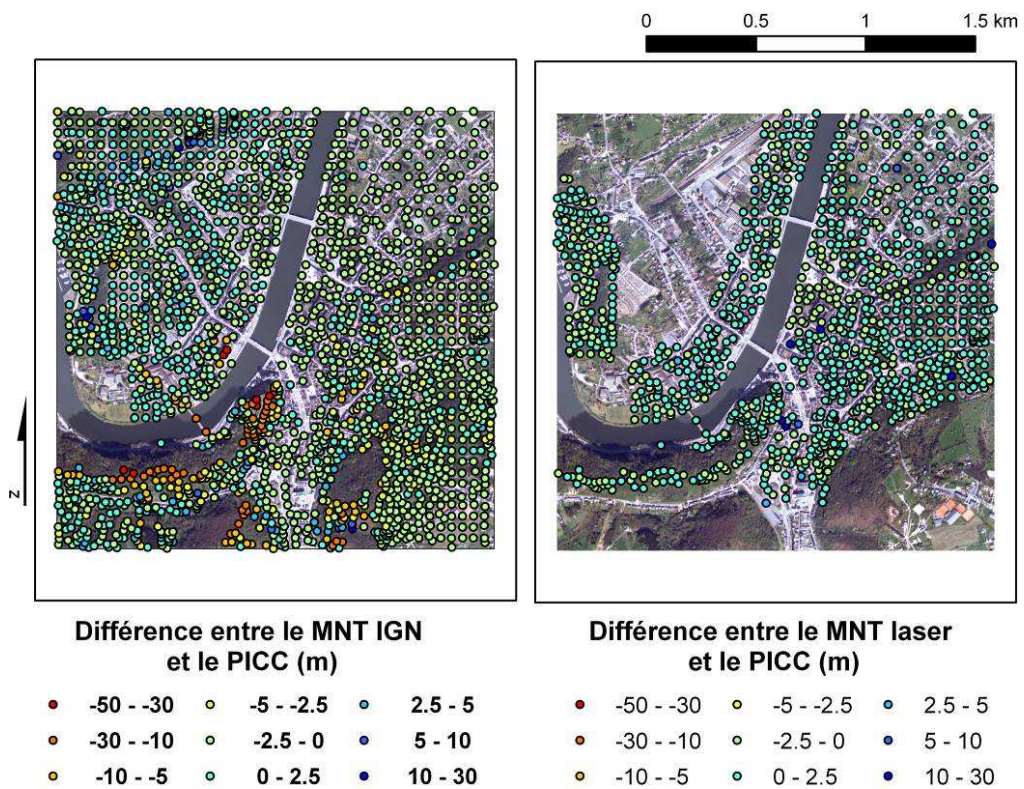


Figure 45. Différences entre le MNT IGN et le PICC et entre le MNT laser et le PICC dans la zone test 4.

La différence entre le MNT IGN et le PICC peut être très forte là où la pente est forte, tandis que la différence entre le MNT laser et le PICC est faible (Figures 42 et 45). Le PICC apporte donc une information non négligeable par rapport au MNT IGN dans les zones de fortes pentes et de falaises, si on considère le MNT laser comme une vérité terrain, et ce même en ne considérant que les points 'pnt' de type 185 (points de niveau de terrain), et sans considérer les points 'ptl' provenant d'éléments linéaires (bas talus, haut talus, falaise pied, falaise haut).

Afin de mettre en évidence cette relation entre la pente et la différence d'altitude entre le MNT IGN et le PICC, nous avons reclassé en 5 catégories les pentes calculées à partir du MNT IGN. Nous avons ensuite calculé les statistiques des différences d'altitude dans chaque catégorie (Tableau 24). On remarque que les moyennes sont très proches de zéro quelle que soit la pente. Par contre, on observe un écart-type de plus en plus important lorsque la pente augmente.

Tableau 24. Statistiques des différences d'altitudes entre le MNT IGN et le PICC pour chaque catégorie de pente.

Catégorie de pente	1	2	3	4	5
Pente	0-10°	10-20°	20-30°	30-50°	50-90°
Nombre de points	286349	329797	178783	157966	64680
Minimum	-81,0	-41,5	-51,6	-59,5	-129,4
Maximum	58,7	59,4	109,5	55,7	95,5
Moyenne	0,11	0,18	0,16	0,09	0,06
Ecart-type	1,16	1,22	1,46	2,11	4,42

d) Analyse du MNT_ERRUISSOL

Deux métadonnées que nous avons générées complètent MNT_ERRUISSOL: la couche vectorielle des « zones types » et l'image de densité de points du PICC. Ils permettent tous les deux d'avoir une information supplémentaire sur la qualité locale de MNT_ERRUISSOL. En effet, ce dernier a été généré à partir de données provenant des MNT IGN, laser et du PICC, qui sont de qualités différentes.

Différence d'altitude entre les MNT_ERRUISSOL et IGN

Nous avons calculé la couche de différence d'altitude entre les MNT_ERRUISSOL et IGN, afin de les comparer.

En règle générale, les différences d'altitude entre les MNT_ERRUISSOL et IGN sont plus faibles là où seul le MNT_IGN existe, par rapport aux zones où le PICC et le MNT_IGN existent. En effet, dans ces premières zones, le MNT_IGN a été pris en compte pour générer le MNT_ERRUISSOL.

Dans les zones où seul le MNT_IGN existe, on a de fortes différences d'altitude là où les pentes sont fortes (Figure 46). Pourtant, d'après les informations disponibles sur le MNT_ERRUISSOL, ce sont les données du MNT_IGN qui ont été prises en compte dans les zones où seul le MNT_IGN existe. Le MNT_ERRUISSOL devrait donc être rigoureusement similaire au MNT_IGN dans ces zones et les différences devraient être proches de zéro. Or, elles atteignent plus de 5 m dans les zones de fortes pentes. Les différences observées sont très probablement dues à la méthode d'interpolation utilisée, soit lors de la conception du MNT_ERRUISSOL, la méthode de Hutchinson appliquée aux données vectorielles de l'IGN.

Dans certaines petites zones où le MNT_IGN seul existe (Figures 47 et 48), on a de fortes différences d'altitude entre les MNT_ERRUISSOL et IGN. Ces différences plus fortes sont bien liées aux MNT et non à l'occupation du sol, comme on le voit sur l'image en vraies couleurs en bas de la figure 47. Ces petites zones de fortes différences sont toutes entourées de zone où le MNT_IGN et le PICC existent (Figure 48). Lors de la génération du MNT_ERRUISSOL, ces zones ont sans doute été considérées comme zones où le PICC existe. En effet, elles se trouvent dans un pavé du PICC existant, mais la densité de points du PICC est très faible dans les petites zones en question que nous n'avons pas reprises dans les zones où les données du PICC existent sur base de notre calcul de densité de points PICC.

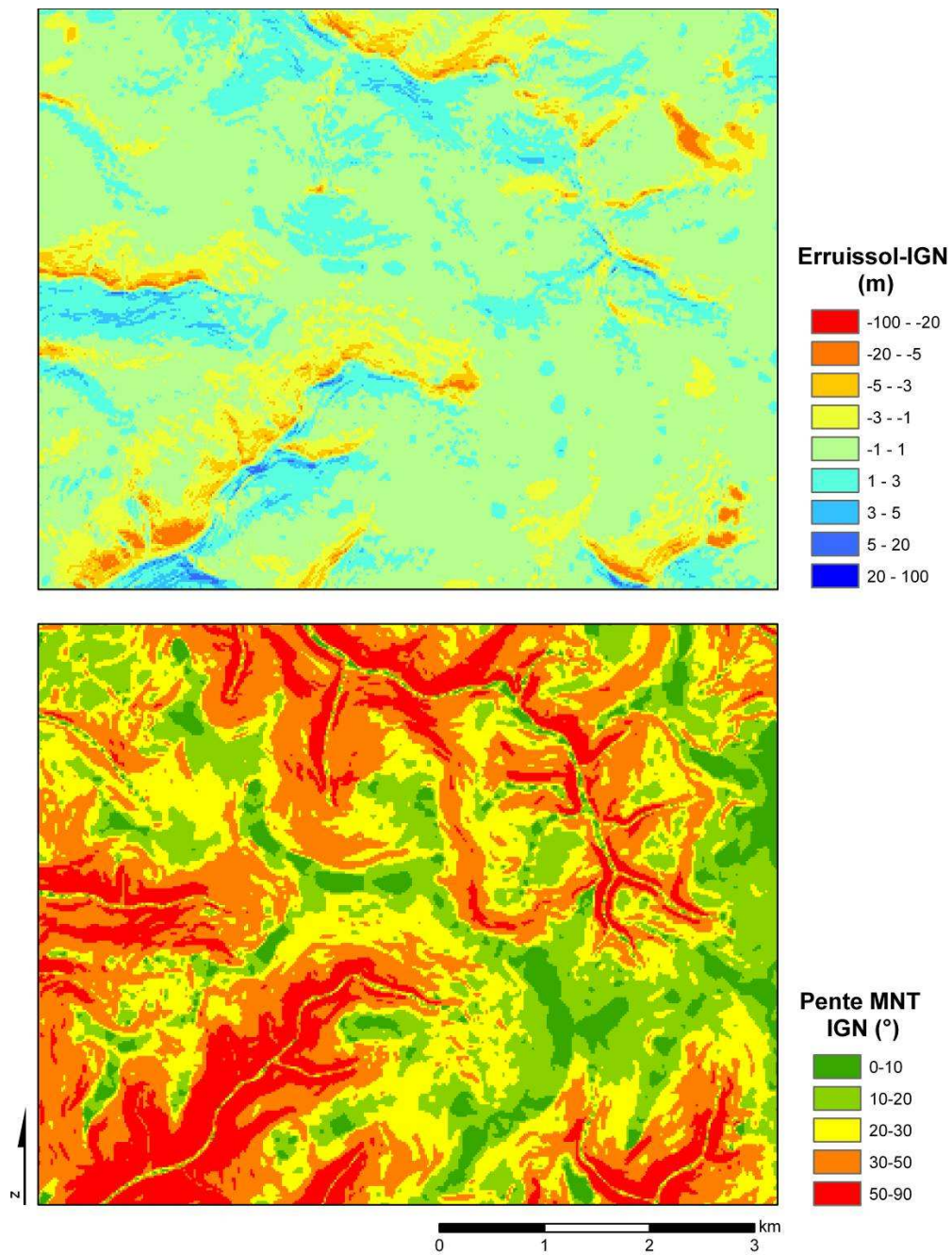


Figure 46. Exemple de différences d'altitude entre les MNT_ERRUISSOL et IGN plus fortes là où les pentes sont fortes, dans les zones où seul le MNT IGN existe.

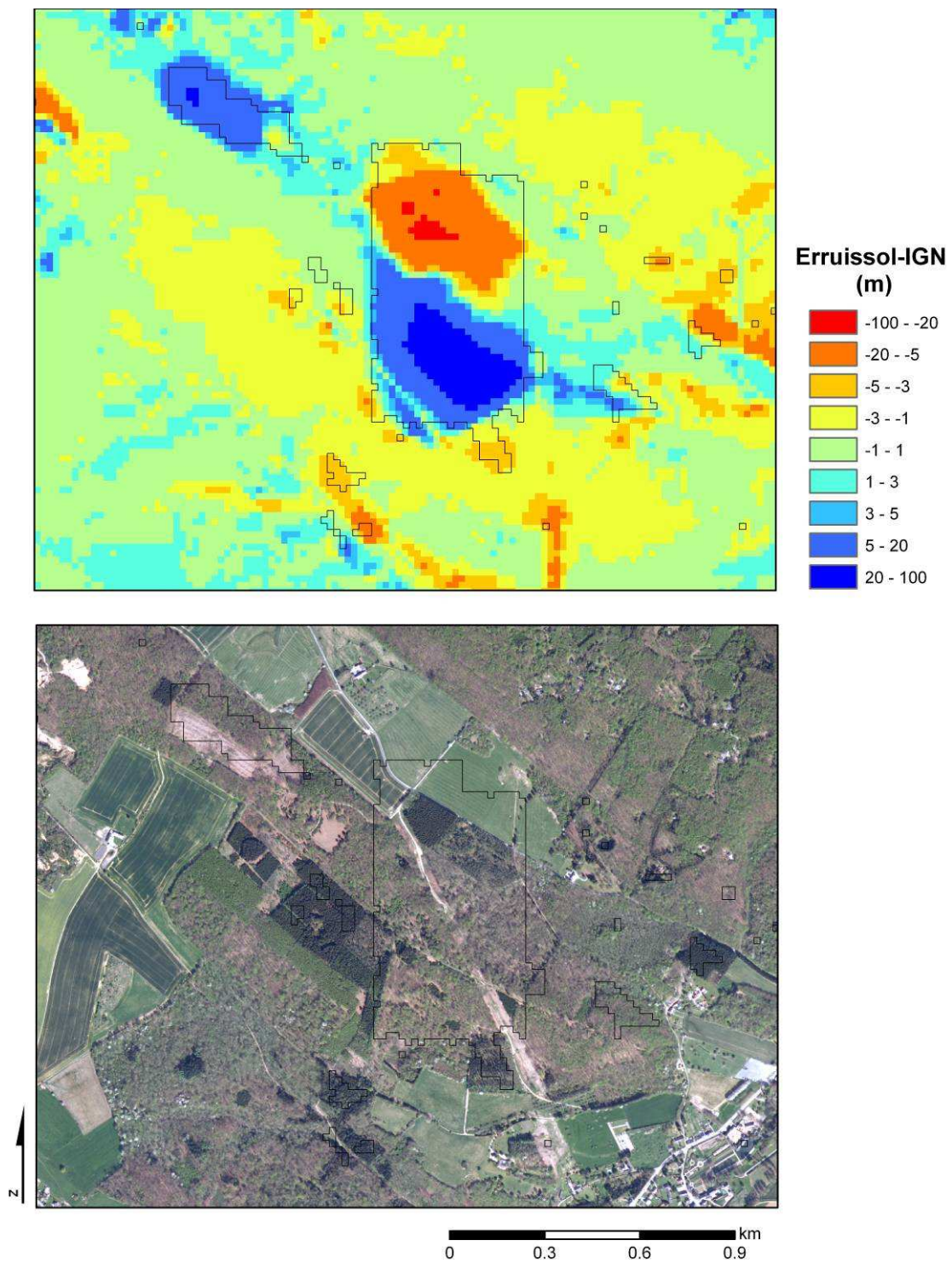


Figure 47. Différence d'altitude entre les MNT_ERRUISSOI et IGN (en haut) plus fortes dans les zones où le PICC n'existe pas (zones à l'intérieur des polygones en traits noirs). Ortho-image dégradée à 2,5 m en vraie couleur de la même zone (en bas).

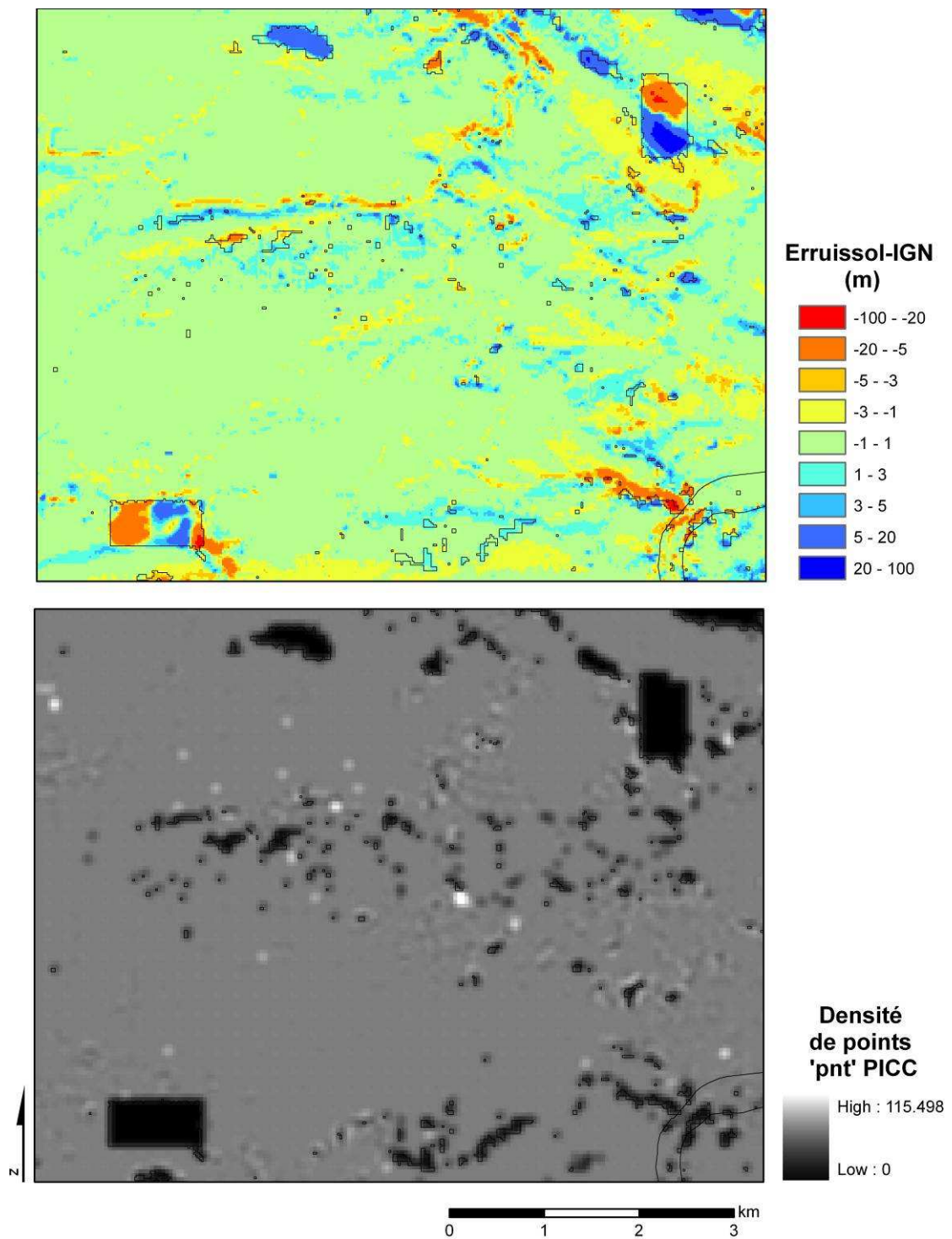


Figure 48. Grandes différences entre les MNT_ERRUISSOI et IGN dans certaines petites zones de faible densité du PICC.

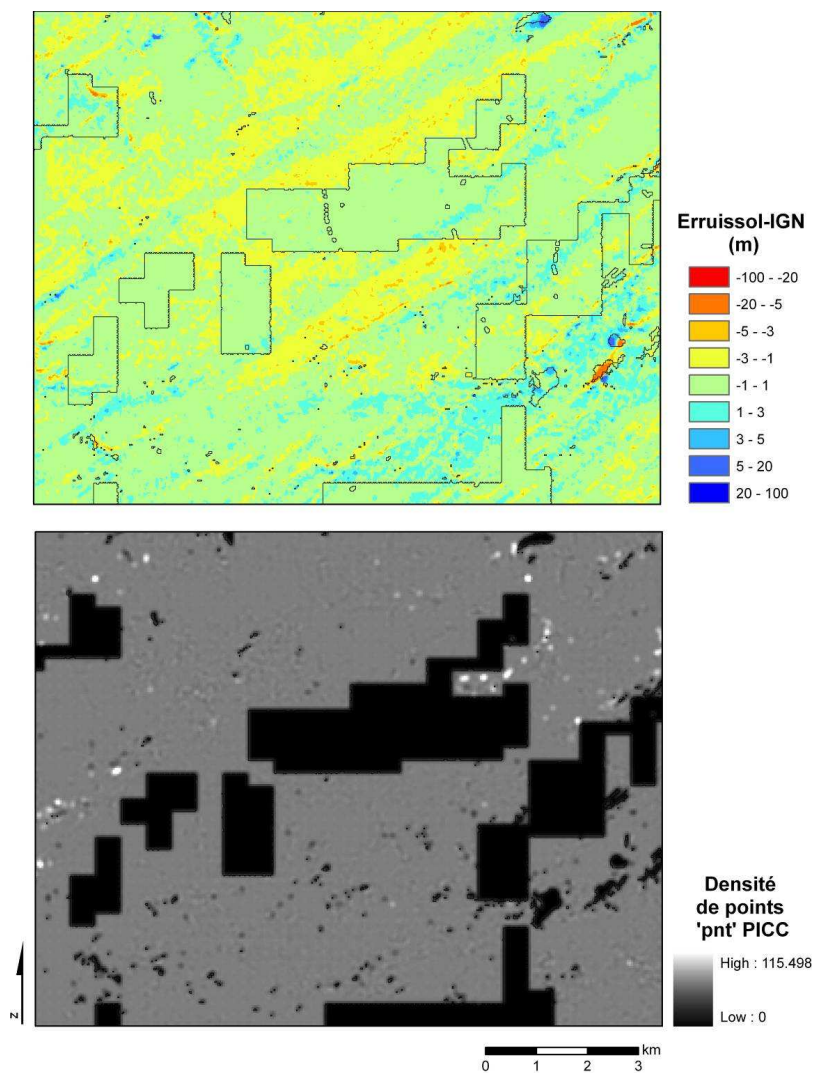


Figure 49. Zones plus grandes où la densité du PICC est très faible, et où les ‘trous’ du PICC ont visiblement été ‘remplis’ par des données IGN lors de l’élaboration du MNT_ERRUISSOL.

Dans d’autres zones un peu plus grandes (Figure 49), les endroits de faible densité de points du PICC ont été considérés comme des ‘trous’, et on voit bien sur la figure que l’on a inséré des données du MNT IGN pour remplir ces ‘trous’, ce qui explique que l’on retrouve des différences entre les MNT ERRUISSOL et IGN plus faibles dans ces ‘trous’ du PICC.

Dans les zones où le MNT laser existe, on retrouve de plus grandes différences d'altitude entre les MNT ERRUISSOL et IGN. En effet, dans ces zones, c'est le MNT laser qui a été considéré pour générer le MNT_ERRUISSOL. Or, le MNT laser est beaucoup plus précis que le MNT IGN, ce qui explique que les différences locales soient bien marquées (Figure 50). Remarquons tout de même que le MNT laser tient compte des routes, ponts, etc. alors que, lorsque le PICC a été utilisé pour générer le MNT_ERRUISSOL, seuls certains points du PICC ont été pris en compte, et notamment, les points relatifs aux voiries n'ont pas été considérés. Par exemple, sur un pont traversant une rivière, le MNT laser donne l'altitude du pont alors que le MNT IGN donne l'altitude de la rivière.

Le MNT_ERRUISSOL donnera donc soit l'altitude du pont, si on se trouve dans une zone où le MNT laser existe, soit l'altitude de la rivière, là où le MNT laser n'existe pas. Même si ces différences sont relativement localisées, il nous paraît indispensable de considérer le MNT_ERRUISSOL accompagné de la couche des 'zones types' que nous avons produite.

De fortes différences d'altitude entre les MNT ERRUISSOL et IGN se retrouvent également dans des cas isolés où des modifications du niveau de terrain ont eu lieu entre les dates des levés du PICC et de l'IGN, comme illustré précédemment (carrières, échangeur d'autoroute, etc.).

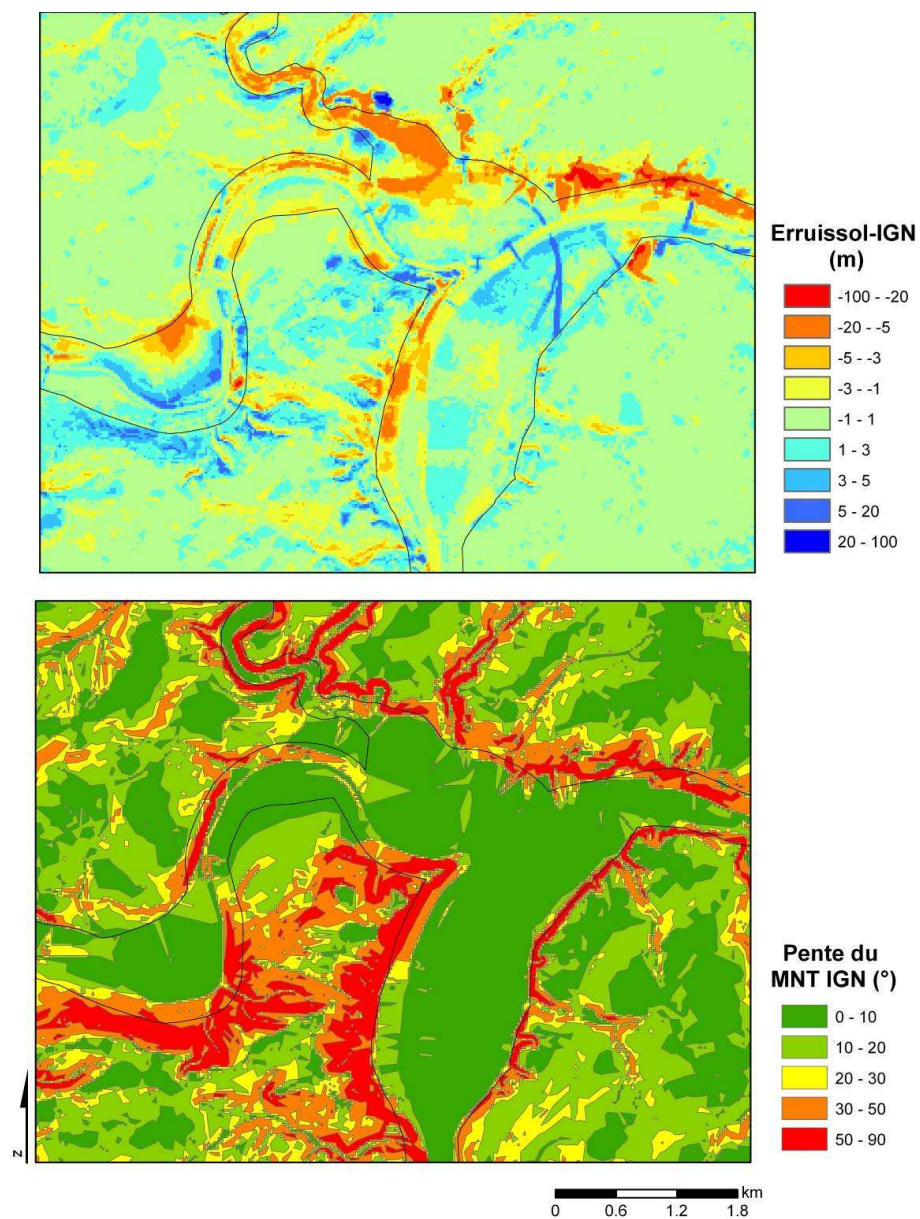


Figure 50. Plus grandes différences d'altitudes localisées dans les zones où le MNT laser existe (à l'intérieur du polygone en trait noir), dues à la précision plus grande du MNT laser. Les fortes différences d'altitudes dans la zone où le MNT laser n'existe pas (à l'extérieur du polygone en trait noir) sont dues aux fortes pentes. On remarque aussi que le MNT laser prend en compte les routes, ponts, etc.

e) Conclusion

Différentes analyses de la qualité des MNT ont été réalisées. Celles-ci ont montré que les données du PICC sont de meilleure qualité que les données IGN, et cela se marque surtout dans les fortes pentes. La précision et l'exactitude de l'altitude donnée par les points du PICC est comparable à celle donnée par le MNT/MNS laser. Le MNT/MNS laser décrit la surface de façon régulière (mailles) alors que les données du PICC sont ponctuelles, et l'interpolation sur ces points donnera bien entendu une moins bonne qualité que le MNT/MNS laser, très précis partout où il existe.

En réalisant ces analyses de qualité des MNT, nous nous sommes rendus compte que la **qualité** du MNT ERRUISSOL n'est **pas homogène** mais dépend des types de données vectorielles utilisées (MNT laser, PICC, courbes de niveau IGN, points et lignes de structures IGN).

2.3.1.4. Production d'un TIN

Remarque préliminaire

En ce qui concerne le **PICC**, les analyses précédentes ont été faites seulement pour les points de type 185 (points de niveau de terrain). Sous le conseil des membres du comité d'accompagnement de notre projet, nous considérons à présent tous les types de points au sol. Ceci implique une probable modification de la qualité des données du PICC dans les zones de fortes densités de points, bien que ces points ajoutés permettent surtout d'injecter une information à très haute fréquence spatiale, peu compatible avec la résolution de travail (5 m).

Nous avons créé un TIN, converti ensuite en MNS par interpolation linéaire, à partir des meilleures données altimétriques disponibles. Le principe est le même que pour la conception du MNT ERRUISSOL en ce qui concerne la combinaison des données altimétriques. Le MNT laser étant le plus précis, c'est lui qui a été pris en compte là où il existait. Pour les zones où il n'existait pas, les données du PICC ont été choisies, et les données IGN ont été utilisées pour combler les « trous » du PICC et couvrir les zones où ni PICC ni le MNT laser n'existent.

Le TIN a d'abord été créé à partir des données altimétriques du **PICC** et des points et lignes structurales **IGN** ou des courbes de niveaux **IGN**. Les lignes de structure du PICC n'ont cependant pas été utilisées comme *breaklines*. Une interpolation linéaire sur ce TIN a permis de produire un MNT à 5 m de résolution. Les données du **MNT/MNS laser** ont été introduites, après avoir subi une dégradation par rééchantillonnage et agrégation des altitudes dans des fenêtres de 25 pixels en calculant la moyenne dans cette fenêtre. Enfin, nous avons remplacé l'altitude du MNT-TIN par le MNT/MNS laser dégradé dans les zones où ces données existent.

Le MNT ainsi obtenu et le MNT ERRUISSOL diffèrent ainsi par le **calcul** (TNI et interpolation linéaire), les **types de points** du PICC pris en considération, et la **résolution**. Le MNT ERRUISSOL ayant pour but de fournir un MNT hydrologiquement correct, la méthode d'interpolation de Hutchinson a été utilisée et les points qui pouvaient modifier les pentes, comme les points des réseaux de voirie, n'ont pas été pris en compte. Le but étant dans ce projet de fournir une couche de hauteur utilisable dans les procédures de classification, nous avons besoin de l'altitude la plus correcte partout, sans trop se soucier de l'intérêt hydrologique du produit final. C'est pour cette raison que nous avons pris en compte tous les points au sol du PICC et que nous avons créé un TIN suivi d'une interpolation linéaire. De plus, la résolution du MNT ERRUISSOL est de 10 m, alors que le MNS_DGA a une résolution de 5 m.

En théorie, le MNT obtenu répond mieux aux objectifs de l'étude. Il a été produit dans le bloc 6 et sera comparé au MNT ERRUISSOL dans cette même zone.

2.3.1.5. Comparaison des MNH construits à partir du TIN et du MNT ERRUISSOL

a) Production des MNH

Deux modèles numériques de hauteur ont été produits pour le bloc 6 :

- le MNH-ERRUI (ou HEST au paragraphe 2.7), obtenu par la soustraction du MNS_DGA et du MNT ERRUISSOL,
- le MNH-TIN, obtenu par la soustraction du MNS_DGA et du TIN.

Les différences de hauteurs constatées entre ces deux MNH n'est donc fonction que du MNT utilisé puisque le MNS constitue le premier terme des deux différences.

Le **MNH-TIN** a une résolution de 5 m, car il provient d'une soustraction entre 2 images à 5 m de résolution, parfaitement cohérentes d'un point de vue géométrique. Pour le **MNH-ERRUI**, on soustrait d'une valeur du MNS_DGA à 5m de résolution, celle du MNT ERRUISSOL à 10 m de résolution. Nous avons gardé la résolution la plus fine, afin d'obtenir un MNH à 5 m de résolution et de pouvoir ainsi le comparer plus facilement au MNH-TIN. Une procédure d'interpolation au plus proche voisin sur le MNT_ERRUISSOL a été adoptée pour égaliser les résolutions lors de la soustraction.

Nous avons **comparé** ces deux MNH afin d'analyser leur qualité et de déterminer si le MNH-TIN se caractérise par un gain réel en **précision** par rapport au MNH-ERRUI et s'il est pertinent ou non de produire un TIN plutôt que d'utiliser le MNT ERRUISSOL. La création d'un TIN se justifie en fait là où les données décrivent très bien les formes du relief. Les valeurs de hauteur données par les deux MNH sont exprimées en **décimètres**, afin de pouvoir être codées en nombres entiers et de réduire la taille des images. Ce codage est suffisamment précis par rapport aux précisions des deux produits (résolution imposée à 5 m).

Néanmoins, l'effet de la procédure de rééchantillonnage au plus proche voisin combiné à cette précision produit l'apparition de structures à très haute fréquence et faible amplitude sur le MNH-ERRUI.

b) Statistiques des différences des hauteurs données par les 2 MNH

Afin de repérer les **différences** entre la hauteur donnée par les 2 MNH, pour chaque pixel, nous avons soustrait de la hauteur donnée par le MNH-TIN celle donnée par le MNH-ERRUI. Cette différence correspond en fait à la différence entre le MNT-TIN et le MNT-ERRUI.

Les statistiques sur les différences de hauteurs entre les 2 MNH dans l'ensemble du bloc 6 sont reprises dans le tableau 25. La **moyenne** est égale à zéro, et l'**écart-type** est très faible (1,5 m), ce qui prouve la bonne correspondance globale entre les 2 MNT et donc les 2 MNH malgré des valeurs extrêmes qui sont très importantes en valeur absolue mais peuvent être qualifiées de marginales ou *outliers*. Il y a donc quelques zones de très grandes différences entre les 2 MNH.

Tableau 25. Moyenne, écart-type et déciles des différences de hauteurs données par les 2 MNH (en décimètres)

moyenne	0
écart-type	15
min	-2199
d1	-9
d2	-4
d3	-2
d4	-1
d5	0
d6	1
d7	3
d8	6
d9	12
d10 =max	2679

Si l'on considère comme *outliers* les valeurs qui n'appartiennent pas à l'intervalle de confiance à 99 %, on constate qu'elles se trouvent surtout dans les carrières, les lacs et réservoirs, dans les zones bâties et sur les routes.

c) Analyse des différences entre les 2 MNH dans les zones où les données de l'IGN ont été utilisées pour la conception des MNT

Il convient de distinguer les zones où des données **identiques** ont été utilisées pour la production du TIN et du MNT ERRUISSOL des zones où des données **différentes** ont été utilisées. Les seules zones où les données sont rigoureusement les mêmes pour les deux modèles sont celles où les données IGN ont été utilisées. Dans les zones où le PICC a été utilisé, des types de points différents ont été pris en compte pour chaque modèle. Dans les zones où les données laser ont été utilisées, les concepteurs du MNT ERRUISSOL ont utilisé un MNT laser corrigé des points ne donnant pas les altitudes au sol, alors que nous avons utilisé le MNT laser non corrigé dans la production du MNT extrait du TIN.

Dans les zones où les données IGN ont été exploitées pour produire les MNT, les différences entre les MNH sont forcément dues à la méthode **d'interpolation**, puisque les données d'altitude utilisées sont les mêmes (points et lignes structurales ou courbes de niveau IGN). Il y a très **peu d'outliers** identifiés dans ces zones, sauf très localement (lacs, réservoirs d'eau). Dans les zones de lacs, réservoirs, ou simplement les rivières qui ne se trouvent pas dans une zone où le MNT/MNS laser existe, on observe des incohérences importantes du MNH-ERRUI.

En effet, la méthode d'interpolation de **Hutchinson** utilisée dans le calcul du MNT ERRUISSOL produit dans ces contextes un profil en dépression fermée ou l'altitude peut être sous estimée (Figure 52). Le MNT-TIN est construit par interpolation **linéaire** ne produisant pas ce type d'effet dans ce contexte. Les incohérences y sont importantes entre le MNT-TIN et le MNT ERRUISSOL.

Dans les **fonds de vallées**, on observe aussi des zones de différences entre les MNH espacées régulièrement (Figure 51). Contrairement au cas précédent, ces incohérences sont dues à l'imperfection du MNT-TIN car le TIN a été créé sans lignes structurales. Cela induit dans ce contexte la création de **triangles horizontaux** responsable des différences de hauteur constatées. Contrairement aux lacs et réservoirs, ici, c'est le MNT ERRUISSOL qui semble donner une interpolation plus proche de la réalité, comme l'attestent les profils de la figure 52. D'autres zones d'incohérence du même type ont été identifiées un peu partout dans l'image, à chaque fois que le tracé des courbes de niveau IGN peut produire la création de triangles horizontaux (Figure 53).

Des différences entre les 2 MNH de l'ordre de 1 ou 2 m existent dans les zones de **forte pente** (Figure 52). Ces différences sont dues à la différence de résolution entre le MNT-TIN et le MNT ERRUISSOL utilisés pour la construction des MNH (Figure 54). Au niveau des fortes pentes, on a un grand nombre de courbes de niveau, très proches les unes des autres, et les erreurs liées à la méthode d'interpolation sont très faibles.

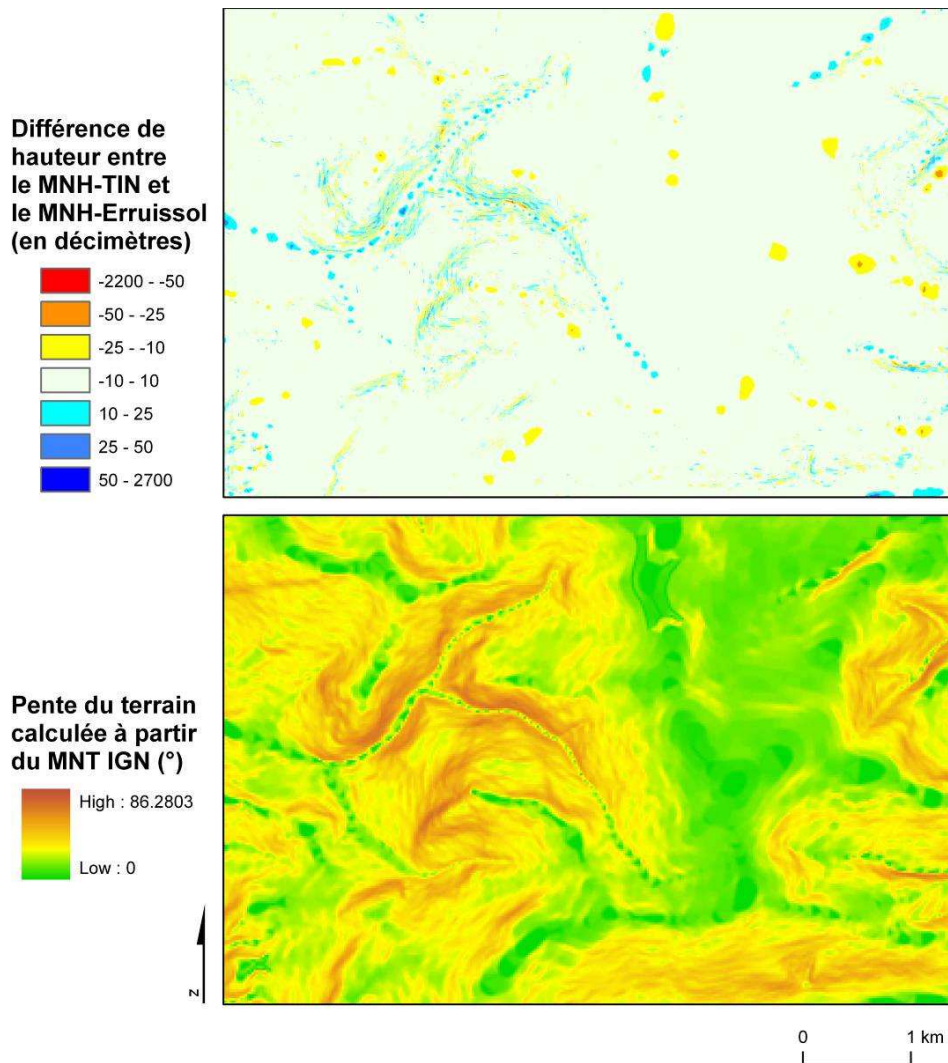


Figure 51. Incohérences entre les MNH régulièrement espacées le long des fonds de vallée dues à la méthode d'interpolation, et différence entre les hauteurs données par les 2 MNH, de l'ordre de quelques mètres, principalement dans des zones de forte pente

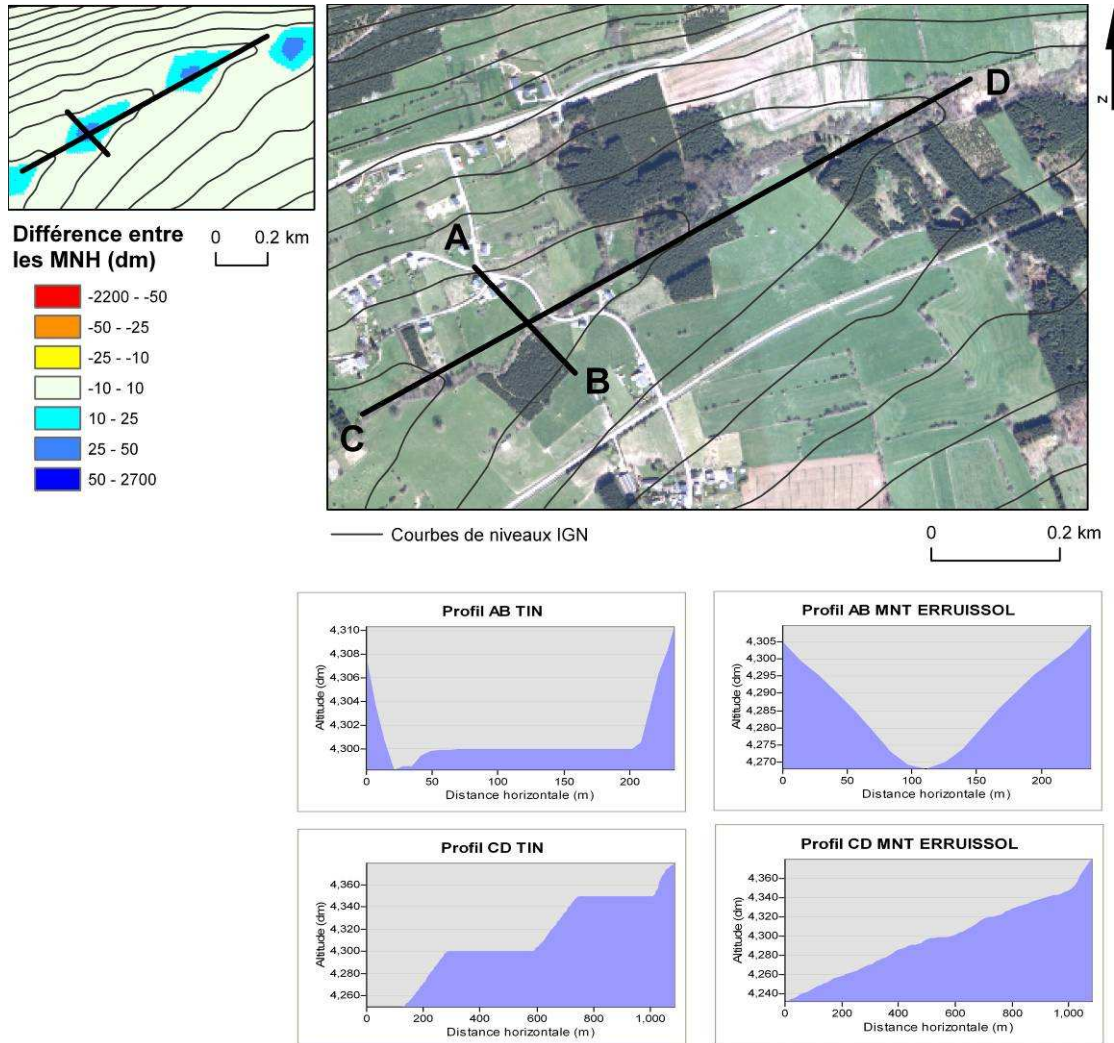


Figure 52. Incohérences entre les 2 MNH liées à la création de triangles horizontaux dans les fonds de vallée où seules les courbes de niveaux IGN ont été utilisées pour créer le MNT-TIN. Dans ces cas, le MNT ERRUISSOL produit les valeurs les plus proches de la réalité.

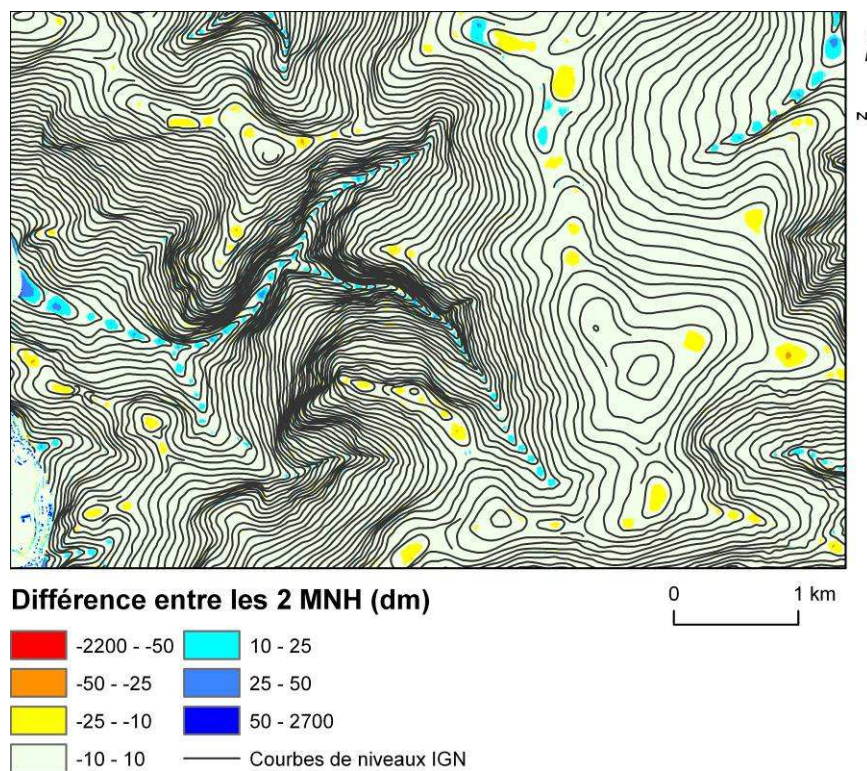


Figure 53. Mise en évidence d'incohérences entre les 2 MNH là où, en dehors des vallées, le tracé des courbes de niveau IGN produit aussi des triangles horizontaux dans la création du TIN. Dans les vallées, suite à l'existence des triangles horizontaux, la hauteur donnée par le MNH-TIN est plus petite que celle donnée par le MNH-ERRUI et inversement en contexte de sommet ou de ligne de crête.

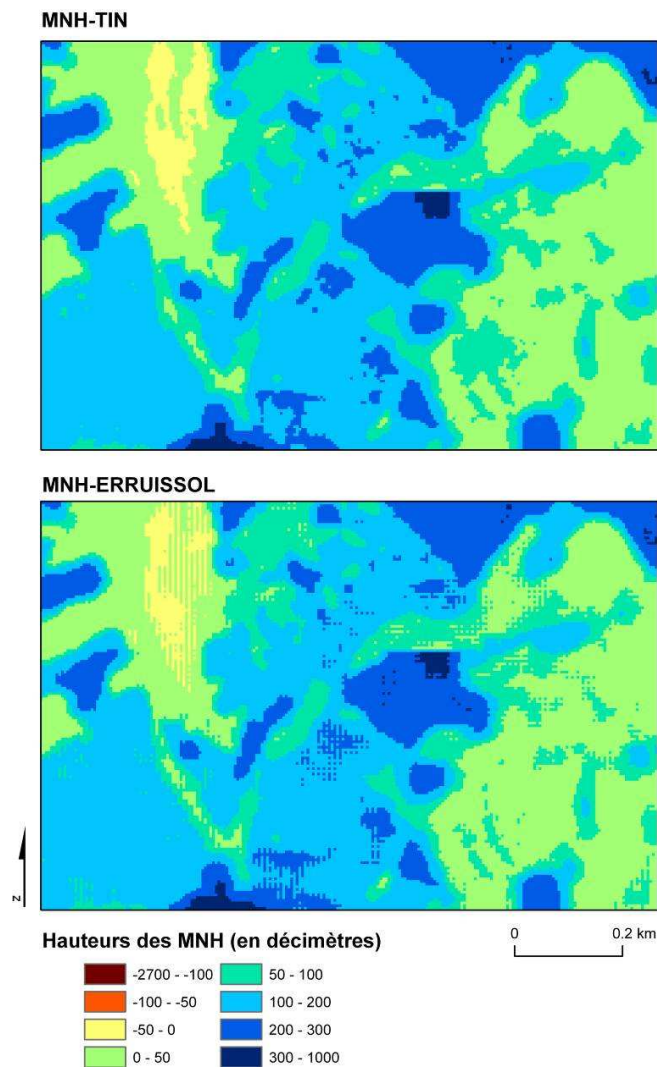


Figure 54. Différences entre les hauteurs données par les 2 MNH dues à la procédure de rééchantillonnage au plus proche voisin à 5 m de résolution du MNT-ERRUI et à la précision décimétrique du calcul des hauteurs

d) Analyse des différences entre les 2 MNH dans les zones où les données du PICC ont été utilisées pour la conception des MNT

Dans les zones où le PICC a été utilisé comme donnée d'altitude, les données du PICC utilisées pour la conception du TIN et du MNT ERRUISSOL ne sont pas toujours les mêmes, car :

- les **types de points** pris en compte ne sont pas les mêmes, pour le TIN, tous les points au sol ont été considérés (Tableau 20), alors que pour le MNT ERRUISSOL, seuls certains points ont été pris en compte (Tableau 19).
- le PICC a été **actualisé** depuis la production du MNT ERRUISSOL, et dans le TIN ont été considérés des points du PICC qui n'existaient pas au moment de la production du MNT ERRUISSOL,
- les données IGN ont été ajoutées dans les « trous » ou zones de faibles densités du PICC, mais ces « trous » peuvent avoir été délimités de façon différente dans les procédures respectives de création du MNT ERRUISSOL et du MNT-TIN.

Les différences remarquées entre les MNH seront donc principalement dues à des données d'**altitude différentes** et seront probablement plus denses dans le cas du MNT-TIN.

Des incohérences très importantes entre les MNH ont été remarquées là où les types de points du PICC utilisés ne sont pas les mêmes dans le MNT-TIN et dans le MNT ERRUISSOL, notamment, dans les carrières (Figure 55) et au niveau des routes (Figure 56). En effet, dans la production du MNT ERRUISSOL, les points de niveau de voirie n'ont pas été pris en compte, ce qui explique les différences au niveau des routes, ni certains types de points comme les axes de ruisseaux, axes de fossés, bords de lacs et étangs, bords de carrières et excavations diverses, ce qui explique les différences dans les carrières.

Dans le pavé IGN n°54 et dans une partie des pavés n°48, 49 et 50, lors de l'élaboration du MNT ERRUISSOL, les données du PICC n'existaient pas. Pour la production du MNT ERRUISSOL, c'étaient donc les courbes de niveau IGN qui avaient été utilisées. Pour la production du TIN, les données du PICC actualisé ont été utilisées là où elles existent, dans une bonne partie de ce pavé IGN. On remarque donc des différences de l'ordre de 1 à 5 m entre les MNH dans ces pavés (Figure 57). **L'actualisation** du MNT et du MNH au fur et à mesure de l'actualisation du **PICC** pourrait donc être envisagée.

Dans les zones où les points du PICC ont été utilisés pour le TIN et le MNT ERRUISSOL, les **types de points** considérés dans le premier mais pas dans le second cas améliorent la qualité de la représentation de l'élévation (Figure 58).

On remarque aussi, dans les zones où le PICC a été utilisé, que les différences entre hauteurs données par les MNH sont plus importantes dans les fortes pentes (Figure 59). On remarque par ailleurs que les données du PICC permettent de mieux représenter les très hautes fréquences liées aux points structuraux au sol utilisés pour créer le TIN mais pas pour produire le MNT-ERRUI.

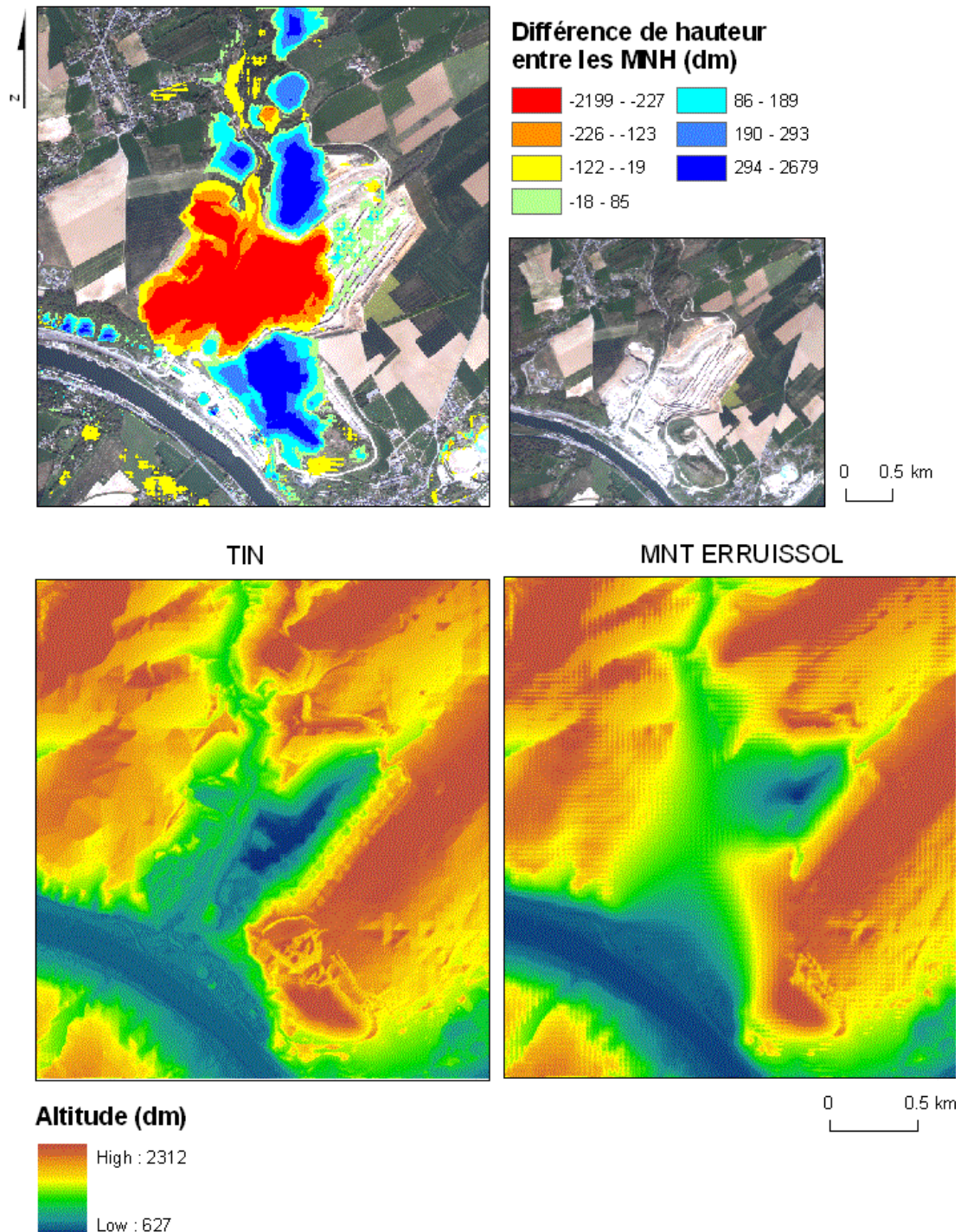


Figure 55. Incohérences importantes entre les deux MNH dans une carrière

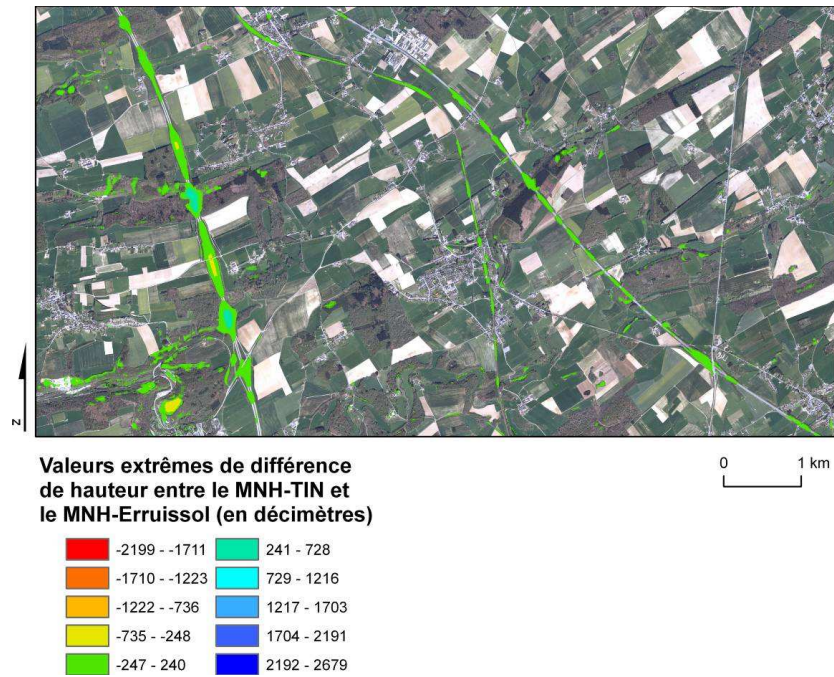


Figure 56. Incohérences entre les MNH au niveau des routes

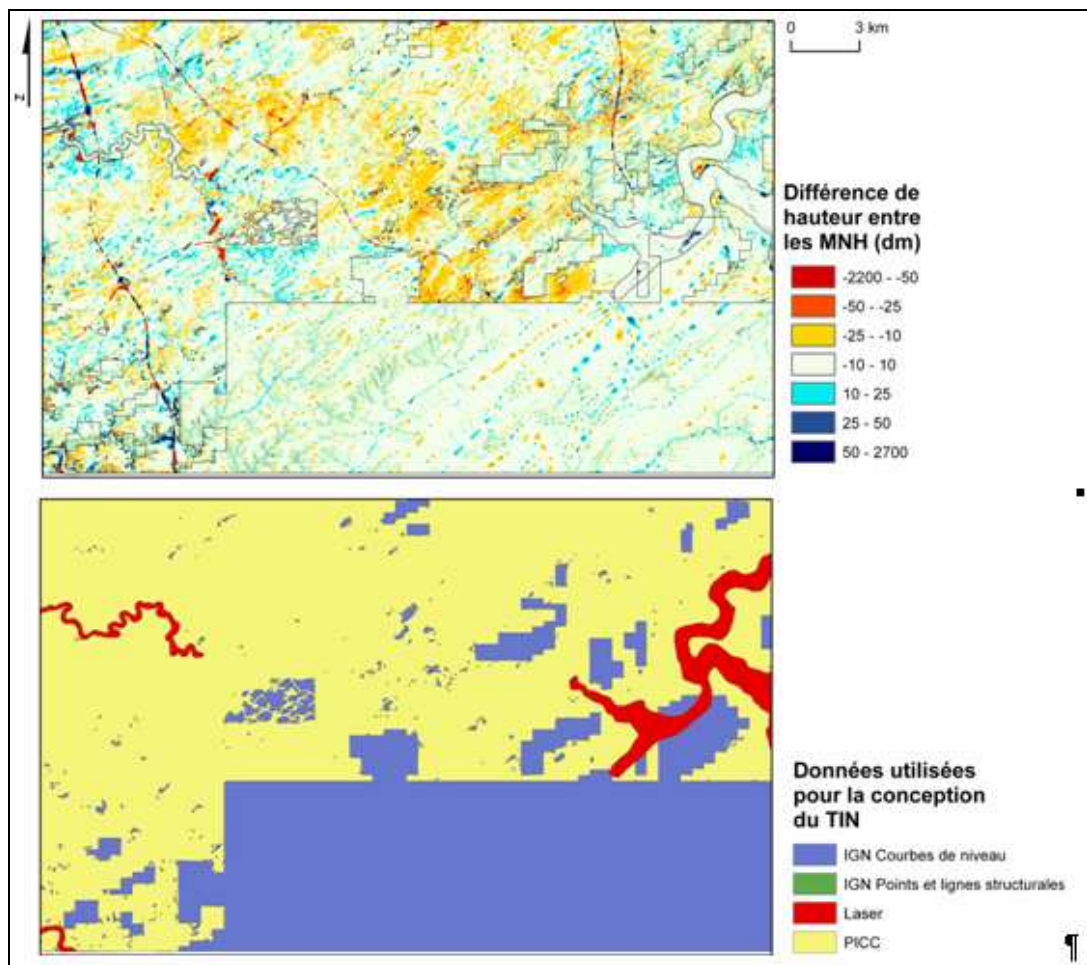


Figure 57. Incohérences de l'ordre de 1 à 5 m plus fréquentes entre les MNH, dans le pavé IGN n°54, là où les données du PiCC actualisé ont été utilisées pour la conception du TIN, alors que pour le MNT ERRUISSOL, les courbes de niveau IGN avaient été utilisées

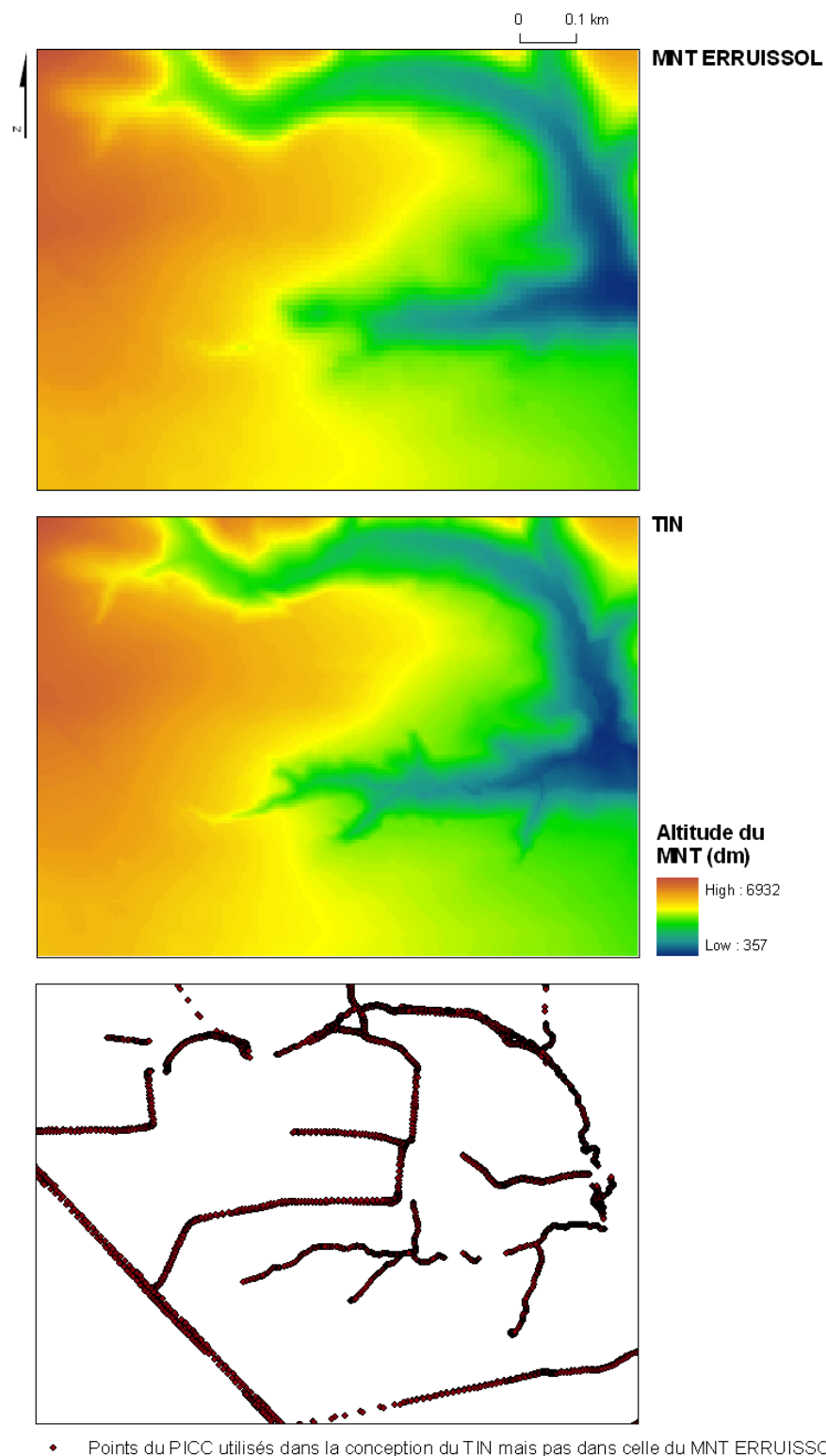


Figure 58. Elévation mieux représentée par le TIN que par le MNT ERRUISSOL dans une zone où les points PICC ont été pris en compte, et où les types de points PICC qui ont été considérés dans la conception du TIN mais pas dans la conception du MNT ERRUISSOL permettent d'améliorer la qualité structurale de la surface représentée.

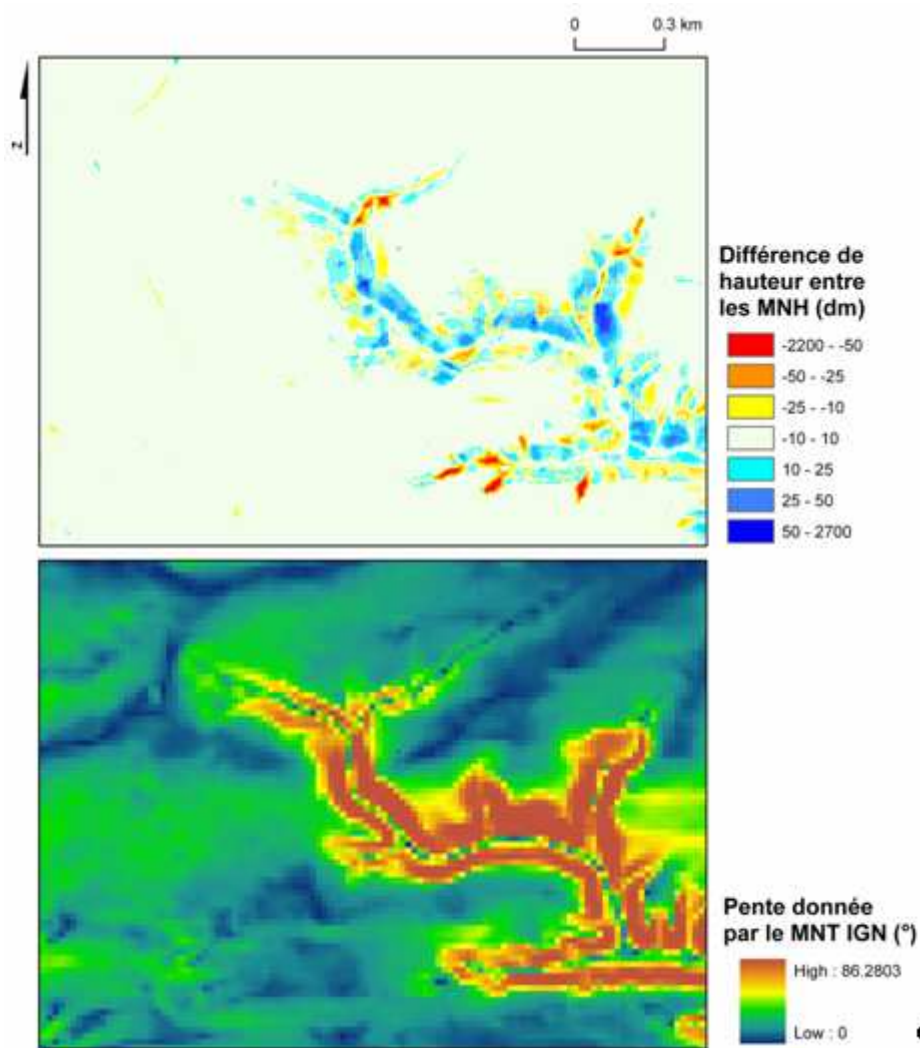
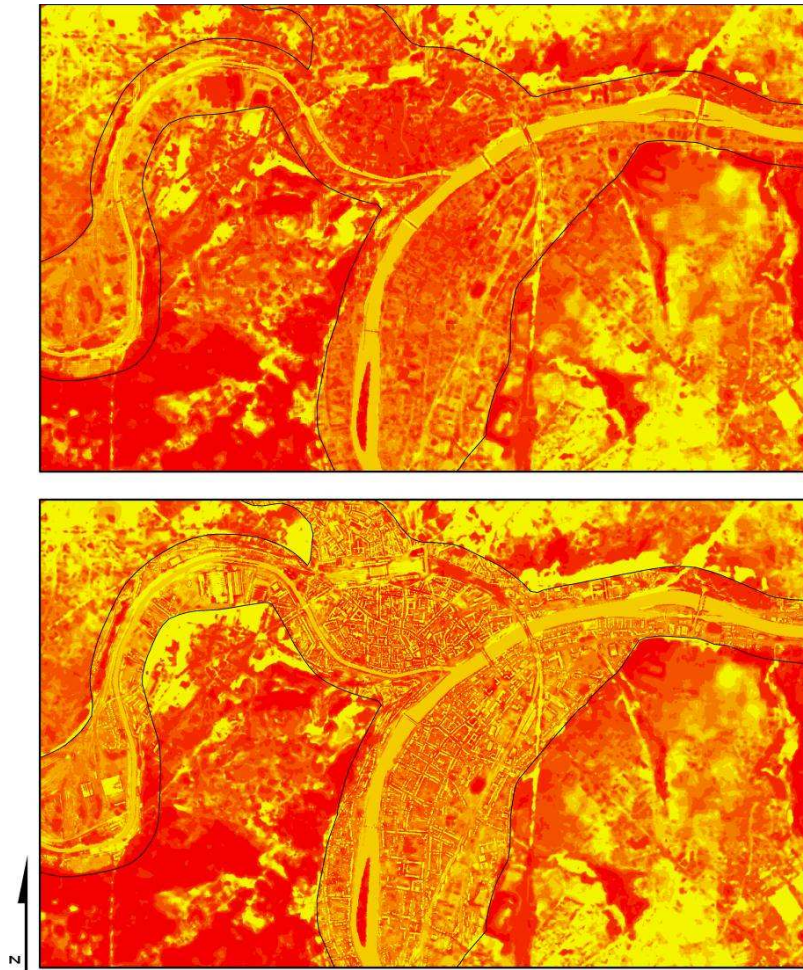


Figure 59. Différences entre les hauteurs données par les MNH plus importantes dans les zones de fortes pentes

e) Analyse des différences entre les 2 MNH dans les zones où le MNT/MNS laser a été utilisé pour la conception des MNT

Dans les zones bâties, les incohérences sont dues au MNT/MNS laser utilisé (Figure 60) : pour le TIN, on a utilisé le MNT/MNS laser **tel quel**, alors que pour le MNT ERRUISSOL, ce MNT laser a d'abord été **corrigé**, en supprimant les données d'altitude qui ne représentent pas des données au sol. En effet, le deuxième écho du laser, représentant approximativement l'altitude au sol dans les zones forestières, est stoppé s'il rencontre une surface qui lui est opaque, tel un bâtiment. Il ne représente donc pas l'altitude au sol dans les zones bâties. Le MNT/MNS laser corrigé pour la conception du MNT ERRUISSOL est donc plus représentatif de l'altitude au sol que le MNT/MNS laser utilisé tel quel.



Différence d'altitude entre le MNS et le MNT Erruissol (en haut) ; le MNS et le TIN (en bas), dans une zone où les données du MNT laser ont été utilisées pour la production des MNH

0 1 km

□ Limite de la zone où le MNT laser a été utilisé

Différence d'altitude (décimètres)

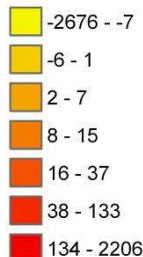


Figure 60. Incohérences entre les MNH dues à l'utilisation d'un MNT/MNS laser corrigé ou non. Dans le MNT-ERRUISSOL, le MNT/MNS laser a été corrigé par élimination des bâtiments, alors que dans le TIN, il a été utilisé tel quel. Les lignes noires sur la carte délimitent les zones où le MNT laser existe. Remarquons que les limites des classes de hauteur (différence d'altitude entre MNS et MNT) ont été établies globalement sur le territoire du bloc 6 sur le MNH-TIN par construction de classes de même fréquence.

2.3.2. Evaluation de l'apport des MNH dans les procédures de classification

L'utilisation d'un MNH peut intervenir pour **lever plusieurs confusions** dans les procédures de classification :

- distinction entre zones forestières, ouvertes et/ou agricoles présentant des signatures spectrales similaires
- appui à la discrimination entre types de surfaces agricoles (au sens de la légende de la CNOSW, telles que surfaces enherbées vs vergers),
- détection des alignements d'arbres et si possible des haies,
- analyse de la structure en hauteur des zones forestières comme complément à l'information spectrale intra- et inter- essences (fonction de l'âge).

Dans une seconde approche, le MNH peut être utilisé pour une analyse plus fine en vue de l'estimation de la **hauteur du couvert forestier** par la mise au point de modèles prédictifs. Une démarche de ce genre, menée par l'unité GRFMN dans le cadre d'un autre projet, a été présentée brièvement dans le rapport précédent (cf. Bonnet S., 2009). L'intégration du MNH dans la démarche vient concrètement en appui dans les tâches 6 et 7.

Nous avons réalisé quelques analyses comparatives afin de déterminer l'apport réel d'un MNH dans les procédures de classification.

2.3.2.1. Analyse des différences des hauteurs données par les MNH dans des zones de résineux

a) A partir d'objets de référence

Afin d'évaluer l'utilité des MNH et de les comparer, nous avons réalisé une analyse des différences des hauteurs données par les MNH dans des zones de **résineux**. Nous exploitons les objets de référence de résineux digitalisés par le GRFMN dans le bloc 6. On dispose d'informations utiles comme les essences et l'année de plantation pour certaines parcelles.

Les **statistiques zonales** des hauteurs données par le MNH-TIN et le MNH-ERRUI ont été calculées pour chaque parcelle.

Nous avons d'abord comparé les **moyennes par parcelle** calculées à partir des deux MNH. Pour la plupart des parcelles, elles sont semblables. Pour une parcelle en particulier, on constate une **différence** de 7 m. En examinant cette parcelle de plus près, on se rend compte qu'elle se trouve dans une zone où l'on a utilisé le PICC actualisé pour la conception du TIN, alors que seules les courbes de niveaux IGN avaient été utilisées pour la conception du MNT ERRUISSOL.

D'autres parcelles montrent également une petite différence liée à la méthode d'interpolation utilisée pour créer le MNT-ERRUI ou à la création de triangles horizontaux dans des zones où le tracé des courbes de niveau IGN utilisées pour produire le TIN est favorable à l'apparition de tels triangles, soit dans des zones de crête ou de sommet soit dans les vallées (Figure 61). Cependant, cette différence n'est que de l'ordre de 1 à 2 m.

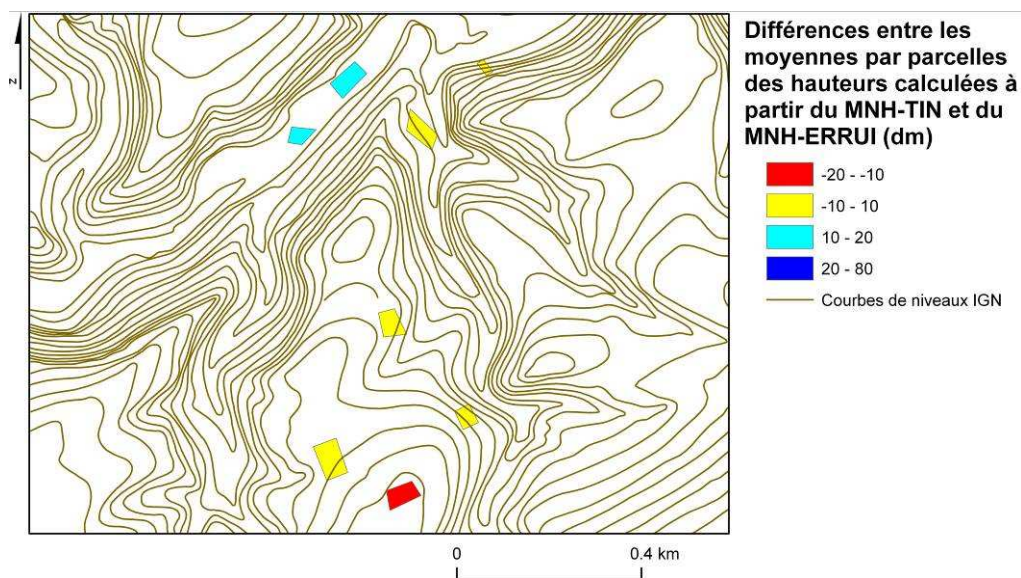


Figure 61. Localisation des parcelles pour lesquelles on observe une différence de l'ordre de 1 à 2 m entre les moyennes par parcelle des hauteurs calculées à partir du MNH-TIN et à partir du MNH-ERRUI. Les erreurs sont liées à la présence de triangles horizontaux dans le TIN là où le tracé des courbes de niveau de l'IGN utilisée est favorable à leur apparition.

Afin de vérifier si la hauteur donnée par le MNH fournit une information utile sur un type de couvert forestier donné, nous avons analysé les moyennes par parcelle des hauteurs données par les MNH en fonction de l'année de plantation, pour les parcelles d'épicéas dont nous disposons de l'année de plantation. Il ressort une proportionnalité intéressante entre la **hauteur** et **l'année de plantation** (Figure 62). Les quelques parcelles pour lesquelles ce n'est pas le cas sont en réalité des parcelles où l'on a un mélange d'épicéas et d'autres essences. Les moyennes par parcelle des hauteurs calculées à partir du MNH-TIN et du MNH-ERRUI sont pratiquement semblables pour la majorité des parcelles.

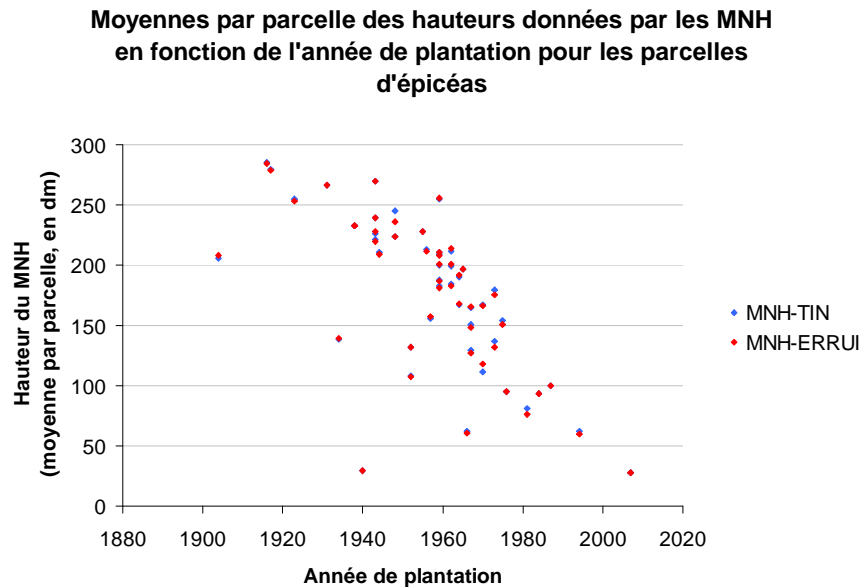


Figure 62. Relation entre les moyennes par parcelle des hauteurs données par les MNH et l'année de plantation, pour les parcelles d'épicéas

b) A partir de mesures ponctuelles de hauteur d'arbres

Nous avons comparé les hauteurs données par les MNH avec des **hauteurs** de résineux **mesurées sur le terrain**. Les hauteurs utilisées comme vérité terrain ont été mesurées par l'unité GRFMN de la GxABT, dans le cadre d'un projet financé par la Division Nature et Forêt (Accord-Cadre « Recherche et Vulgarisation Forestières 2004 – 2009 » - Rapport d'activité Juin 2008). Dans le bloc 6, nous disposons de mesures dans la Donation Royale de Ciergnon et à Elsenborn.

Pour tous les points où nous disposons de mesures de hauteur d'arbres dans le bloc 6, nous avons comparé la hauteur donnée par le MNH-ERRUI à la hauteur mesurée sur le terrain. En général, les **hauteurs** sont **sous-estimées**.

Plusieurs paramètres interviennent dans l'explication de cette sous-estimation :

- le MNS généralise la surface enveloppe du sommet des arbres à une résolution de 5 m,
- le MNS n'est pas fiable partout,
- le MNT n'est pas une représentation fidèle de la surface du sol, surtout dans les zones où il a été produit à partir des courbes de niveau IGN,
- les points de mesure ont été localisés avec un GPS peu performant et sous couvert forestier, il peut donc y avoir des erreurs de localisation des points mesurés.

Même si les hauteurs données par le MNH ne correspondent pas directement aux hauteurs des sommets des arbres mesurés, Bonnet S. (2009) a montré que l'on pouvait établir une **relation linéaire** empirique entre le MNH et la hauteur dominante.

Nous avons également remarqué que les zones situées en **bordure de forêt** sont souvent entachées d'une erreur qui peut être très importante. Ces zones de lisières de forêt avaient déjà été mises en évidence dans le contrôle de la qualité du MNS (DGA et Unité de Géomatique, 2008), mais elles sembleraient être beaucoup plus larges que prévu à certains endroits.

2.3.2.2. Analyse dans des zones agricoles

Quelques tests ont également été réalisés dans des zones agricoles. Dans la majorité de ces zones, la hauteur des objets devrait être nulle ou comprise entre 0 et 2 à 3 m d'altitude. Or, en réalisant des profils topographiques, nous remarquons à de nombreux endroits des **valeurs négatives** de hauteur. La plupart sont de l'ordre de -2 à -3 m, mais certaines valeurs peuvent atteindre une dizaine de mètres ou plus.

Face à ce constat, nous recommandons de traiter le MNH dans les procédures de classification avec une grande prudence et de ne pas l'utiliser dans les processus de segmentation.

Dans le but d'évaluer son apport dans les procédures de classification par objet, après la segmentation, nous avons calculé la hauteur moyenne par parcelle, dans des objets de référence agricoles. 351 objets de référence dans les zones agricoles du bloc 6 ont été ajoutés (cf. 2.4.3.2). Il s'agit de parcelles du SIGEC, érodées d'une zone tampon intérieure de 15 m (buffer négatif), afin d'éliminer les erreurs cartographiques et les effets de bords éventuels, qui ont été sélectionnées sur base de l'observation de l'ORTHOS_DGA en vraie couleur.

Des parcelles homogènes et bien représentatives ont été choisies pour diverses affectations SIGEC. La hauteur moyenne dans ces parcelles (Figure 63) montre à nouveau des incohérences parfois importantes à certains endroits. On observe une importante proportion de valeurs négatives jusqu'à -15 m ou plus. La plupart des erreurs sont tout de même de l'ordre de 2 à 3 m. Des incohérences plus légères sont remarquées également là où les courbes de niveau IGN ont été utilisées dans la conception du MNT, à des endroits où le tracé de ces courbes de niveau est favorable à l'apparition de triangles horizontaux sur le TIN.

Les erreurs importantes sont dues à des imperfections dans la recherche automatique de points homologues lors de la production du MNS_DGA.

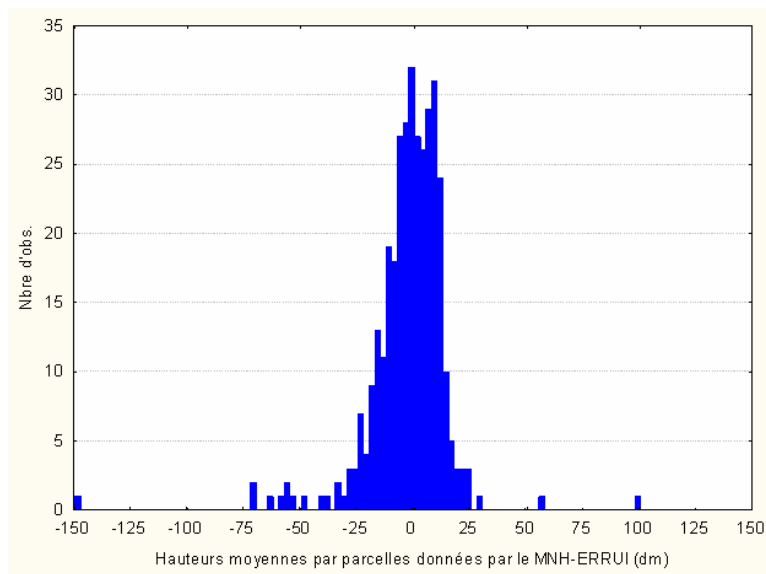


Figure 63. Hauteurs moyennes par parcelle des objets de référence agricoles dans le bloc 6

2.3.2.3. Conclusion

Le MNH peut être utilisé dans les procédures de classification par objet comme couche complémentaire pour aider à distinguer certaines occupations du sol, mais il n'est pas assez précis pour être utilisé dans une procédure de segmentation proprement dite. En effet, les valeurs sont souvent erronées en lisière de forêt, parfois sur une superficie relativement importante.

L'incertitude des hauteurs données par le MNH pour les zones agricoles est de l'ordre de 2 à 3 m pour une majorité de parcelles. Elle peut être due au MNT, lorsqu'il a été produit à partir de courbes de niveau IGN.

Localement, on retrouve pour certaines parcelles des incohérences spatiales très importantes par rapport aux parcelles voisines, de l'ordre de la dizaine ou vingtaine de mètres. Elles sont probablement dues à une erreur dans la recherche automatique de points homologues lors de la production du MNS_DGA et apparaît donc sur les 2 MNH produits à partir du MNT-TIN ou du MNT-ERRUI.

Dans les zones forestières, la hauteur donnée par le MNH n'est pas la hauteur du sommet des arbres, mais une hauteur moyenne par pixel de 5 m de côté. Cette hauteur permet cependant, dans certains cas de peuplements de résineux, d'estimer la hauteur de la canopée à l'aide d'un modèle numérique de la canopée (Bonnet S., 2009). Cependant, dans les zones à la lisière des forêts, et ce dans une bande assez large, on peut retrouver des erreurs importantes, de l'ordre de plusieurs dizaines de mètres. L'utilisation du MNH dans les procédures de classification est donc à considérer avec prudence.

2.4. Préparation d'une base de données d'objets de référence (GRFMN)

2.4.1. Objectif

L'objectif de cette partie de l'étude est de constituer une collection d'objets de référence susceptibles d'être exploités dans les différentes étapes de mise au point des procédures d'exploitation des couches ORTHOS_DGA et satellitaires (segmentation, classification, validation, ...). Ils ont pour vocation de servir de « vérité terrain » et de zones d'entraînement pour construire et/ou valider les règles de segmentation - classification. Des objets de référence peuvent aussi être utilisés dans une approche par pixel.

2.4.2. Méthodologie

Les objets de référence sont de type « polygone » mais peuvent être créés à partir de données initiales « point » ou « polygone ». Ces points ou polygones initiaux sont géoréférencés et caractérisés par une information sémantique décrivant la situation « terrain » pour cette localisation. L'objet de référence produit à partir de ces données initiales correspond à la délimitation (digitalisation), sur base de la photo-interprétation des ORTHOS_DGA, d'un élément du paysage décrit par une information sur l'« occupation du sol » (pessière de 60 ans, parcelle de froment d'hiver,...).

Les sources de données utilisées pour alimenter cette base de données sont :

- les bases de données « SIGEC_2006 » et « SIGEC_2007 » (contiennent un inventaire annuel du parcellaire et des superficies agricoles) ;
- les parcellaires forestiers produits par les services extérieurs du Département Nature et Forêts ;
- la base de données « IPRFW » (Inventaire Permanent des Ressources Forestières de Wallonie).

Les objets de référence créés sont stockés dans une géodatabase. Chaque objet est décrit par un ensemble d'attributs. Les objets de référence forestiers (un exemple en figure 64) sont caractérisés au minimum par le type de peuplement (feuillus, résineux ou mixte) et, quand l'information existe, par des champs précisant l'essence (jusqu'à trois essences possibles), les années de plantation respectives, leurs pourcentages, la source de l'information, l'exposition et la pente du peuplement.

Les objets agricoles sont décrits par les attributs disponibles à partir du SIGEC. Il s'agit principalement, pour une année donnée, de la culture présente sur une parcelle (Figure 65).

La création d'objets de référence repose sur un travail de photo-interprétation (correspondance image - attributs) et de digitalisation à partir d'une des trois sources citées précédemment.

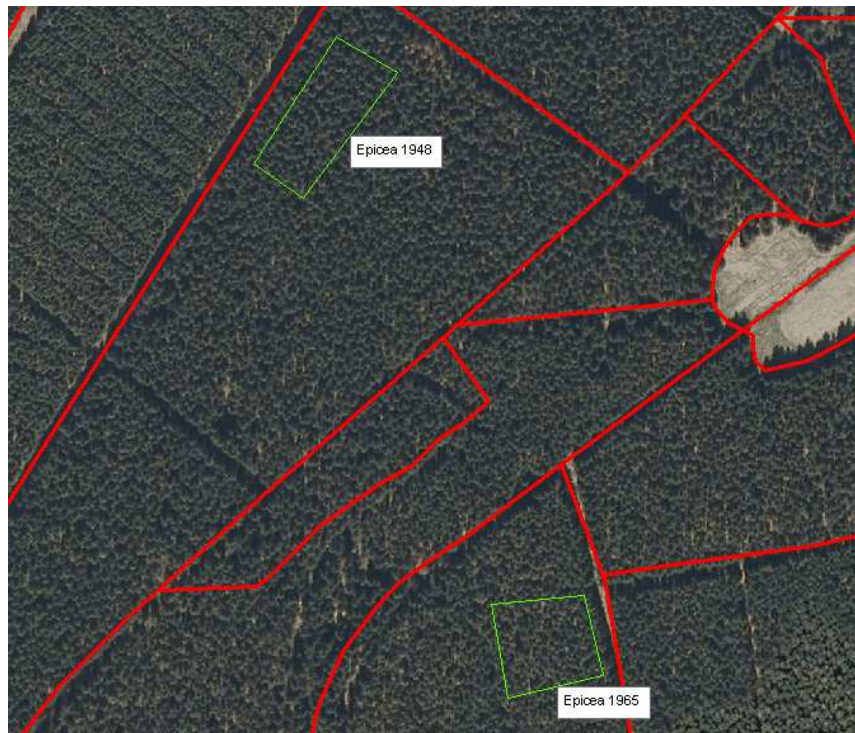


Figure 64. Utilisation du parcellaire forestier (cantonement de Houffalize, limites en rouge) comme appui à la digitalisation d'objets de référence en zone forestière.



Figure 65. Extrait du SIGEC_2007 fournissant le parcellaire agricole, source de données pour l'extraction d'objets de référence en zones agricoles.

2.4.3. Types d'objets de référence

Deux types d'objets de référence sont créés, en fonction de l'étape de la procédure où ils peuvent intervenir : les objets « géométriques » (établissement des règles de segmentation et de classification) et les objets « spectraux » (études radiométriques diverses). La création de ceux-ci est décrite aux paragraphes suivants.

2.4.3.1. Objets « géométriques »

a) Description

Ces objets ont une double vocation : tout d'abord, ils permettent l'étude de la qualité de la segmentation via la comparaison quantitative – à l'aide d'indices de qualité (cf. paragraphe 2.5.3.5) – des limites de ces objets avec celles des résultats de la segmentation.

Ensuite, ils peuvent, en grâce à l'information thématique détaillée qui leur est attribuée à la création, intervenir dans le processus d'établissement des règles de classification.

b) Sélection d'îlots de la carte d'occupation des sols existante et alimentation de la base de données

L'idée de s'appuyer sur des unités constitutives de la carte d'occupation des sols produite dans le cadre du projet CNOSW pour procéder aux phases de segmentation et de classification a été émise. Ces unités de base sont des « îlots », un îlot étant constitué d'un groupe de parcelles, bâties ou non, entièrement délimité par de l'espace non cadastré. Ceci se justifie notamment par un souci de cohérence et d'intégration de la démarche d'analyse dans la méthode de production de la CNOSW ainsi que par une meilleure « signification » des limites des unités de base traitées. En effet, le traitement par tuile d'ortho-image génèrerait des limites géométriques artificielles qu'il ne serait pas forcément évident de gérer. C'est donc sur base des îlots que les tests de segmentation et de mesure de qualité de celle-ci ont été menés (cf. paragraphe 2.5).

Sur base de l'expérience acquise au cours de cette convention, il n'est toutefois pas exclu d'envisager dans une étude ultérieure un autre critère de découpage pour aborder le traitement et l'analyse des ortho-images. Parmi les différentes solutions envisageables, on peut envisager :

- un regroupement d'îlots ;
- les pavés images proprement dits ;
- un regroupement de pavés ;
- les polygones correspondant au découpage des *seamlines* (disponibles pour les ortho-images 2009).

Dans le cadre de cette démarche, afin d'alimenter la base de données d'objets de référence en objets « géométriques », des îlots de la carte d'occupation du sol existante ont été sélectionnés dans les blocs « juin-juillet 2006 » et leur occupation du sol interne délimitée et caractérisée via photo-interprétation.

Les îlots ont été sélectionnés en respectant une certaine typologie d'occupation du sol (Tableau 26) afin de rencontrer différentes situations suffisamment contrastées pour analyser les possibilités de discrimination dans les phases de test.

Tableau 26. Typologie retenue pour la sélection des îlots à partir de la CNOSW

Code	Type
1	Mélange feuillus/résineux avec différences de hauteurs
2	Mélange feuillus/résineux sans différences de hauteurs
3	Feuillus uniquement
4	Résineux uniquement
5	Feuillus et zones agricoles
6	Résineux et zones agricoles
7	Mélange feuillus/résineux et zones agricoles
8	Zones agricoles uniquement

Tableau 27. Nombre d'îlots par type sélectionnés, première phase de collecte

Type	Nombre d'îlots
1	10
2	8
3	9
4	9
5	7
6	9
7	10
8	9
Total	71

Une première sélection et caractérisation d'îlots a été opérée. Le nombre d'îlots sélectionnés par type est présenté au tableau 27. Les essences forestières dominantes sont listées au tableau 28.

Tableau 28. Liste des principales essences forestières représentées dans les îlots sélectionnés

résineux	DO	Douglas
	EP	Epicéa
(année de plantation de 1920 à 2006)	MZ	mélèze
	PS	Pin sylvestre
feuillus	CA	Charme
	CH	Chêne
	HE	Hêtre
	ME	Merisier
	FD	Feuillus divers

L'analyse visuelle des ORTHOS_DGA s'est appuyée notamment sur les parcellaires forestiers à disposition de l'unité GRFMN ainsi que sur le SIGEC 2006. Lors de la sélection, chaque polygone représentant les îlots a été divisé en zones d'occupation du sol homogènes selon la typologie du tableau 26 (cf. figure 67).

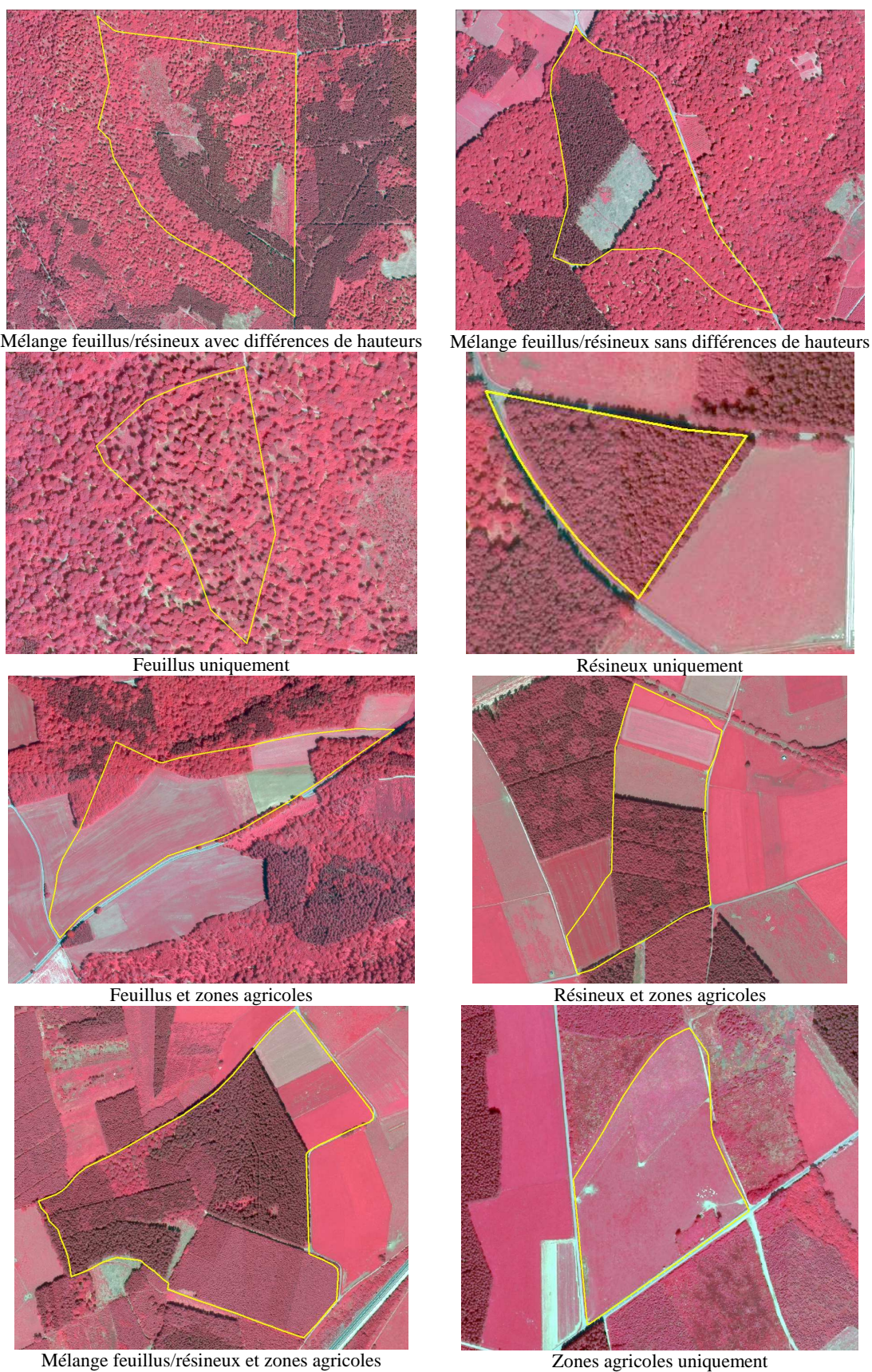


Figure 66. Illustration des 8 types d'îlots sélectionnés



Figure 67. Exemple d'îlot divisé en zones homogènes d'occupation du sol

Notons qu'une contrainte supplémentaire a également été fixée lors de la sélection des îlots. Dans le cadre de cette étude, différentes images SPOT 5 ont été utilisées afin de disposer de données complémentaires d'origine satellitale et présentant en outre un caractère multi-dates. Les tableaux 29 et 30 en rappellent les principales caractéristiques.

Tableau 29. Synthèse des caractéristiques des images SPOT 5 commandées

Références image	50462490510121026481J	50462490607191042132J
Satellite	SPOT5	SPOT5
K/J	46 / 249	46 / 249
Date d'acquisition	12-10-2005	19-07-2006
Résolution spatiale	10 m C	10 m C
% Nuages	0 %	0 %
Notation nuageuse	AAAAAAA	AAAAAAB
Décalage	0	3
Centre	N49°54 E5°41	N49°44 E5°50
Angle d'incidence	-24.72°	-4.18°
Nb bandes spectrales	4	4
Commande	Niv. 3 Orthorectifiée avec points d'appuis et DSM fournis par le client - Lambert Belge 1972	Niv. 3 Orthorectifiée - Lambert Belge 1972
	Niv. 3 Orthorectifiée - Lambert Belge 1972	
	Niv. 1B	

Tableau 30. Synthèse des caractéristiques des images SPOT 4 commandées

Références image	40442480705011112062I	40462480705011112041I
Satellite	SPOT4	SPOT4
K/J	44 / 248	46 / 248
Date d'acquisition	01-05-2007	01-05-2007
Résolution spatiale	20 m C	20 m C
% Nuages	0 %	0 %
Notation nuageuse	AAAAAAAA	AAAAAAAA
Décalage	1	3
Centre	N50°19 E5°16	N50°13 E5°56
Angle d'incidence	26.15°	29.27°
Nb bandes spectrales	4	4
Commande	Niv. 3 Orthorectifiée - Lambert Belge 1972	

Comme le montre la figure 68, deux scènes SPOT 5 (octobre et juillet) sont disponibles dans le bloc «juin-juillet 2006». Les deux images se superposent en une zone commune dans ce bloc (cf. figure 68). Il nous a semblé intéressant de concentrer l'effort de sélection des îlots dans cette zone commune aux deux images SPOT 5. Ainsi, 43 îlots sur les 71 sélectionnés se trouvent dans cette zone que nous appellerons « zone SPOT » dans la suite du rapport pour plus de facilité.

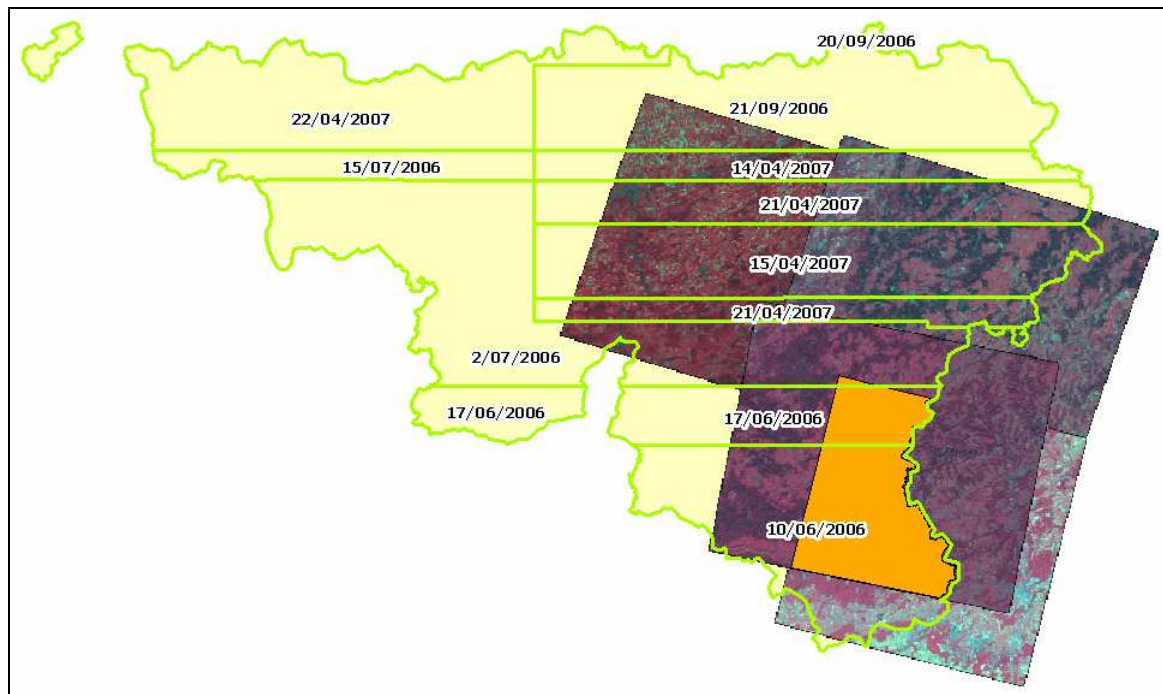


Figure 68. Localisation des scènes SPOT 4 et SPOT 5 commandées, limites des blocs d'ortho-images et dates de vols et en orange, zone commune aux deux scènes SPOT 5 disponibles dans le bloc « juin-juillet 2006 »

En résumé, pour analyser les possibilités de segmentation – classification dans les îlots de cette zone, nous disposons des ORTHO-DGA, du MNH, de l'image SPOT 5 d'octobre, de l'image SPOT 5 de juillet.

Une deuxième phase de sélection d'îlots a permis d'étayer les données de « vérité terrain » disponibles dans cette même zone.

Nous disposons donc au total dans la zone SPOT de 123 îlots répartis spatialement tel que présenté en figure 69. La superficie de cette couche constituant la « vérité terrain » est de 43,7 km² (cf. tableau 31).



Figure 69. Localisation du set d'îlots sélectionnés dans la zone "SPOT"

Tableau 31. Superficie couverte par les différentes classes d'occupation du sol, set d'îlots zone « SPOT »

Occupation	Surface (km ²)	pourcentage
Agricole	13.0	29.7
Autre	0.4	0.9
Feuillus	13.5	30.9
Mixte	2.2	5.0
Résineux	11.7	26.8
Vide	2.9	6.6
Total	43.7	100

2.4.3.2. Objets « spectraux »

a) Description

Tout au long de l'étude, des objets spectraux ont été créés à différentes fins : contribution à l'étude de la qualité radiométrique de la couche ORTHOS_DGA (paragraphe 2.1), contribution à l'étude du MNH en zone agricole (2.3.2) et enfin, contribution à l'analyse des signatures spectrales des différentes classes qui interviennent dans le processus de classification. Toutefois, dans ce dernier cas, les objets géométriques leur ont été préférés dans la phase d'élaboration des règles de classification par souci de cohérence avec la chaîne de traitement globale (segmentation et classification). Ceci sera détaillé au paragraphe 2.7.

Ces objets spectraux se divisent en deux groupes :

- des polygones délimitant des zones considérées comme homogènes en termes d'occupation du sol (cf. figures 70 et 71)
- des objets représentant une seule occupation mais contenant des éléments « contextuels » amenant de la variabilité dans la signature spectrale. Par éléments « contextuels », il faut entendre des zones de trouées dans des peuplements forestiers, des layons d'exploitation dont l'orientation peut influencer la texture du segment (cf. figure 72),... Ces objets sont représentatifs de la situation que l'on peut rencontrer dans une parcelle.



Figure 70. Exemple d'objets de référence spectraux et homogènes, digitalisés à partir d'un parcellaire DNF.



Figure 71. Exemples d'objets de référence spectraux et homogènes, digitalisés à partir du SIGEC 2006



Figure 72. Exemples d'objets de référence contextuels avec présence de layons d'exploitation pouvant influencer la texture.

b) Alimentation de la base de données

L'alimentation de la base de données d'objets de référence peut être considérée comme une tâche de fond, comme illustré précédemment pour les objets « géométriques ». En ce qui concerne les objets « spectraux », différentes phases de collecte d'objets ont été réalisées.

Une première phase de création d'objets de référence a été mise en place pour venir en appui à la tâche 1. A cet effet, une série d'objets « spectraux » avaient été digitalisés dans le bloc « avril 2007 » pour lequel les *seamlines* sont disponibles. Pour ce bloc « avril 2007 », il a été décidé de se limiter aux zones agricoles et aux forêts résineuses. Le cas de forêts feuillues a volontairement été écarté en raison de la date d'acquisition des données; la plupart des essences feuillues n'étant pas encore « en feuilles », les possibilités de discrimination sont limitées. Pour ce bloc, des objets ont été créés à partir:

- des parcellaires DNF pour les forêts communales de Houffalize et de Manhay ;
- des parcellaires de la Donation Royale (Villers-sur-Lesse) ;
- de l'IPRFW.

Une seconde phase de création d'objets de référence « spectraux » a ensuite été menée afin d'enrichir la base de données pour les blocs « juin-juillet 2006 ». En effet, le bloc « avril 2007 » est problématique dans le cas des forêts feuillues puisque les possibilités de discrimination sont limitées par l'absence de feuillage. Un des objectifs de l'étude est d'analyser les possibilités de discrimination de différentes classes d'occupation du sol. A cette fin, la première étape est de commencer par distinguer de grandes catégories clés d'occupation du sol telles que « non-forêt/forêt » et, dans le dernier cas, « feuillus/résineux ». Les distinctions entre classes s'affinent au fur et à mesure du processus de classification. Pour satisfaire à cette étape exploratoire, les blocs « juin-juillet 2006 » sont donc préférentiellement choisis pour la sélection de zones de référence et de tests.

Une troisième phase de création d'objets de référence « spectraux » en zone agricole a été menée pour les besoins de l'analyse du MNH. Dans ce cadre, 351 objets de référence ont été ajoutés dans les zones agricoles du bloc 6. Il s'agit de parcelles du SIGEC, érodées d'une bande de 15 m (*buffer* négatif), afin d'éliminer les erreurs cartographiques et les effets de bords éventuels, qui ont été sélectionnées sur base de l'observation de l'ORTHOS_DGA en vraie couleur.

Une quatrième phase de création d'objets de référence « spectraux » a été réalisée dans la zone commune aux deux scènes SPOT 5 (cf. figure 68).

2.5. Mise au point des procédures de segmentation d'images (GRFMN)

2.5.1. Introduction

L'analyse orientée-objet, dont la segmentation est une des étapes, prend de plus en plus d'importance dans le contexte de la mise à disposition de données de télédétection à très haute résolution (Blaschke, 2010).

Dans le cadre de cette étude, nous nous sommes focalisés sur deux logiciels commerciaux parmi les différentes solutions de segmentation existantes : Definiens Developer 8.1 et ENVI 4.7 couplé au module ENVI EX (cf. paragraphe 2.5.2). Pour chacun d'eux, les procédures de segmentation reposent sur un algorithme particulier, lui-même défini par des paramètres. Ces paramètres sont utilisés pour découper une image et peuvent influencer tantôt sur la taille des objets, tantôt sur l'importance donnée à la forme et aux caractéristiques spectrales des objets créés.

La mise au point des procédures de segmentation (i.e. la fixation des valeurs des paramètres des algorithmes de segmentation) dont les objets résultant seront injectés dans la phase de classification est la phase centrale de la tâche 5. Traditionnellement, cette mise au point se base sur une évaluation par essais-erreurs. Cette méthode est assez subjective et dépendante de l'expérience de l'opérateur. Afin de rendre le processus plus objectif et de traiter des quantités de résultats plus importantes (multiplier les combinaisons testées), une démarche quantitative d'évaluation de la qualité de la segmentation a été utilisée.

2.5.2. Solutions logicielles pour la segmentation

2.5.2.1. Definiens Developer 8.1 (eCognition)

Ce logiciel utilise un algorithme de segmentation de type *Region Growing*, multi-résolution (c'est-à-dire qui permet d'intégrer diverses sources d'information) et multi-échelle (c'est-à-dire qui permet de mettre en œuvre plusieurs échelles de segmentations hiérarchisées sur une même image, ce qui offre la possibilité d'exploiter le niveau le plus adéquat selon l'élément à cartographier) (Baatz and Schäpe, 2000).

Le *Region Growing* démarre d'un pixel source et va l'agréger au pixel voisin le plus similaire, pour former un segment ou objet qui va grandir au fur et à mesure, jusqu'à atteindre la limite d'homogénéité définie par un seuil fixé par l'utilisateur (Baatz and Schäpe, 2000).

La figure 73 illustre le principe de segmentation multi-échelle de l'image. Les relations hiérarchiques entre les objets sont très importantes et constituent des caractéristiques à part entière pouvant intervenir comme critère de classification (Definiens 2007 a et b).

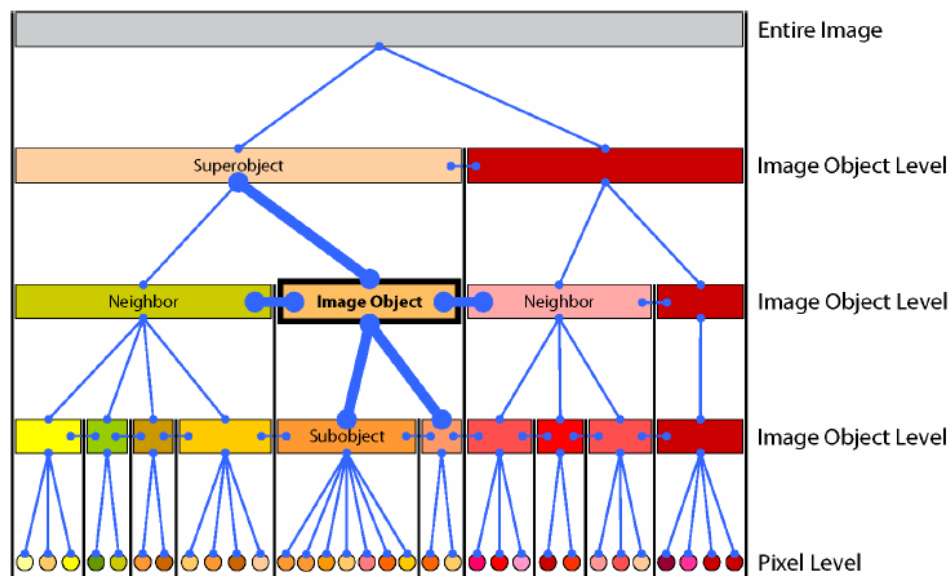


Figure 73. Schéma illustrant la structure hiérarchique entre les différents niveaux d'objets.

La segmentation au sein de Definiens fait intervenir trois paramètres (Figure 74) : *scale parameter*, *color/shape*, *smoothness/compactness*. Le premier est une valeur d'homogénéité qui sert de seuil lors de la croissance des objets. Les deux paramètres suivants permettent de jouer sur l'importance accordée à la forme des régions, en opposition à leur couleur (*color/shape*) et leur compacité en opposition à la sinuosité de leur pourtour (*smoothness/compactness*) (Baatz and Schäpe, 2000).

Le traitement d'une image dans Definiens Developer est organisé sous forme d'un *Process Tree*. Il s'agit de l'ensemble des étapes de travail structurées pas à pas, selon le principe d'un organigramme. De nombreux algorithmes peuvent être combinés en vue d'extraire les objets d'intérêt. Plusieurs méthodes de classification sont disponibles : classification par seuillage, classification floue (*fuzzy classification*), classification supervisée (*Nearest Neighbour*, sélection d'échantillons) (Definiens 2007 a et b).

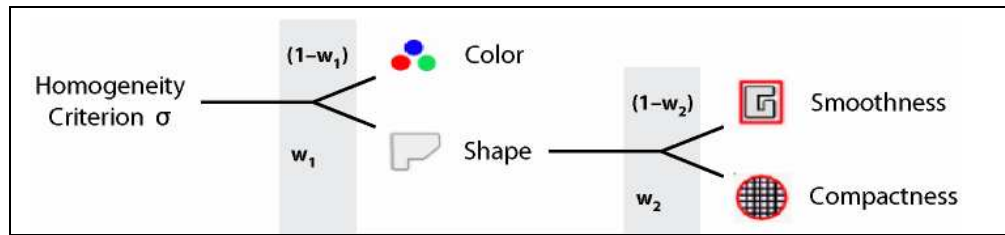


Figure 74. Paramètres de la segmentation (extrait du guide utilisateur Definiens Developer 7.0).

Definiens Developer est un logiciel performant pour la segmentation, utilisé dans de multiples études. Il permet l'intégration de plusieurs algorithmes dans une chaîne de traitement (*rule set*) qui peut être exportée et exécutée sur d'autres images. La facilité de travailler avec plusieurs niveaux de segmentation hiérarchisés est un avantage important dans l'optique d'une cartographie d'occupation du sol (légende hiérarchisée, à plusieurs niveaux).

2.5.2.2. ENVI Feature Extraction (ENVI Fx)

ENVI 4.7 est un logiciel de traitement d'images de télédétection complet qui possède de nombreuses fonctionnalités utilisées en télédétection (*pansharpening*, orthorectification, de nombreux filtres, segmentation, prétraitements, classification...) et supporte de nombreux formats d'image en plus de son format propre (ENVI 2008 a et b). Pour pouvoir faire de la segmentation, ENVI 4.7 doit être couplé au module ENVI EX, constitué de plusieurs *workflow*. « *Feature Extraction* » est le *workflow* contenant les outils de segmentation. Il permet l'extraction d'objets à partir de nombreux types d'images (panchromatique, multispectrale, radar, hyperspectrale,...).

La figure 75 présente les étapes de travail du processus de segmentation dans ENVI EX. La segmentation repose sur un algorithme de détection des contours (algorithme *watershed*, type *edge-based*). Un seul paramètre de segmentation, le *scale level*, est à fixer au départ. Cette étape est facilitée par l'usage de la fenêtre de visualisation, permettant de voir, sur une zone « échantillon », le résultat de la segmentation avec le paramètre en cours. Le *scale level* varie de 0 à 100 : la valeur 0 est la segmentation la plus fine (un pixel par segment) et la valeur 100 correspond à la segmentation la plus grossière (l'ensemble de l'image est un segment) (ENVI 2008 a et b).

Ensuite, à partir de ce résultat, on peut choisir d'affiner la segmentation, en effectuant des regroupements (*Merging segments* et *Refining segments*). Une fois le post-traitement des segments terminé, des attributs spectraux, spatiaux, texturaux peuvent être calculés et exportés. Une classification peut également être mise en œuvre avant l'exportation.

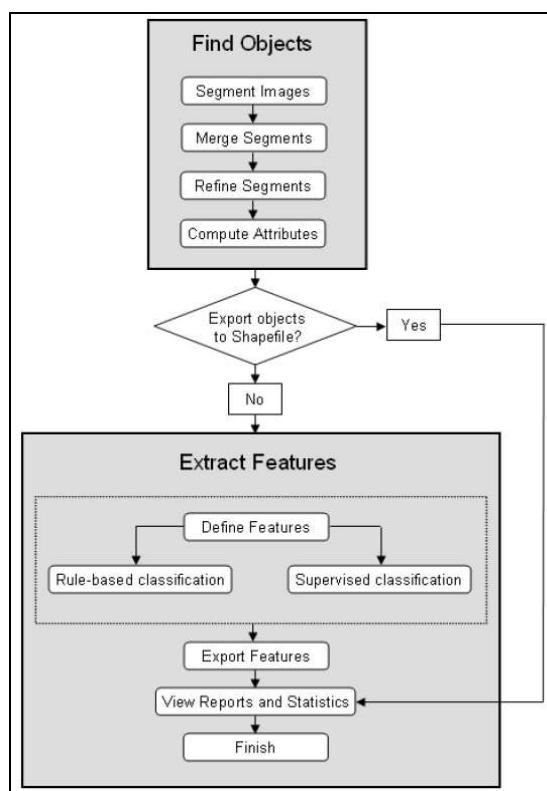


Figure 75. Feature extraction workflow (ENVI 2008 a).

2.5.3. Mise au point des procédures de segmentation

2.5.3.1. 1^{ère} étape : choix de la ou des image(s) de base

Pour rappel, les images disponibles sur un îlot sont – en fonction de sa localisation – ORTHOS_DGA, SPOT d’octobre, SPOT de juillet et MNH. Une couche de données supplémentaires a été produite et ajoutée aux précédentes : les ORTHOS_DGA dégradées à 2.5 m de résolution. A ces 5 couches de données peuvent également s’ajouter un ensemble de néo-canaux, plus communément appelés « indices de végétation », calculés à partir des bandes spectrales des différentes images (cf. paragraphe 2.6.2.1).

Dans le cadre de la segmentation proprement dite et afin de bénéficier d’une précision plus importante dans la délimitation des objets, il a été décidé de travailler préférentiellement sur les ORTHOS_DGA à très haute résolution.

2.5.3.2. 2^{ème} étape : phase préliminaire et exploratoire de segmentation

La seconde étape consiste en une phase préliminaire de tests de segmentation. Elle prend la forme d'une analyse visuelle des segments obtenus en fixant différentes valeurs des paramètres des logiciels de segmentation. Cette étape est nécessaire à plus d'un titre puisqu'elle va permettre de :

- renforcer les connaissances de l'utilisateur dans la prise en main des logiciels ;
- se familiariser avec la sensibilité des algorithmes aux valeurs des paramètres
- déterminer les gammes de valeurs des paramètres à retenir pour passer à la phase de test plus avancée.

Ces objectifs ont été rencontrés et les gammes de valeurs de paramètres à tester sont présentées au tableau 32.

Tableau 32. Combinaisons de paramètres de segmentation à tester dans Definiens et ENVI EX.

	DEFINIENS			ENVI Ex	
	<i>scale parameter</i>	<i>shape</i>	<i>compactness</i>	<i>scale level</i>	<i>merge level</i>
intervalle de valeurs testées	75 - 275	0.2 - 0.8	0.2 - 1.0	50 - 86	86 - 98
pas	50	0.2	0.2	4	2

Notons également que cette étape a permis d'observer, dans le cas d'ENVI EX, la tendance de l'algorithme de segmentation à produire des segments plus petits (surtout en milieu forestier où les contrastes sont plus forts). La difficulté étant de fixer un *merge level* adéquat afin d'éviter une *sursegmentation* qui pourrait nuire ultérieurement à l'efficacité, voire à la qualité du processus de classification.

2.5.3.3. 3^{ème} étape : tests de segmentation et évaluation de la qualité

La troisième étape de la démarche est la phase principale de test : plus systématique, elle va analyser la qualité de la segmentation en rapport avec les valeurs des paramètres. Cette phase passe par l'utilisation d'indices quantitatifs pour l'évaluation de la qualité de la segmentation et le choix de combinaisons de paramètres à maintenir pour la suite de l'étude.

Pour faciliter la lecture du rapport, cette 3ème étape est scindée en trois parties (cf. paragraphes suivants) :

- Mise en œuvre des tests de segmentation : il s'agit d'une description de la mise en œuvre des tests dans les deux logiciels utilisés (paragraphe 2.5.3.4)
- Analyse de la qualité de la segmentation : méthodologie et mise en œuvre (paragraphe 2.5.3.5)
- Choix des paramètres de segmentation : résultats des analyses pour orienter le choix des combinaisons de paramètres (paragraphe 2.5.3.6)

2.5.3.4. Mise en œuvre des tests de segmentation

a) Définiens

La succession des étapes mises en œuvre dans Définiens 8.1 pour tester un ensemble de combinaisons de paramètres de segmentation sur différents îlots est la suivante :

- création de « x¹ » « projets » (éventuellement via la création d'un *workspace*) où sont chargées les couches de base servant à la segmentation (4 canaux des ORTHOS_DGA prédécoupées à l'extension de l'îlot traité ainsi qu'une couche binaire servant à masquer les portions externes à l'îlot lors du traitement).
- composition et sauvegarde d'un *rule set* permettant pour chacune des combinaisons choisies (cf. figure 76) :
 - la création d'une couche d'objets répondant au critère d'homogénéité défini par la combinaison de paramètres testée via un algorithme de segmentation multi-résolution
 - l'export et la sauvegarde de cette couche sous forme d'un *shapefile* de polygones
 - la suppression de cette couche dans le projet en cours
- chargement de ce *rule set* dans chacun des « x » projets établis et exécution de celui-ci.

¹ Nombre d'îlots testés

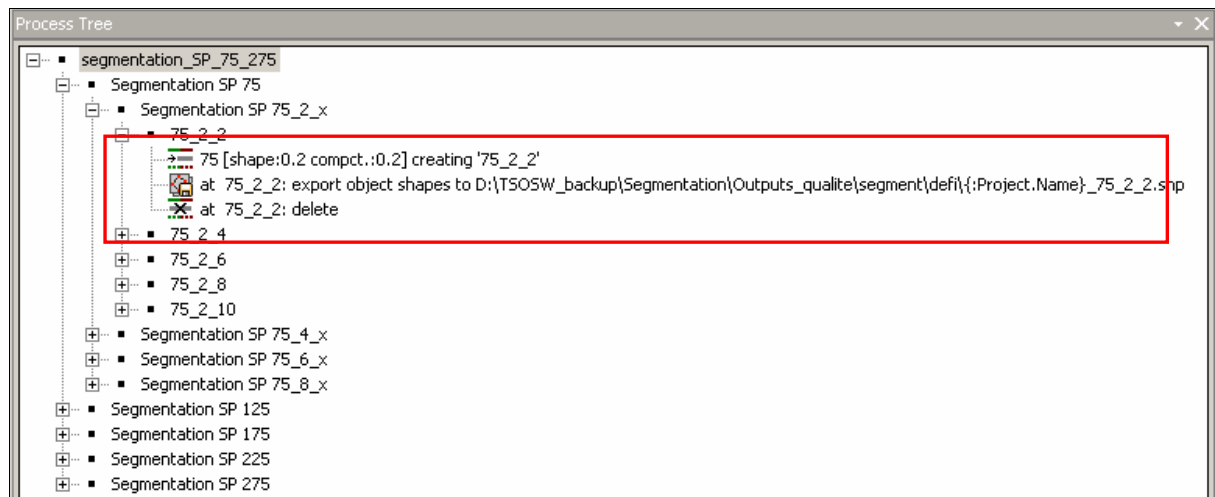


Figure 76. Rule set de segmentation dans Definiens 8.1, zoom sur les étapes programmées pour une combinaison de *scale parameter*, *shape* et *compactness* (respectivement 75, 0.2, 0.2)

Notons que la raison pour laquelle l'ensemble du traitement n'a pas pu être intégré au sein d'un seul projet est lié au fait que la licence Definiens utilisée ne comporte pas de langage d'écriture de macros permettant cette automatisation.

Cette **remarque est importante** pour la suite du projet si on devait envisager une production à grande échelle (ex. sur l'ensemble de la Région Wallonne).

b) ENVI 4.7 + ENVI EX

Le logiciel ENVI comporte un langage de programmation intégré. Il est donc possible d'automatiser complètement les opérations de segmentation (ensemble des combinaisons de *scale level* et de *merge level*) sur tous les îlots « test ». Le langage de programmation utilisé dans ce logiciel est le langage IDL (*Interactive Data Language*). Afin de comparer les résultats de segmentation des deux logiciels, la programmation a été écrite de manière à exporter les résultats de la partie « *Find Objects* » illustrée à la figure 75.

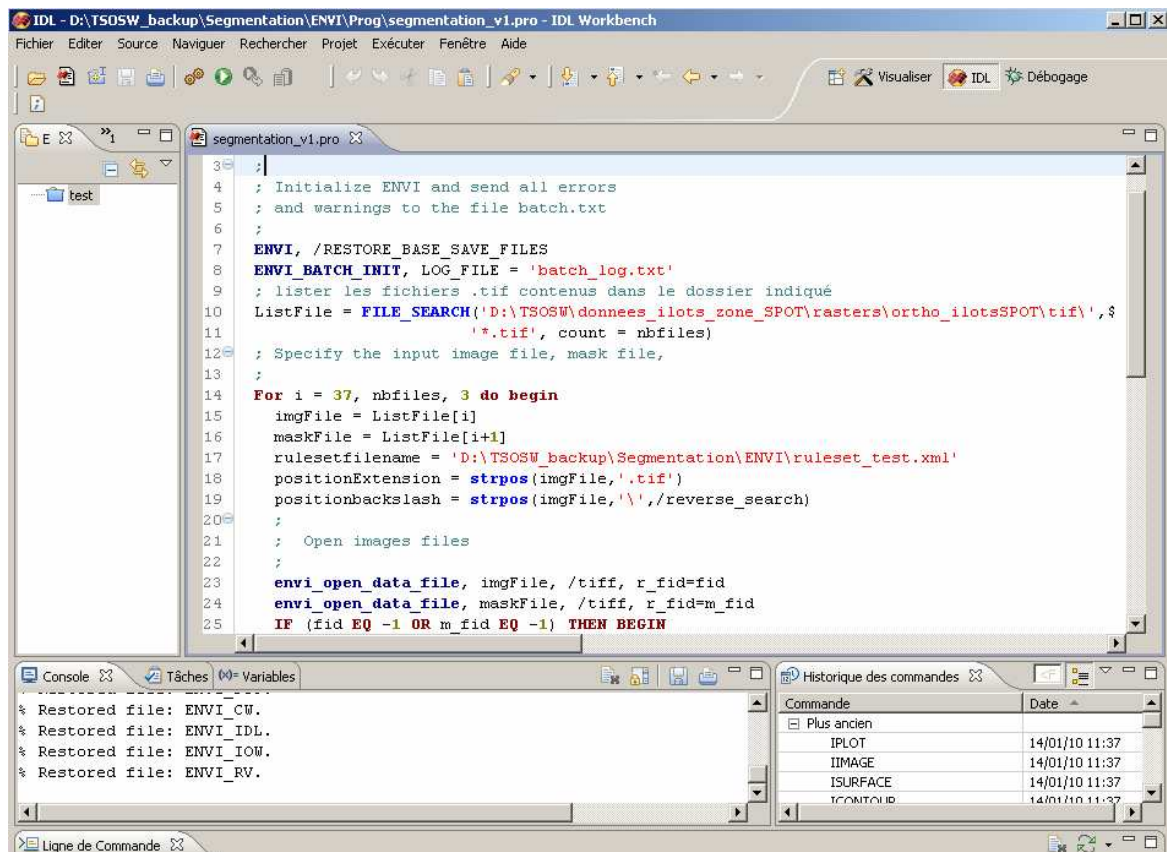


Figure 77. Environnement de programmation dans ENVI 4.7

2.5.3.5. Analyse de la qualité de la segmentation

a) Introduction : indices quantitatifs et classes visées

La qualité de la classification est directement affectée par la qualité de la segmentation. L'évaluation de la qualité de la segmentation revêt donc une grande importance (Möller, Lymburner et al., 2007). Cette évaluation de la segmentation peut se faire de manière qualitative et quantitative, c'est-à-dire basée sur une interprétation visuelle ou au contraire sur des indices quantitatifs calculés pour chaque segment produit.

Il a été décidé, dans cette étude, de guider le choix des paramètres de segmentation à l'aide d'une analyse quantitative donc plus objective et répliquable. Cette approche nous a semblée nécessaire étant donné la volonté de tester, pour les deux logiciels retenus, un grand nombre de combinaisons de paramètres.

Les indices quantitatifs analysent le résultat d'une segmentation par rapport à des objets de référence en calculant un indice de « similarité » géométrique. Différents indices ont été présentés dans les rapports intermédiaires : l'Overlapping Area (Neubert and Herold, 2008), le Shape Index (Neubert *et al.*, 2006), l'area-fit-Index (Lucieer, 2004) et l'index D (Clinton *et al.*, 2008).

Dans le cadre de cette tâche, l'index D a été préféré aux autres et plus particulièrement sa sous-composante undersegmentation puisqu'elle permet une analyse plus fine de la qualité de la segmentation (cf. infra).

Dans un premier temps (avant de lancer une première classification), l'obtention des limites des classes d'occupation du sol suivantes a été visée et analysée :

- Agricole,
- Résineux,
- Feuillus,
- Mixte,
- Gagnage (prairies situées en forêt),
- Vide (c'est-à-dire toute zone de trouée dans une zone forestière)
- Autre (comprenant essentiellement les zones bâties en zone agricole).

b) Index D : définition et mise en œuvre

Définition

L'index D mesure la qualité de la segmentation à partir d'une estimation de l'*oversegmentation* (image trop segmentée) et de l'*undersegmentation* (image pas assez segmentée, i.e. limite des objets de référence non matérialisées). Ces deux indices se calculent à partir des surfaces des zones d'intersection entre l'objet de référence et les segments (Clinton, Holt et al., 2008).

Le calcul de D se fait par les formules présentées aux équations 1 et 2. Une valeur de D égale à zéro correspond à une concordance parfaite entre les segments et les zones d'entraînements ; pour optimiser la segmentation, D doit donc être minimisé. Les X_i sont les objets du jeu d'entraînement et les Y_j sont les objets issus de la segmentation.

Pour un objet X_i , plusieurs Y_j peuvent intervenir dans le calcul : la figure 78 illustre les critères pris en compte (inclusion des centroïdes et superficies communes).

Équation 1. Formule de calcul de l'*oversegmentation* et de l'*undersegmentation* d'après (Clinton, Holt et al., 2008).

$$\text{OverSegmentation}_{ij} = 1 - \text{area}(x_i \cap y_j) / \text{area}(x_i).$$
$$\text{UnderSegmentation}_{ij} = 1 - \text{area}(x_i \cap y_j) / \text{area}(y_j).$$

Équation 2. Formule de calcul de l'index D (Clinton, Holt et al., 2008).

$$D = \sqrt{\text{OverSegmentation}^2 + \text{UnderSegmentation}^2}$$

$$\begin{aligned} Y_a &= \{\text{all } y_j \text{ where the centroid of } x_i \text{ is in } y_j\} \\ Y_b &= \{\text{all } y_j \text{ where the centroid of } y_j \text{ is in } x_i\} \\ Y_c &= \{\text{all } y_j \text{ where } \text{area}(x_i \cap y_j) / \text{area}(y_j) > 0.5\} \\ Y_d &= \{\text{all } y_j \text{ where } \text{area}(x_i \cap y_j) / \text{area}(x_i) > 0.5\} \end{aligned}$$

Figure 78. Critères de prise en compte des objets Yj pour chaque objet de référence Xi. Les objets Yj d'intérêt sont les Yj contenant le centroïde de Xi, les objets Yj dont les centroïdes sont contenus dans Xi ainsi que les objets Yj dont au moins 50% de la surface coïncident avec Xi (Clinton, Holt et al., 2008).

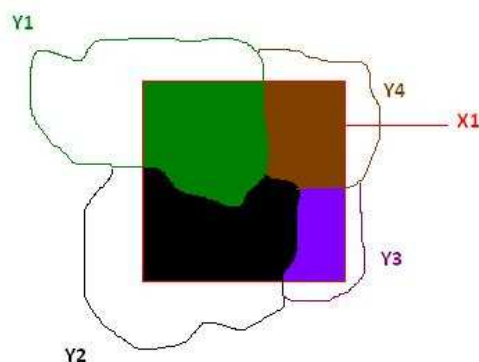


Figure 79. Illustration du calcul de l'oversegmentation et de l'undersegmentation, par comparaison des surfaces de l'objet de référence et des objets issus de la segmentation.

La figure 79 illustre le principe de calcul. Pour l'objet de référence X1, quatre valeurs *d'oversegmentation* et quatre valeurs *d'undersegmentation* sont calculées : X1 par rapport à Y1, à Y2, à Y3 et à Y4, les surfaces prises utilisées dans la formule sont les surfaces d'intersections entre X1 et les quatre Y (en couleurs pleines). Ensuite, les moyennes *d'oversegmentation* et *undersegmentation* et D sont calculées. Le calcul de D peut se faire pour tous les objets ou par type d'objets, c'est-à-dire par classes (arbres, bâtiments, ...), ce qui permet une étude plus fine de la segmentation.

Nous avons choisi de considérer prioritairement le critère *d'undersegmentation* apparaissant dans l'index D. En effet, une « mauvaise » *undersegmentation* (i.e. une valeur élevée de ce paramètre) signifie que certains segments produits contiennent plusieurs occupations du sol (ils sont à cheval sur des limites des objets de la couche de référence) (cf. figure 80). Par contre, une valeur élevée *d'oversegmentation* traduit le fait que la segmentation a produit un grand nombre de segment au sein d'un objet de référence. Cette dernière situation ne constitue pas un risque direct d'altération du résultat final (en termes de qualité de délimitation des objets). Il peut cependant nuire à l'efficacité globale du processus et dans une certaine mesure à la qualité de la classification (cf. figure 81). Nous fixerons donc des valeurs limites *d'oversegmentation* moyenne par combinaison afin de ne pas segmenter exagérément les images.

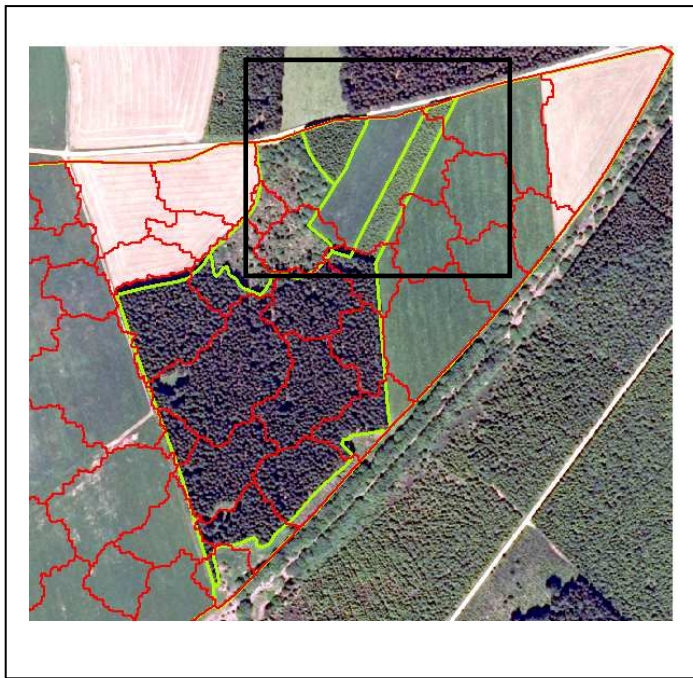


Figure 80. En vert : vérité terrain, en rouge résultat de la segmentation (Definiens, 150, 0.9, 0.6). Encadré noir : polygones de référence agricole et résineux avec *undersegmentation* élevée (0,65 à 0,85)

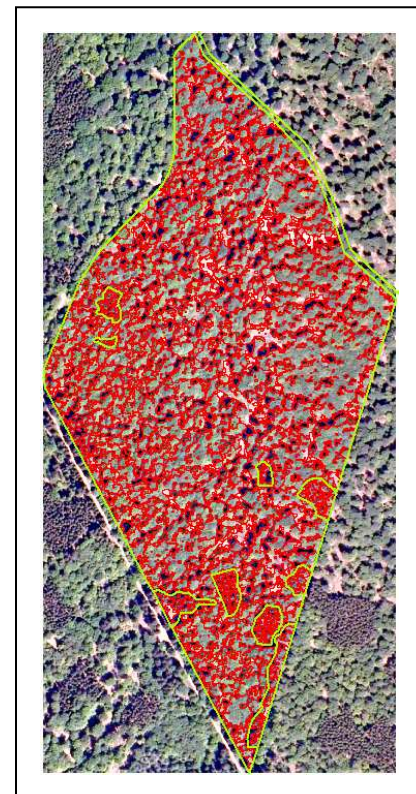


Figure 81. Segmentation (ENVI 34, 86) créant une *oversegmentation* très élevée pour tous les polygones de référence (~0.97)

Mise en œuvre

Afin de pouvoir analyser un grand nombre de combinaisons de paramètres de segmentation créées à la fois par Definiens et par ENVI, une application d'automatisation du calcul de l'index D (et donc, entre autre, de *l'undersegmentation*) a été développée. Celle-ci l'a été dans l'environnement du tableur de Microsoft, Excel, via un code écrit en langage Visual Basic for Applications (VBA). Ce code fait appel à des outils de geoprocessing d'ArcGis (*Intersects,...*), des sélections (par attributs et spatiales, points dans polygones,...), de rapport de surfaces,...

Cette application travaille pour chacune des combinaisons de paramètres testée sur la comparaison de deux *shapefiles* (cf. définition de l'Index D, supra): le premier, de référence, est unique (*shapefile* représentant la « vérité terrain » pour l'ensemble des 43 îlots « test », constitués en phase 1 de digitalisation, cf. paragraphe 2.4.3.1), le deuxième étant le *shapefile* résultat de la segmentation (exporté de Definiens ou d'ENVI).

Les résultats numériques *d'undersegmentation* et *d'oversegmentation* calculés par cette application sont stockés dans une base de données Access. Dans la table de résultats, pour chaque objet de référence (i.e. un polygone d'une occupation du sol homogène (cf. classes retenues, supra)), on retrouve les informations suivantes (cf. figure 82) :

- l'identifiant de l'îlot dans lequel l'objet se situe,
- le logiciel utilisé,
- les paramètres de segmentation utilisés : pour Definiens, les paramètres 1, 2 et 3 étant respectivement, le *scale parameter*, le *shape* (multiplié par 10, pour facilité de dénomination de fichier) et le *compactness* (également multiplié par 10), la classe d'occupation du sol (vérité terrain, cf. paragraphe 2.4.3.1) et pour ENVI, seuls les deux premiers paramètres sont remplis, représentant les *scale* et *merge levels*,
- et les valeurs obtenues *d'under* et *d'oversegmentation*

id_test_seg	id_ilot	id_ref	id_soft	param1	param2	param3	occupation	under_s	over_s
4851	51932	1	defi	125	6	2	AGRICOLE	0.0012055	0.9149298
5011	51932	1	defi	75	6	2	AGRICOLE	0.0012055	0.9643034
7235	57449	1	defi	225	6	6	AGRICOLE	0.00121135	0.02551896
37168	54155	3	defi	125	4	3	AGRICOLE	0.00121305	0.7875597
37672	54155	3	defi	150	4	3	AGRICOLE	0.00121384	0.4112329
33300	51932	1	defi	125	8	7	AGRICOLE	0.00121588	0.8652882
33462	51932	1	defi	150	8	7	AGRICOLE	0.00121588	0.7710122
4728	51261	5	defi	275	8	8	RESINEUX	0.00121703	0.4577735
36328	53453	2	defi	125	7	6	AGRICOLE	0.00121972	0.9694986
4891	51932	1	defi	175	6	2	AGRICOLE	0.00121977	0.8455951
26765	44057	2	defi	125	7	7	RESINEUX	0.00122435	0.8798382
26703	44057	2	defi	125	3	6	RESINEUX	0.00122555	0.9061151
7191	57449	1	defi	175	6	2	AGRICOLE	0.00123034	0.2711137
25008	43630	6	envi	50	94	0	VIDE	0.9810484	0.04944535
25009	43630	1	envi	50	98	0	AGRICOLE	0.03727363	0.03860512
25023	43630	3	envi	54	82	0	FEUILLUS	0.0954228	0.9322593
25011	43630	3	envi	50	98	0	FEUILLUS	0.9803463	0.00448466
25022	43630	2	envi	54	82	0	AUTRE	0.9244155	0.4803037
25013	43630	5	envi	50	98	0	VIDE	0.05245959	0.1878802
25014	43630	6	envi	50	98	0	VIDE	0.9836681	0.00998882
25065	43630	3	envi	58	86	0	FEUILLUS	0.1221173	0.9136192
25016	43630	2	envi	54	78	0	AUTRE	0.9243671	0.4849446
25175	43630	5	envi	70	86	0	VIDE	0.3964833	0.01488839
25018	43630	4	envi	54	78	0	RESINEUX	0.01339748	0.9836835
25019	43630	5	envi	54	78	0	VIDE	0.2260461	0.3277982
25020	43630	6	envi	54	78	0	VIDE	0.364982	0.2118035

Figure 82. Exemple de table de résultat de calcul d'under et oversegmentation

2.5.3.6. Choix de combinaisons de paramètres de segmentation

Pour rappel, le tableau 32 (supra), présente les combinaisons de paramètres de segmentation testées dans les deux logiciels. Ceci représente pour les 43 ilots tests sélectionnés dans la zone commune aux deux scènes SPOT (1^{ère} phase de collecte), 4300 *shapefiles* produits dans Definiens et 2580 produits dans ENVI EX.

a) Definiens

Analyse graphique des résultats

La figure 83 met clairement en évidence la relation décroissante et attendue entre les *oversegmentation* et *undersegmentation* moyennes par combinaison de paramètres.

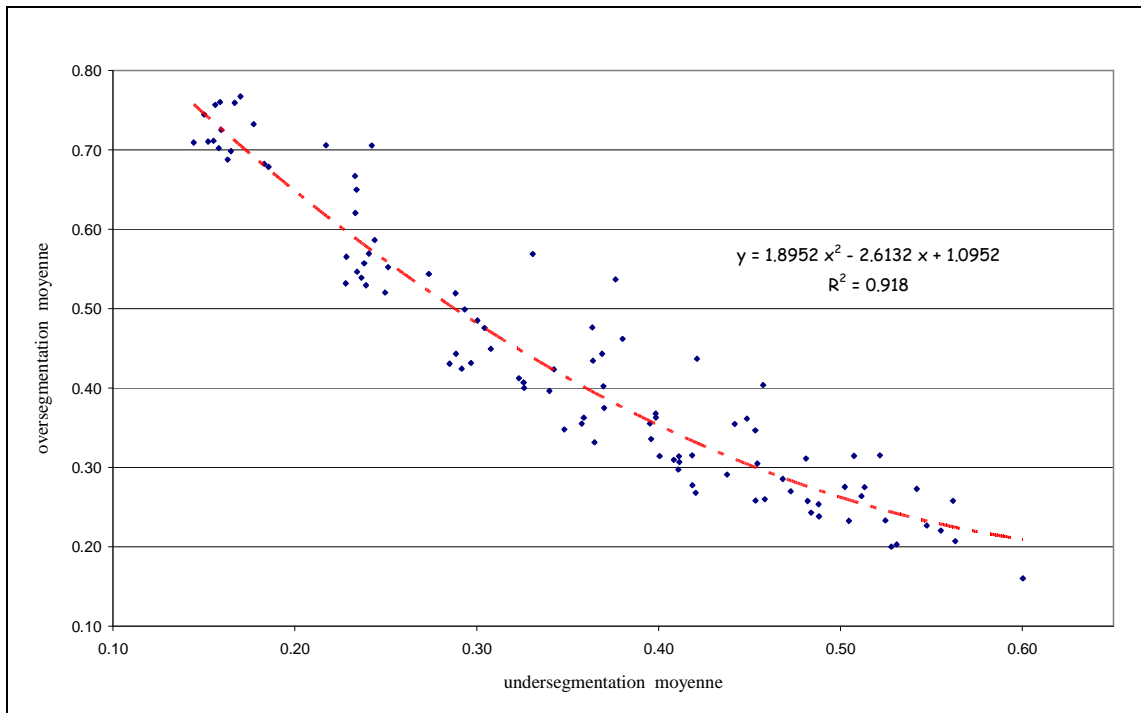


Figure 83. Relation entre les *oversegmentation* et *undersegmentation* moyennes par combinaison

La figure 84 illustre l'évolution de l'*undersegmentation* en fonction de différentes combinaisons de paramètres. Notons que celles-ci sont classées par *scale parameter*, pour un même *scale parameter*, par *shape* croissant et enfin, pour un même *shape*, par *compactness* croissant.

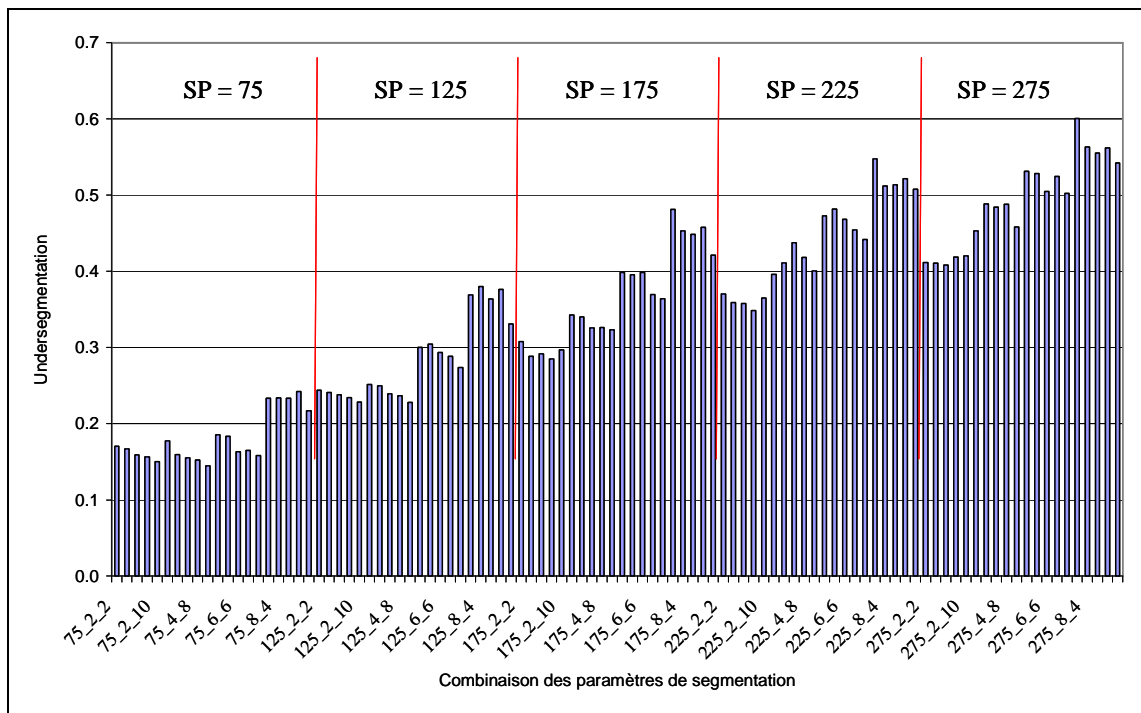


Figure 84. Evolution de l'*undersegmentation* en fonction des combinaisons de paramètres dans Definiens.

Sans être d'un grand détail, cette figure nous permet d'ores et déjà d'observer quelques tendances. Tout d'abord, plus le *scale parameter* est petit, meilleure est l'*undersegmentation* (en moyenne). Ceci est logique puisque les segments étant plus petits, la probabilité de marquer une limite donnée est plus forte (d'autant plus si la distinction spectrale entre classes n'est pas clairement marquée). On peut également observer que dans chaque ensemble de combinaisons du même *shape* et d'un même *scale parameter*, les qualités de segmentation se valent. Ceci sera démontré statistiquement (cf. infra).

Pour éviter de produire trop de petits segments (cf. justification supra), nous avons choisi de ne considérer que les combinaisons dont l'*oversegmentation* était inférieure à 0,7. Cela a pour effet d'exclure une grande partie des combinaisons avec un *scale parameter* de 75.

Pour affiner l'analyse de l'influence des paramètres de segmentation sur la qualité de celle-ci, les combinaisons à *scale parameter* de 125 (faible *undersegmentation* et *oversegmentation* inférieure à 0.7) ont été sélectionnées. L'ensemble des combinaisons indiquées au tableau 33 a été généré et l'*undersegmentation* calculée.

Tableau 33. Combinaisons de paramètres de segmentation testés dans Definiens, deuxième phase

	DEFINIENS		
	<i>scale parameter</i>	<i>shape</i>	<i>compactness</i>
intervalle de valeurs testées	125	0.1 - 0.9	0.1 - 1.0
pas		0.1	0.1

Les figures suivantes illustrent, pour un *scale parameter* de 125, l'influence des deux derniers paramètres (*shape* et *compactness*) sur l'*undersegmentation* et ce, en relation avec les différentes occupations du sol.

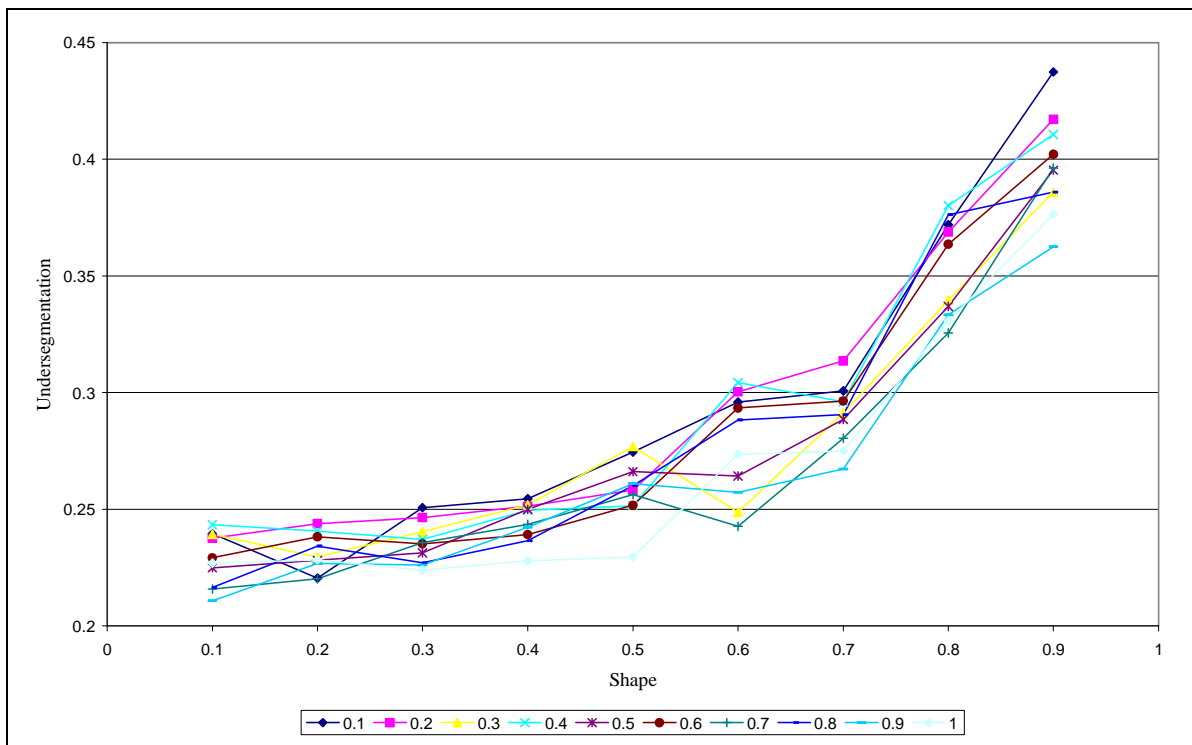


Figure 85. *Undersegmentation* en fonction du paramètre *shape* pour différentes valeurs du paramètre *compactness* et pour une valeur de *scale parameter* égale à 125.

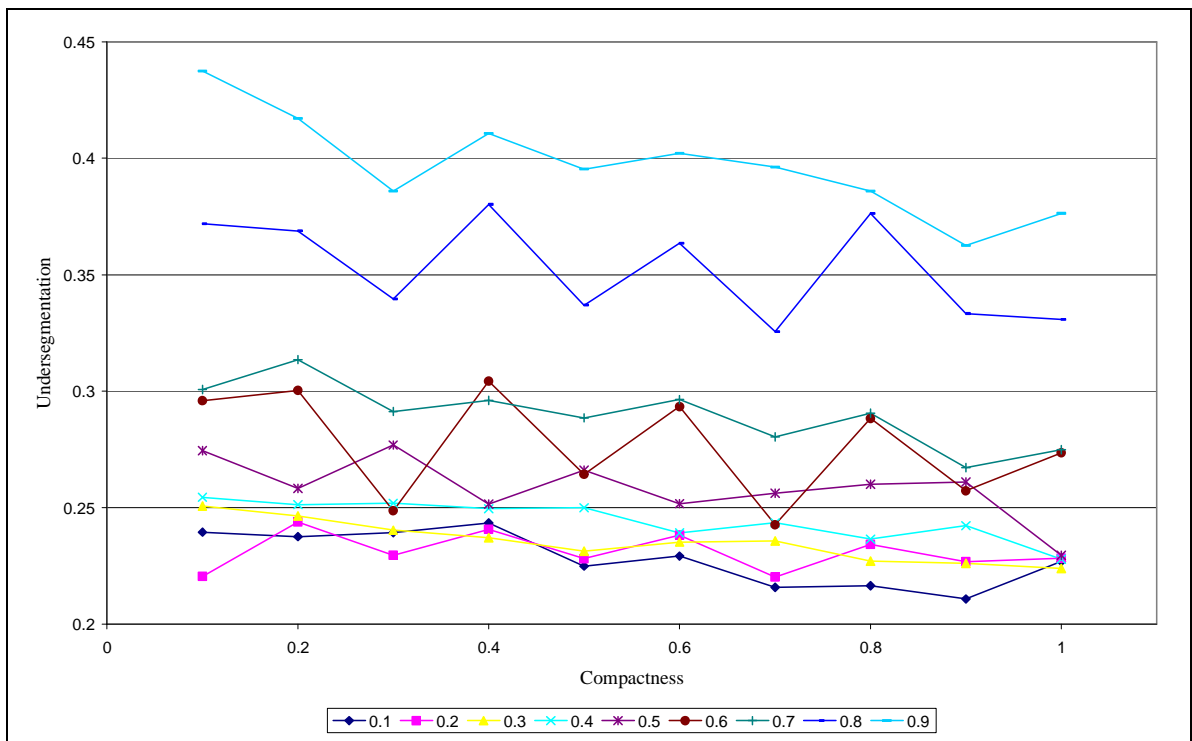


Figure 86. *Undersegmentation* en fonction du paramètre *compactness* pour différentes valeurs du paramètre *shape* et pour une valeur du *scale parameter* égale à 125.

En ce qui concerne le paramètre *shape*, on observe aux figures 85 et 86 que *l'undersegmentation* reste stable et autour de la valeur 0,25 pour des facteurs de *shape* de 0,1 à 0,4 pour ensuite augmenter jusqu'à des valeurs proches de 0,4 pour des facteurs de *shape* de 0,9 à 0,1. Par contre, on n'observe aucune tendance nette d'influence du paramètre *compactness* sur les valeurs d'*undersegmentation*.

La figure 87, à l'instar de la figure 85, illustre l'évolution de *l'undersegmentation* en fonction du paramètre *shape* mais cette fois pour les différents types d'occupation du sol visés à ce stade et pour un paramètre *compactness* égal à 1.0.

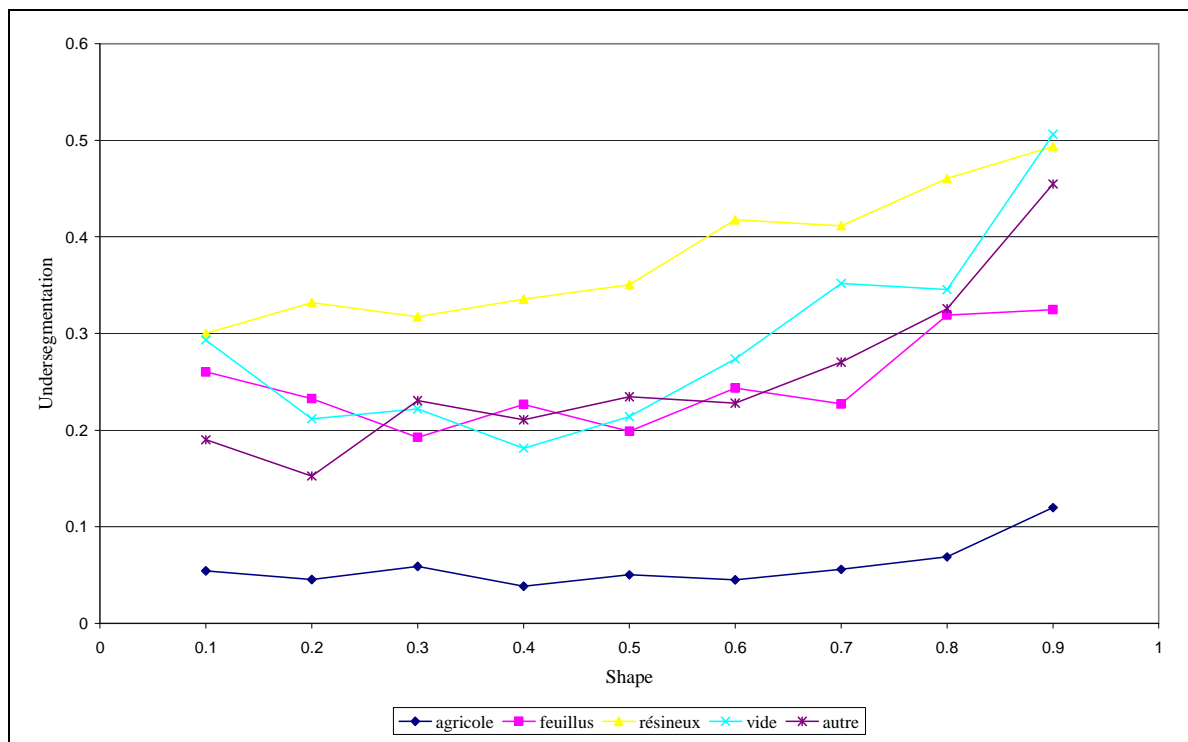


Figure 87. Undersegmentation pour différents types d'occupation du sol en fonction du paramètre *shape*, pour un *scale parameter* égal à 125 et un paramètre de *compactness* égal à 0,2

On peut observer que la classe « agricole » est caractérisée par une *undersegmentation* en moyenne très faible (donc de très bonne qualité) tandis que les limites des zones de type « résineux » sont plus difficiles à identifier. Ceci peut être expliqué par le fait que les limites de zones agricoles sont souvent tranchées (agricole-forêt,...) tandis que plusieurs zones de résineux définies lors de la tâche 4 sont des petites inclusions dans des ensembles forestiers.

Analyse statistique des résultats

▪ INTRODUCTION

Des analyses statistiques ont été menées sur le jeu de données avec deux objectifs :

- attester de la nécessité ou non de traiter les types d'îlots différemment en ce qui concerne l'étape de segmentation
- orienter le choix des combinaisons de segmentation à retenir pour la suite du traitement

Notons qu'afin de ne pas donner trop d'importance au critère *color* (complémentaire à 1 du paramètre *shape*) et inversement, il a été décidé d'exclure du jeu de données les combinaisons ayant 0.1, 0.2 et 0.9 pour paramètre *shape*. Ceci se justifie notamment par les variations spectrales marquées le long des *seamlines* n'expliquant aucunement un changement d'occupation du sol.

▪ OPPORTUNITE DE REALISER UNE ANALYSE DIFFERENTIEE PAR TYPE D'ILOT

La question du traitement différentiel ou non en fonction du type d'îlot peut être résolue en testant l'interaction des facteurs « combinaison » (étant l'association des paramètres de segmentation utilisés dans Definiens, 60 différents dans ce cas) et « type » d'îlot (8 types). Ceci a été fait via une analyse de la variance à deux facteurs croisés fixes.

Les résultats fournis par le logiciel Minitab sont illustrés à la figure 88.

Analysis of Variance for under, using Adjusted SS for Tests						
Source	DF	Seq SS	Adj SS	Adj MS	F	P
Type	7	225.8293	225.8293	32.2613	427.83	0.000
combi	59	26.5661	11.9569	0.2027	2.69	0.000
Type*combi	413	13.9370	13.9370	0.0337	0.45	1.000
Error	11700	882.2600	882.2600	0.0754		
Total	12179	1148.5924				

S = 0.274603 R-Sq = 23.19% R-Sq(adj) = 20.04%

Figure 88. Résultats de l'ANOVA à deux facteurs fixes pour un *scale parameter* de 125 (facteurs : combinaison de paramètres de segmentation et type d'îlot), non significativité de l'interaction des deux facteurs

L'interaction entre les facteurs testés dans l'explication de la variabilité de l'*undersegmentation* est non significative [hypothèse rejetée avec un p de 1.00 (cf. Figure 88)]. Ceci nous permet donc de rechercher, tous types confondus, la ou les meilleures combinaisons à sélectionner pour la suite des traitements.

▪ CHOIX DE COMBINAISONS

La méthode Gupta a été appliquée au jeu de donnée pour orienter le choix des combinaisons de paramètres de segmentation (Dagnelie, 2006). Cette méthode permet, dans le cadre de cette analyse, la détermination de l'intervalle de valeurs contenant les *undersegmentations* statistiquement similaires à l'*undersegmentation* minimale.

La borne supérieure de cet intervalle se calcule comme suit :

$$y_{\min} + g_{1-\alpha} \sqrt{\frac{CM}{n}}$$

avec y_{\min} : la valeur minimale d'*undersegmentation* moyenne par combinaison

CM : le carré moyen résiduel de l'ANOVA à 2 facteurs fixes (cf. supra)

n : le nombre d'observations ayant servi à calculer chaque moyenne (ici, 203)

$g_{1-\alpha}$: la valeur de la variable de Gupta avec $p=n$ et $k =$ nombre de degrés de liberté du CM

Dans le cas traité, la meilleure combinaison (i.e. ayant l'*undersegmentation* la plus faible) est la combinaison : *scale parameter* : 125, *shape* : 0.3 et *compactness* : 1.0 ayant une valeur d'*undersegmentation* de 0.22. La borne supérieure calculée via la méthode Gupta vaut 0.3.

Ce qui revient à dire que les combinaisons indiquées au tableau 34 sont équivalentes statistiquement (parmi celles testées). Remarquons que ces combinaisons sont au nombre de 40 alors que le nombre de combinaisons testées est de 60. Ceci relativise les différences moyennes d'*undersegmentation* pour un même *scale parameter*.

Parmi ces combinaisons, celle ayant un ***scale parameter* de 125, un *shape* de 0.3 et un *compactness* de 1.0** a été sélectionnée pour initier la phase d'établissement des règles de classification, tout en gardant à l'esprit que 39 autres combinaisons donnent des résultats de qualité de segmentation équivalents.

Tableau 34. Combinaisons ayant pour scale parameter 125 équivalentes statistiquement

shape	compactness	undersegmentation	shape	compactness	undersegmentation
0,3	1	0,222	0,4	0,1	0,252
0,3	0,8	0,225	0,4	0,5	0,252
0,4	1	0,226	0,5	0,4	0,253
0,3	0,9	0,227	0,5	0,7	0,256
0,3	0,6	0,232	0,5	0,2	0,261
0,3	0,5	0,234	0,5	0,9	0,261
0,4	0,8	0,234	0,5	0,8	0,261
0,3	0,4	0,236	0,5	0,5	0,266
0,3	0,7	0,237	0,6	0,7	0,269
0,4	0,6	0,238	0,6	1	0,271
0,3	0,3	0,239	0,5	0,1	0,273
0,5	1	0,241	0,6	0,5	0,275
0,4	0,9	0,242	0,6	0,3	0,277
0,3	0,2	0,246	0,5	0,5	0,278
0,4	0,4	0,248	0,6	0,9	0,286
0,4	0,2	0,249	0,6	0,8	0,287
0,4	0,7	0,249	0,6	0,6	0,292
0,3	0,1	0,249	0,7	0,9	0,298
0,4	0,3	0,251	0,6	0,2	0,298
0,5	0,6	0,251	0,7	0,7	0,301

b) ENVI EX

La figure 89 illustre l'évolution de l'*undersegmentation* en fonction de différentes combinaisons testées. Les plages de paramètres testées ont été choisies via une exploration visuelle de résultats exploratoires fournis avec une grande plage de paramètres. Notons que les combinaisons sont classées, à la figure 89, par *scale level* et pour un même *scale level*, par *merge level* croissant. Cette figure illustre l'augmentation générale, et jusqu'à des valeurs élevées, de l'*undersegmentation* avec le *scale level*. Notons que pour des valeurs de *scale level* inférieures à 50, l'*oversegmentation* devient supérieure à 0.7 (limite fixée, cf. supra).

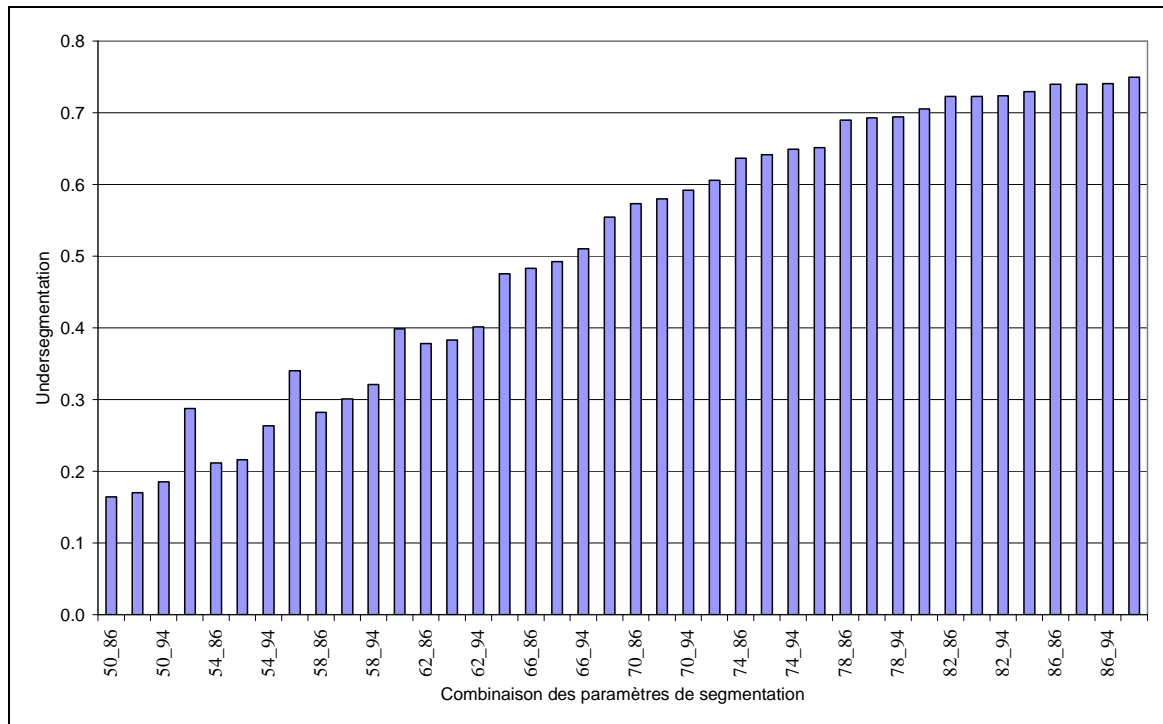


Figure 89. Evolution de l'undersegmentation en fonction des combinaisons de paramètres dans ENVI

L'étude des combinaisons à maintenir a été moins « poussée » quantitativement que pour Definiens et ce, pour plusieurs raisons :

- qualitativement, ENVI semble donner de moins bons résultats : notamment, il produit des segments qui, *a priori*, semblent moins faciles à classer
- quantitativement, les combinaisons caractérisées par une *undersegmentation* inférieure à 0.4 et une *oversegmentation* inférieure à 0.7 sont moins nombreuses que pour Definiens (cf. figure 89).

Malgré le ressenti de moindre qualité par l'opérateur, nous avons néanmoins choisi de maintenir une combinaison de paramètres de segmentation pour ENVI EX afin de ne pas exclure un logiciel à un stade non abouti de l'étude et de pouvoir ainsi comparer les résultats de classification des outputs de segmentation des deux logiciels.

Le choix s'est orienté sur une combinaison de segmentation associée à une valeur d'*undersegmentation* similaire à celle retenue pour Definiens (c'est-à-dire autour de 0.2) :

→ scale level = 50 et merge level = 90.

c) Illustration

Après avoir décrit l'orientation des choix des paramètres de segmentation, voici une illustration de résultats obtenus avec ceux-ci. Pour chaque îlot présenté, la « vérité terrain » (basée sur l'interprétation de différentes couches de données) est présentée, suivie des résultats de segmentation de Definiens (combinaison retenue : *scale parameter* = 125, *shape* = 0.3 et *compactness* = 1.0) et de ceux d'ENVI (combinaison retenue : *scale level* = 50 et *merge level* = 90).



Figure 90. Ilot n°53453, vérité terrain et résultats des segmentation sur Definiens et Envi Ex.

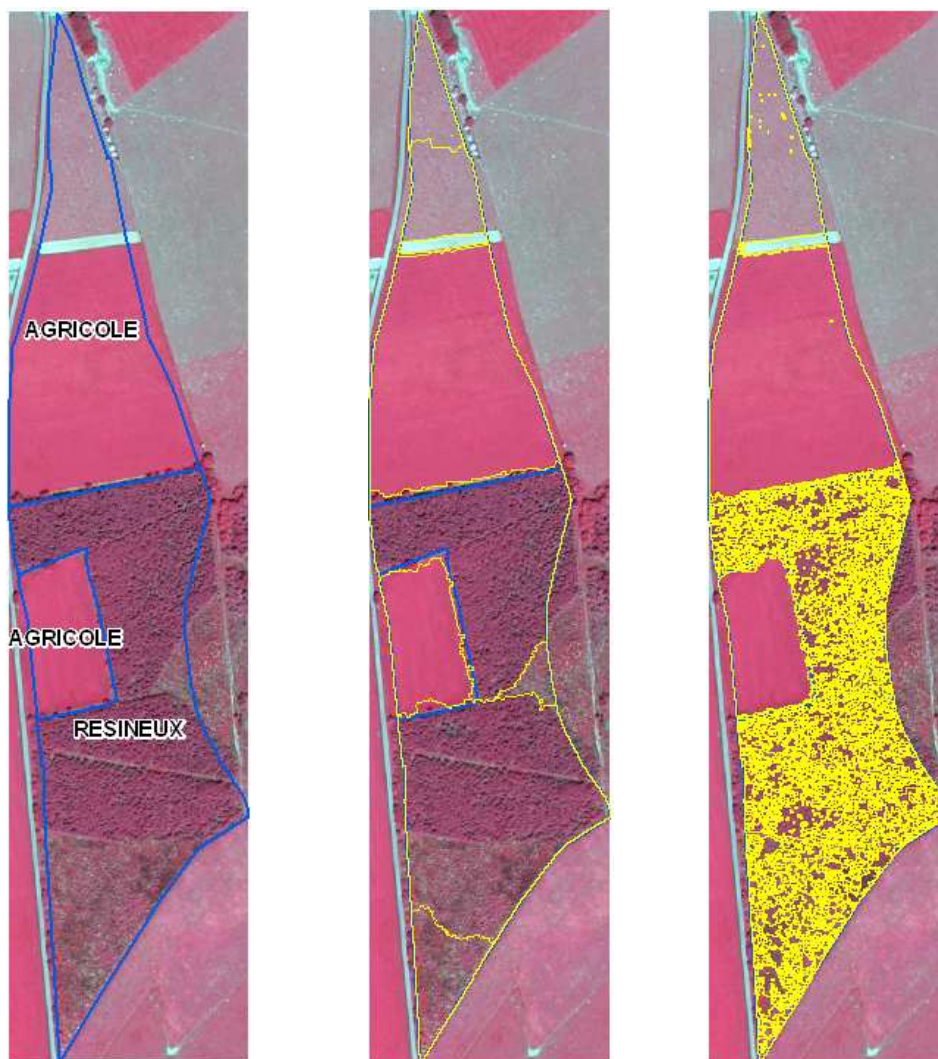


Figure 91. Ilot n°53317, vérité terrain (gauche) et résultats des segmentations sur Definiens (centre) et Envi Ex (droite).



Figure 92. Ilot n°43669, vérité terrain (gauche) et résultats des segmentations sur Definiens (centre) et Envi Ex (droite)

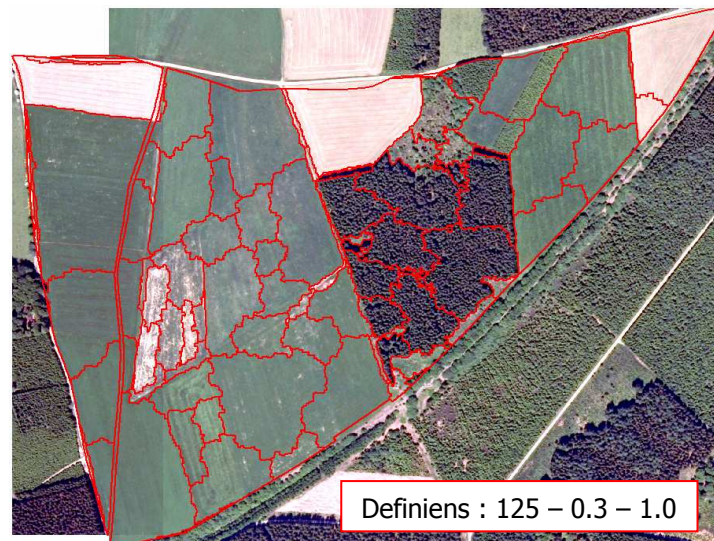
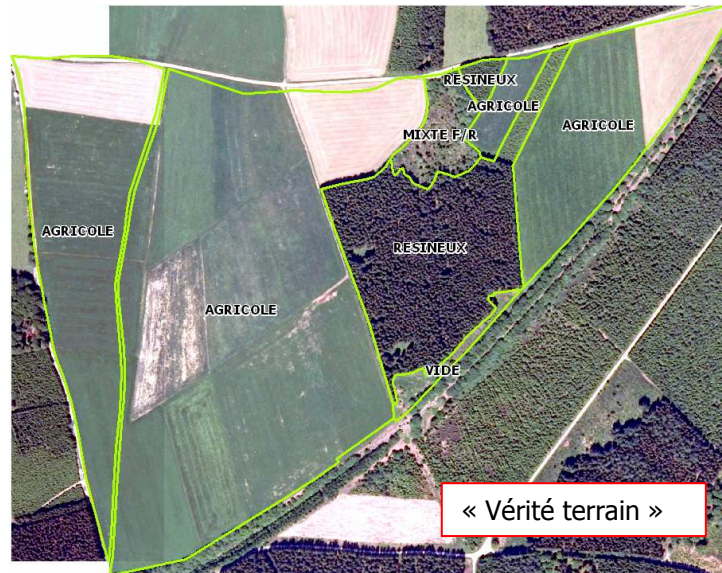


Figure 93 (1^{ère} partie). Ilots n°54144 et 54155, vérité terrain et résultats des segmentations sur Definiens

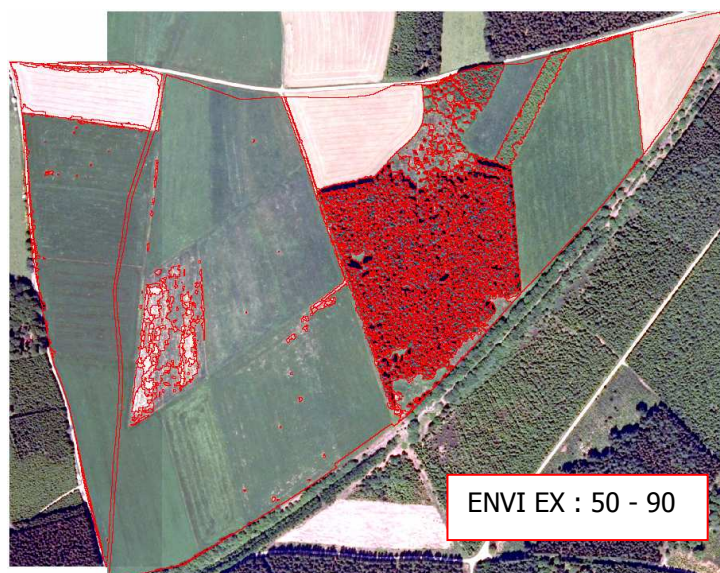


Figure 93 (suite). Ilots n°54144 et 54155, résultats des segmentations sur Envi Ex

Outre les différences qualitatives notables entre les segmentations fournies par les deux logiciels, nous pouvons remarquer à la figure 90, l'effet de la différence radiométrique liée au mosaïquage des ortho-images le long vraisemblablement d'une *seamline*.

2.6. Préparation de la base de données « images » (GRFMN)

2.6.1. Introduction

L'objectif de cette tâche est de définir les propriétés des objets qui présenteront un intérêt pour la classification. Par propriétés, on entend des attributs visant à la caractérisation des objets générés dans la phase de segmentation. Ces attributs portent sur la radiométrie, la texture ou la forme des objets.

Pour calculer ces attributs, différentes sources de données « images » sont disponibles et ont déjà été évoquées dans les paragraphes précédents : ORTHOS_DGA, MNH, SPOT 5 d'octobre, SPOT 5 de juillet. A ces quatre types d'images, la création d'une couche ORTHO dégradée à 2,5m de résolution a été ajoutée afin d'étudier l'apport dans la classification d'une couche où la grande variabilité spectrale induite par la taille réduite des pixels a été diminuée. Cette dégradation d'information peut également se justifier dans l'optique d'une utilisation par pixel.

Le paragraphe 2.6.2 présente les différents types d'attributs utilisés pour caractériser les objets et le paragraphe 2.6.3 rappelle brièvement l'application informatique développée pour venir en aide à la préparation de la base de données « images ».

2.6.2. Type d'attributs

2.6.2.1. Attributs radiométriques

Les attributs radiométriques sont calculés à partir des bandes spectrales des différentes images :

- R, V, B, IR des ORTHOS_DGA et ORTHOS dégradées
- IR, R, V et MIR des SPOT

Ceux-ci peuvent également être obtenus sur base de néocanaux (ou indices de végétation).

Classiquement, la moyenne des valeurs des pixels à l'intérieur d'un objet est calculée. D'autres attributs sont également utilisés et présentés au tableau 35. En dehors des quatre canaux classiques des images, d'autres couches « raster » peuvent également être prises en compte pour le calcul d'attributs radiométriques. Il peut s'agir par exemple du MNH (calcul d'une hauteur moyenne d'un îlot) mais également de néo-canaux (cf. tableau 36).

Tableau 35. Formule de calcul d'attributs radiométriques couramment utilisés

Nom de l'indice	Formule	Signification
Mean _c	$\mu_c = \frac{\sum_{i=1}^n v_i}{n}$	Valeur moyenne d'un canal unique (c), avec v_i , la valeur du pixel i de l'objet n, le nombre de pixels dans l'objet
Spectral Mean or Brightness	$b = \frac{\sum_{i=1}^{n_c} \mu_i}{n_c}$	Brillance de l'objet, avec μ_i , la valeur moyenne du canal i, n_c , le nombre de canaux
Spectral ratio	$ratio_c = \frac{\mu_c}{\sum_{i=1}^{n_c} \mu_i}$	Moyenne d'un canal unique normé
Standard Deviation	$\sigma_c = \frac{\sum_{i=1}^{n_c} (v_i - \mu_c)^2}{n - 1}$	Ecart-type d'un canal unique

Tableau 36. Définition de néocanaux utilisés dans le cadre de l'étude

Nom de l'indice	Formule	Signification
Simple Ratio	$\frac{\mu_R}{\mu_{IR}}$	Avec μ_R , la valeur moyenne du canal rouge μ_{IR} , la valeur moyenne du canal infrarouge
Normalized Difference Vegetation Index	$\frac{\mu_{IR} - \mu_R}{\mu_{IR} + \mu_R}$	Indice de végétation
Difference Vegetation Index	$\mu_{IR} - \mu_R$	Indice de végétation
Green Difference Vegetation Index	$\mu_{IR} - \mu_G$	Indice de végétation
Ratio Vegetation Index	$\frac{\mu_{IR}}{\mu_R}$	Indice de végétation
Green Ratio Vegetation Index	$\frac{\mu_{IR}}{\mu_G}$	Indice de végétation
Green Normalized Difference Vegetation Index	$\frac{\mu_{IR} - \mu_G}{\mu_{IR} + \mu_G}$	Indice de végétation
Soil adjusted Vegetation Index	$\frac{\mu_{IR} - \mu_R}{\mu_{IR} + \mu_R + 0.5} \times 1.5$	Indice de végétation
Green Soil adjusted Vegetation Index	$\frac{\mu_{IR} - \mu_G}{\mu_{IR} + \mu_G + 0.5} \times 1.5$	Indice de végétation
Optimized Soil adjusted Vegetation Index	$\frac{\mu_{IR} - \mu_R}{\mu_{IR} + \mu_R + 0.16}$	Indice de végétation
Green Optimized Soil adjusted Vegetation Index	$\frac{\mu_{IR} - \mu_G}{\mu_{IR} + \mu_G + 0.16}$	Indice de végétation
Forest Differentiation Index	$\mu_{IR} - (\mu_B + \mu_R)$	Indice de végétation forestière, avec μ_B , la valeur moyenne du canal bleu

2.6.2.2. Attributs «texturaux »

La texture peut être définie comme la fréquence de changement de tons dans une image. Les indices de texture calculés se basent sur la méthode développée par Haralick (1973) (Tableau 37). Il existe en réalité plusieurs paramètres. Chacun d'eux est calculé sur base de la GLCM, Gray-Level Co-occurrence Matrix (matrice de co-occurrence des niveaux de gris). Parmi ces indices de texture, on peut citer contraste, homogénéité, entropie.

Tableau 37. Formule de calcul d'attributs de texture couramment utilisés

Nom de l'indice	Formule	Signification
GLCM Mean	$\mu_{i,j} = \frac{\sum_{i,j=0}^{N-1} P_{i,j}}{N^2}$	Moyenne de la distribution des niveaux de gris, avec <i>i</i> , le numéro de la ligne <i>j</i> , le numéro de la colonne <i>P_{i,j}</i> , la valeur normalisée dans la cellule <i>i,j</i> <i>N</i> , le nombre de lignes ou de colonnes
GLCM Correlation	$\sum_{i,j=0}^{N-1} P_{i,j} \left[\frac{(i - \mu_i)(j - \mu_j)}{\sqrt{(\mu_i^2)(\mu_j^2)}} \right]$	Mesure de l'uniformité de la distribution des valeurs dans la matrice, avec <i>μ_{i,j}</i> , GLCM mean
GLCM StdDev	$\sqrt{\sum_{i,j=0}^{N-1} P_{i,j} (i, j - \mu_{i,j})^2}$	Mesure de la dispersion des valeurs autour de la moyenne (GLCM mean)
GLCM Contrast	$\sum_{i,j=0}^{N-1} P_{i,j} (i, j)^2$	Mesure de la concentration des valeurs hors de la diagonale
GLCM Entropy	$\sum_{i,j=0}^{N-1} P_{i,j} (-\ln P_{i,j})$	Mesure de l'égalité de la repartition des éléments de la matrice

Nous envisagerons également l'utilisation des variances – notamment – comme indices alternatifs de texture. En effet, les indices de texture basés sur la GLCM impliquent une quantité de calculs impressionnante tandis que les variances et covariances sont calculées beaucoup plus rapidement et apportent néanmoins une information pertinente sur la texture des images au sein des objets.

2.6.2.3. Attributs de forme

Ces attributs sont calculés sur base des caractéristiques géométriques des segments. Le tableau 38 présente quelques attributs courants de forme.

Tableau 38. Formule de calcul d'attributs de forme couramment utilisés

Nom de l'indice	Formule	Signification
Length-to-width ratio	$\lambda = \frac{l}{w}$	Rapport entre la longueur et la largeur du plus petit rectangle encadrant l'objet, avec l , la longueur w , la largeur
Area	$A = \sum_{i=1}^n a_i$	Superficie, avec a_i , la superficie couverte par le pixel i
Length	$L = \sqrt{A \cdot \gamma}$	Indice de longueur
Width	$W = \sqrt{\frac{A}{\gamma}}$	Indice de largeur
Border Length	$bl = \sum_{i=1}^{n_i} e_i$	Périmètre de l'objet, avec e_i , la longueur du côté i de l'objet n_i , le nombre de cotés de l'objet
Shape Index	$si = \frac{bl}{4 \times \sqrt{A}}$	Indice de forme, calculé comme le rapport entre bl et le périmètre du carré de même superficie que l'objet
Border Index	$bo = \frac{bl}{2 \times (l + w)}$	Autre indice de forme, calculé comme le rapport entre bl et le périmètre du rectangle encadrant l'objet

2.6.2.4. Autres

L'utilisation d'autres types de « transformations » appliquées aux images ont été envisagées, notamment des filtres de détection des contours tels que SOBEL et passe-haut. Le filtre passe-haut a été appliqué sur les ORTHOS_DGA et a mené au calcul d'attributs (moyenne et écart-type) sur les segments.

2.6.3. Développement d'une application informatique pour la préparation de la base de données « images »

Une application informatique a été développée pour venir en appui à la préparation de la base de données « images ». Elle a été créée en langage VBA dans l'environnement Excel et fait appel à des outils de geoprocessing d'ArcGis et des geodatabases ESRI.

Cette application travaille à partir d'un *shapefile* de polygones (par exemple des îlots ou des objets de référence), des catalogues des ORTHOS_DGA, des ORTHOS dégradés et des deux scènes SPOT 5.

Les fonctionnalités de cette application sont :

extraire des différentes sources de données, les images dont les limites correspondent aux polygones du *shapefile* de base ;

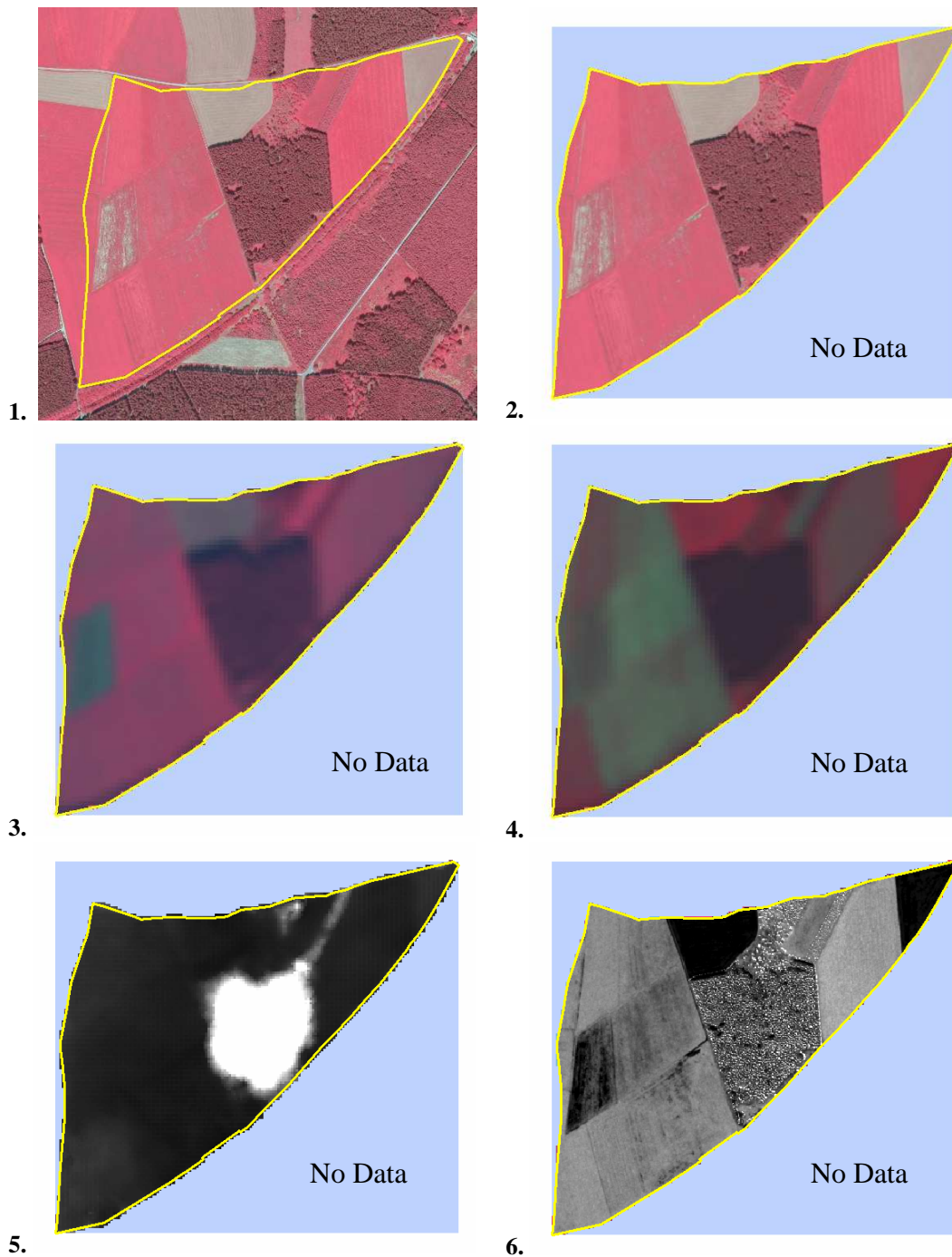
créer les couches correspondant à l'ensemble des néo-canaux retenus ;

calculer, pour chaque polygone, une série de statistiques des valeurs des pixels de l'ensemble des couches et bandes spectrales créées et les stocker dans une base de données afin notamment d'analyser des signatures spectrales.

La figure 94.1 illustre l'exemple d'un polygone représentant un îlot de la CNOSW avec en fond la couche formée des tuiles ORTHOS_DGA ; celui-ci pouvant se trouver à cheval sur plusieurs images. Les figures 94.2 à 94.6 illustrent différents exemples de résultats fournis par l'application.

L'application va créer une mosaïque des images constituant l'îlot. Cette mosaïque va être découpée selon les limites du polygone de référence avec comme résultat une image composée des mêmes canaux qu'au départ. Les pixels en dehors des limites du polygone sont transformés en *No Data* (figure 94.2). Ces opérations permettent de réduire le volume des données à traiter lorsqu'on décide de travailler à l'échelle d'un îlot par exemple. Les figures 94.3, 94.4 et 94.5 présentent l'îlot sélectionné avec en fond respectivement les images SPOT d'octobre, de juillet et le MNH HEST (issu de la soustraction du MNT ERRUISSOL au MNS_DGA).

Parallèlement au découpage des images, une série de néo-canaux sont également créés (voir la liste au tableau 36), par exemple le NDVI en figure 94.6.



Figures 94.1. Îlot sur fond ORTHOS_DGA, 2. Découpage de l'ORTHOS_DGA aux limites de l'îlot, 3. Découpage de la SPOT d'octobre, 4. Découpage de la SPOT de juillet, 5. Découpage du MNH, 6. Découpage du NDVI

N°	POLY_ID	ZONE_CODE	COUNT	AREA	MIN	MAX	RANGE	MEAN	STD	SUM
1	1	1	2218697	554674	0	255	255	98.7268	27.0922	219045000
2	2	1	263507	65876.8	1	245	244	113.334	16.8075	29864300
3	3	1	288643	72160.8	33	208	175	113.296	22.5612	32702100
4	4	1	1393019	348255	0	241	241	104.948	20.2333	146194000
5	5	1	253409	63352.3	0	251	251	81.5383	23.6102	20662500
6	6	1	1533405	383351	0	255	255	101.461	24.4019	155580000
...										
...										

Figure 95. Exemple d'une table résultat du calcul de statistiques du canal rouge des ORTHO_DGHA pour quelques îlots

Cette application a permis, pour tous les îlots sélectionnés (cf. paragraphe 2.4.3.1 b), la création des images à l'extension des îlots telles que les ORTHOS_DGA à 25cm et 2.5m de résolution, les deux scènes SPOT 5, la couche MNH et les néocanaux.

2.7. Analyses thématiques (GRFMN – Unité de Géomatique)

2.7.1. Introduction

Les analyses thématiques, dont il est question dans la tâche 7, ont pour objectif d'évaluer les possibilités de classification automatique ou semi-automatique relative à 3 types d'éléments présents dans le paysage rural : les territoires agricoles (correspondant à la classe 2 de la CNOSW), les forêts et milieux semi-naturels (correspondant à la classe 3 de la CNOSW) et les éléments arborés de type linéaire (alignements, haies, ...) ou ponctuel (arbres ou groupes d'arbres isolés) non repris dans la légende actuelle de la CNOSW.

La classification au sens de la télédétection correspond à la détermination de règles (statistiques, arithmétiques et logiques) qui aboutissent au choix des classes retenues ainsi qu'au mode de classement (Girard, 2004). Les règles de classification peuvent être regroupées en deux ensembles : les classifications supervisées et les classifications non supervisées.

Dans le cadre de cette tâche 7, les méthodes envisagées sont de type supervisé, c'est-à-dire basées sur une recherche d'éléments semblables à des éléments connus (objets de référence). Par souci de cohérence avec la chaîne de traitement visant la cartographie de l'occupation du sol qui passe par un stade de segmentation, les éléments d'entraînement visant l'apprentissage / l'établissement de l'algorithme de classification sont les segments eux-mêmes caractérisés par une occupation du sol. Cette dernière information est obtenue par le croisement de la couche « segments » avec la couche « vérité terrain » définie au préalable (cf. paragraphe 2.4.3.1).

D'un point de vue mise en œuvre, la tâche 7 a été menée en parallèle par les deux équipes. Deux approches statistiques ont été investiguées. L'équipe de l'Unité de Géomatique a évalué, de manière exploratoire, les possibilités offertes par les techniques d'analyse en composantes principales (paragraphe 2.7.3.) L'Unité GRFMN a quant à elle étudié les potentialités d'extraction des attributs les plus discriminants et de classification via une approche par arbres de décision (Classification and Regression Trees) (paragraphe 2.7.4.).

2.7.2. Calcul des attributs

2.7.2.1. Terminologie

Pour rappel, les attributs spectraux testés dans les phases de classification sont présentés au tableau 39 (formules, cf. paragraphe 2.6.2). Chaque canal ou néocanal est caractérisé par deux valeurs : la moyenne et l'écart-type calculés sur les pixels contenus dans les objets.

Pour faciliter la lecture du rapport, les appellations synthétiques des attributs sont renseignées sur une fiche annexée sous forme de fiche cartonnée « volante » (cette fiche est également fournie en annexe 1). Notons que les différents attributs comportent un suffixe permettant de différencier la moyenne (« _M ») de l'écart-type (« _Sd »). Par exemple, Ogdvi_M et Ogdvi_Sd correspondent respectivement à la moyenne et l'écart-type du néocanal GDVI calculé sur base des ORTHOS_DGA.

Tableau 39. Attributs spectraux calculés pour les tests de classification (moyenne et écart-type)

Sources de données				
	ortho_DGA	orthoDGA dégradées	SPOT 5 - octobre 2005	SPOT 5 - juillet 2006
bandes spectrales	R, V, B, NIR, bright	R, V, B, IR	NIR, R, V, MIR	NIR, R, V, MIR
	bright	bright		
	filtre passe-haut			
néocanaux	gdvi	gdvi	gdvi	gdvi
	gndvi	gndvi	gndvi	gndvi
	grvi	grvi	grvi	grvi
	ndvi	ndvi	ndvi	ndvi
	nir	nir	nir	nir

Les correspondances entre numéros de bandes et longueur d'onde captée pour les ORTHOS_DGA d'une part et les images SPOT 5 d'autre part, sont renseignées au tableau 40.

Tableau 40. Correspondance spectrale des bandes de l'ortho-image et des images SPOT.

	ORTHOS_DGA	SPOT
Bleu	Bande 3	
Vert	Bande 2	Bande 3
Rouge	Bande 1	Bande 2
Proche infrarouge (NIR)	Bande 4	Bande 1
Infrarouge moyen (MIR)		Bande 4

Les attributs texturaux calculés sont au nombre de 5 et sont basés sur les matrices de co-occurrence de niveaux de gris (GLCM) : contraste, entropie, corrélation, moyenne et écart-type (cf. tableau 37 pour définition).

Les attributs géométriques sont de deux types. Les premiers concernent la forme et la surface des segments en plan (*Area*, *Border index*, *Border length*, *Length*, *Length-to-width ratio*, *shape index*, *Width* ; cf. tableau 38 pour définition), tandis que les seconds concernent l'extraction de statistiques zonales de hauteur du sursol obtenue par comparaison du MNS_DGA et du MNT_ERRUISSOL (HEST, cf. paragraphe 2.3).

2.7.2.2. Mise en œuvre

a) Attributs spectraux

Une application VBA a été développée dans l'environnement du tableur Excel afin d'automatiser le calcul, pour chaque segment produit via Definiens ou Envi Ex, des attributs spectraux (cf. tableau 39). Celle-ci utilise des *geoprocessing* du logiciel ArcGIS et stocke les résultats produits dans une base de données Access. Préalablement, les ORTHOS_DGA dégradées, les images SPOT et les néocanaux produits à partir de ces images ont dû être rééchantillonnées à 50 cm afin de permettre le calcul de ces attributs pour tous les objets (tailles souvent petites et formes irrégulières).

b) Attributs texturaux et de forme

Les attributs de forme et de texture sont calculés à l'aide du logiciel Definiens 8.1. Pour poursuivre la comparaison des logiciels, nous avons forcé le calcul des attributs caractérisant les segments d'*Envi Ex* par le logiciel Definiens 8.1. Ceci a été réalisé via une « *chessboarding* » segmentation où intervient comme couche thématique le *shapefile* des objets produits par *Envi Ex*.

2.7.3. Analyse statistique des attributs des segments (Unité de Géomatique)

2.7.3.1. Introduction

Ce paragraphe rapporte l'analyse statistique des attributs extraits par segment dans le cadre des tâches 5 et 6. Son objectif est de comprendre les interactions entre attributs et de déterminer les attributs les plus discriminants dans une procédure de classification des segments afin d'éviter la redondance qui conduit à surestimer l'effet de certains facteurs dans la classification des segments.

Les données de référence qui ont été utilisées dans le cadre de cette partie d'étude sont celles qui étaient disponibles début juin 2006, c'est-à-dire les segments produits sous Definiens et Envi accompagnés des attributs présentés ci-avant ainsi que la vérité « terrain » sur les 43 îlots situés en zone SPOT produits lors de la première phase de récolte (cf. paragraphe 2.4.3.1b) et utilisés en phase de segmentation.

2.7.3.2. Analyse et résultats

Les segments produits par Envi et Definiens ont été étudiés séparément. Seuls les segments dont la vérité terrain est géométriquement la plus certaine ont été conservés. Il s'agit de ceux dont le segment est couvert à 90-100% par le polygone vérité terrain auquel il se superpose majoritairement. Certaines variables ont dû être soumises à une transformation logarithmique en raison de distributions asymétriques. Les matrices des corrélations et covariances peuvent être utilisées pour analyser les interactions entre variables attributaires des segments par ACP à la française permettant de choisir la matrice d'inertie à traiter.

La **matrice des covariances** a été exploitée dans un premier test car elle rend compte de l'importance de la variabilité des variables alors que les corrélations égalisent l'influence des variables combinées. La figure 96 montre la décroissance de la valeur propre d'une composante à la suivante pour les 10 premières composantes.

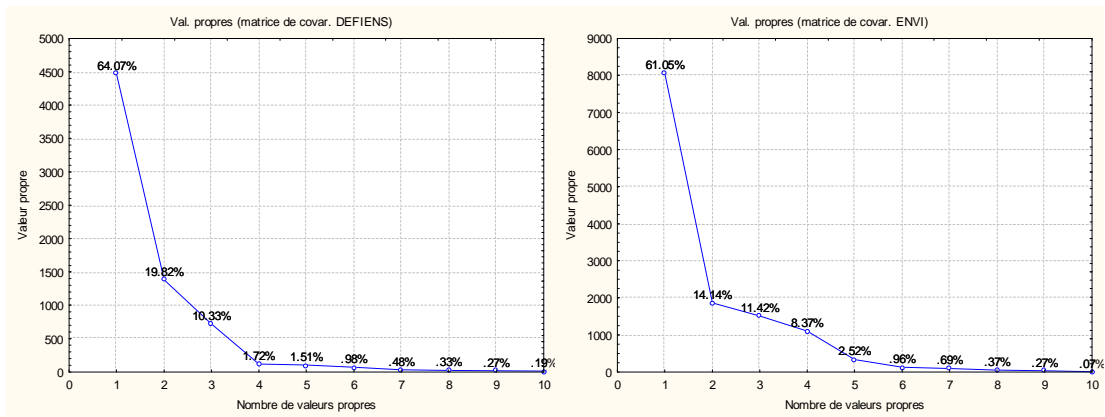


Figure 96. Valeur propre des 10 premières composantes extraites de la segmentation sous Definiens et sous Envi

Les associations entre variables attributaires ne sont pas clairement différentes selon que les produits de Definiens ou Envi soient exploités. La figure 97 montre ces associations pour Definiens. L'analyse des communautés sur 9 facteurs et des coordonnées factorielles des variables attributaires fait ressortir préférentiellement les variables qui ont de grandes variances, soit les attributs spectraux et dans une moindre mesure certains attributs texturaux. Une analyse basée sur les variables standardisées par exploitation est donc nécessaire pour mieux comprendre les associations. Elle constitue le second test effectué.

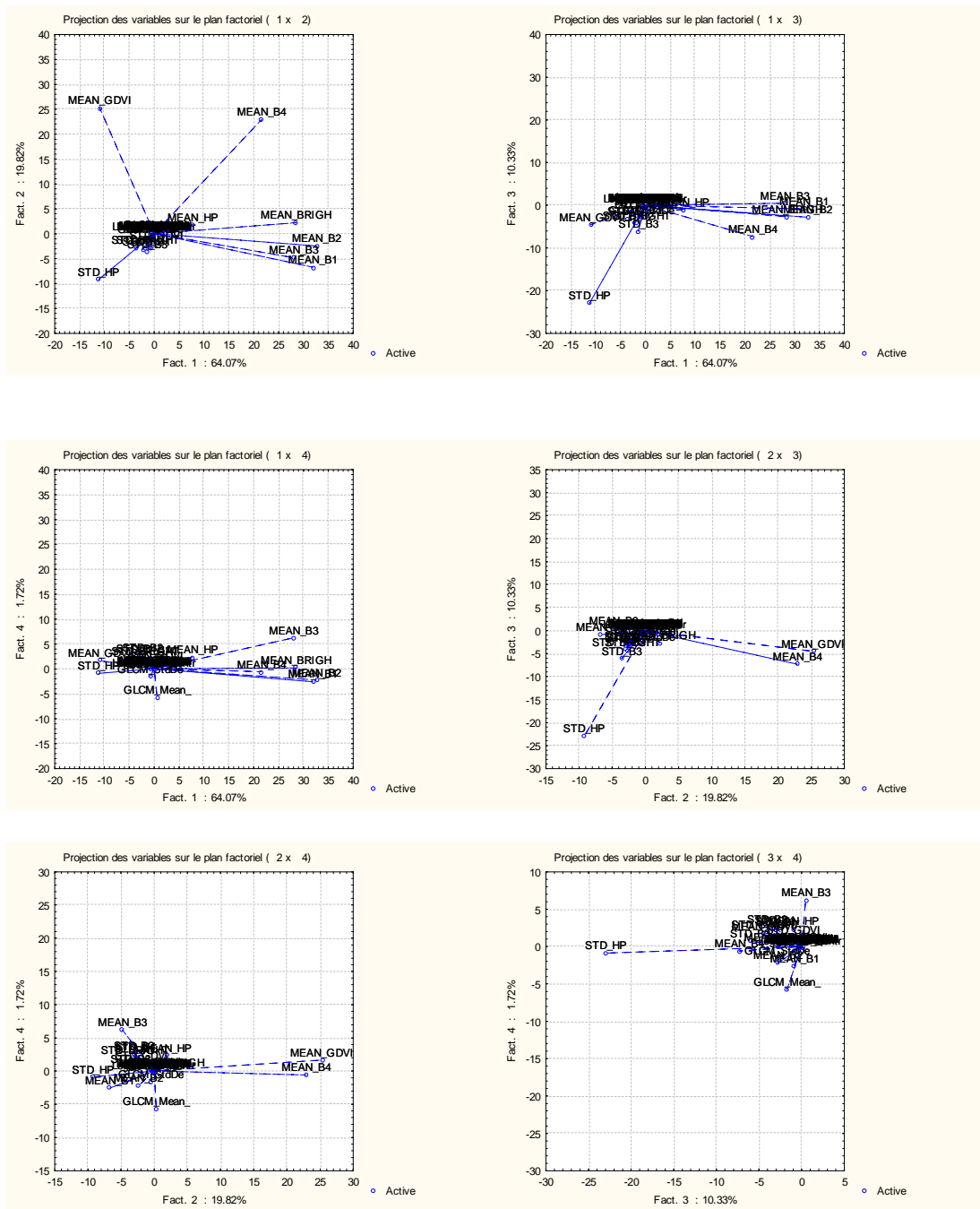


Figure 97. Coordonnées factorielles des variables attributaires dans l'espace des 4 premières composantes obtenues par segmentation sous DEFIENS en exploitant la matrice d'inertie des covariances

Le second test exploitant la matrice de corrélations met en évidence une matrice de corrélations singulière liée à la très forte redondance entre variables attributaires. La diagonalisation est donc basée sur le calcul de la matrice inverse généralisée selon l'aide en ligne du logiciel Statistica 8.0.

Toujours selon cette aide, cette transformation matricielle ne devrait théoriquement pas avoir de conséquences néfastes sur l'extraction des composantes et la signification des associations entre variables, mais elle peut produire des imprécisions dans les calculs, notamment celui des scores factoriels.

Onze composantes ont été extraites du produit de la segmentation sous Definiens. Elles expliquent 95.3 % de la variance. Une rotation de type varimax-normalisée a été appliquée aux 11 facteurs pour maximiser les saturations de chaque variable de départ vis-à-vis des composantes extraites.

Les communautés obtenues sur le produit de la segmentation sous Definiens (Tableau 43) montrent que les variables de variance zonales des attributs spectraux sont généralement introduites très tôt dans le modèle. Viennent ensuite les moyennes zonales des attributs spectraux. Suivent les attributs de dimension des segments, puis ceux caractérisant la longueur du périmètre et les attributs texturaux. Les attributs décrivant la forme des segments sont généralement introduits dans le modèle en dernier lieu. Certaines variables ne sont représentées par le modèle que très tard dans le modèle, soit à partir de l'introduction des composantes 9 à 11. Il s'agit des variables GLCM_Corre, GLCM_Mean et STD_GDVI.

Les valeurs propres et poids factoriel (ou saturations) des attributs après rotation sont reproduits dans le tableau 42. Sur base de ce tableau, une brève discussion de la signification des facteurs est proposée ci-dessous. Une illustration graphique de ces poids factoriels est présentée aux figures 98 et 99.

Les 7 premiers facteurs présentent après rotation une valeur propre très nettement supérieure à l'unité, variance correspondant à chacune des variables de départ puisque la matrice des corrélations est utilisée comme matrice d'inertie dans la procédure. Les facteurs 10 présentent quant à eux des valeurs proches et même inférieures à l'unité.

Tableau 41. Communautés obtenues par ACP sur la matrice des corrélations entre variables attributaires de la segmentation sous Définiens

Variable	Communautés (segments_definiens.sta) Extraction : Compos. principales Rotation : Varimax normalisé Condition d'inclusion : $\sqrt{\lambda} \geq 1$											
	Pour 1 Facteur	Pour 2 Facteurs	Pour 3 Facteurs	Pour 4 Facteurs	Pour 5 Facteurs	Pour 6 Facteurs	Pour 7 Facteurs	Pour 8 Facteurs	Pour 9 Facteurs	Pour 10 Facteurs	Pour 11 Facteurs	R-deux Multiple
Log_Area_pxl	0.026	0.027	0.910	0.911	0.935	0.941	0.962	0.984	0.990	0.991	0.995	1.000
Log_Border_ind	0.030	0.039	0.050	0.074	0.979	0.980	0.982	0.993	0.993	0.993	0.995	0.979
Log_Border_len	0.007	0.008	0.905	0.905	0.969	0.974	0.989	0.990	0.992	0.993	0.995	1.000
Log_GLCM_Contr	0.002	0.002	0.044	0.853	0.862	0.864	0.868	0.869	0.943	0.957	0.957	0.938
GLCM_Corre	0.092	0.100	0.173	0.176	0.184	0.185	0.190	0.209	0.925	0.930	0.934	0.841
GLCM_Entro	0.069	0.075	0.120	0.900	0.932	0.932	0.933	0.938	0.942	0.943	0.944	0.900
GLCM_Mean_	0.001	0.001	0.006	0.240	0.240	0.242	0.243	0.245	0.248	0.933	0.935	0.564
GLCM_StdDe	0.070	0.074	0.078	0.853	0.855	0.857	0.857	0.857	0.881	0.949	0.954	0.930
Log_Length_Pxl	0.018	0.023	0.889	0.891	0.898	0.903	0.915	0.992	0.992	0.992	0.993	1.000
Log_LengthWidt	0.000	0.014	0.027	0.029	0.044	0.044	0.045	0.985	0.993	0.994	0.996	1.000
Log_Shape_inde	0.030	0.031	0.032	0.046	0.903	0.904	0.906	0.988	0.991	0.991	0.994	1.000
Log_Width_Pxl	0.016	0.016	0.746	0.746	0.746	0.750	0.764	0.980	0.988	0.989	0.992	1.000
MEAN_B1	0.003	0.510	0.515	0.522	0.525	0.553	0.914	0.915	0.926	0.955	0.955	0.975
STD_B1	0.647	0.682	0.722	0.722	0.733	0.770	0.770	0.778	0.867	0.867	0.885	0.925
MEAN_B2	0.006	0.410	0.427	0.430	0.433	0.487	0.950	0.950	0.957	0.980	0.980	1.000
STD_B2	0.755	0.756	0.764	0.777	0.782	0.849	0.866	0.868	0.948	0.953	0.954	0.978
MEAN_B3	0.018	0.570	0.570	0.571	0.571	0.573	0.960	0.961	0.963	0.963	0.964	0.993
STD_B3	0.906	0.907	0.912	0.937	0.946	0.946	0.947	0.947	0.947	0.955	0.956	0.962
MEAN_B4	0.056	0.078	0.123	0.124	0.125	0.152	0.958	0.964	0.964	0.976	0.976	1.000
STD_B4	0.471	0.481	0.494	0.505	0.510	0.510	0.634	0.636	0.767	0.783	0.874	0.897
MEAN_GDVI	0.032	0.870	0.878	0.879	0.880	0.887	0.942	0.950	0.965	0.968	0.968	1.000
STD_GDVI	0.155	0.235	0.268	0.272	0.284	0.285	0.285	0.292	0.300	0.302	0.960	0.795
MEAN_GNDVI	0.001	0.781	0.793	0.794	0.798	0.921	0.944	0.945	0.955	0.960	0.964	0.991
STD_GNDVI	0.224	0.225	0.232	0.232	0.233	0.638	0.794	0.795	0.798	0.798	0.908	0.939
MEAN_GRVI	0.002	0.172	0.196	0.197	0.198	0.905	0.935	0.935	0.935	0.935	0.938	0.958
STD_GRVI	0.027	0.084	0.089	0.089	0.089	0.953	0.966	0.966	0.966	0.967	0.967	0.948
MEAN_HP	0.024	0.174	0.208	0.219	0.219	0.236	0.785	0.785	0.793	0.821	0.823	0.781
STD_HP	0.667	0.703	0.710	0.799	0.821	0.823	0.843	0.845	0.897	0.901	0.904	0.869
MEAN_NDVI	0.022	0.891	0.893	0.893	0.893	0.894	0.901	0.903	0.918	0.944	0.954	0.996
STD_NDVI	0.448	0.485	0.513	0.518	0.525	0.528	0.891	0.891	0.902	0.915	0.941	0.979
MEAN_NIR	0.006	0.906	0.908	0.910	0.911	0.950	0.971	0.972	0.973	0.978	0.983	0.998
STD_NIR	0.454	0.527	0.538	0.543	0.548	0.634	0.889	0.889	0.896	0.904	0.944	0.982

Tableau 42. Poids factoriels et valeurs propres après rotation obtenues par ACP sur la matrice des corrélations entre attributs des segments produits par Definiens.

Variable	Poids Factoriels(Varimax normalisé (segments_definiens.sta) Extraction : Compos. principales (Poids marqués >.700000) Condition d'inclusion : v6=1										
	Facte ur 1	Facte ur 2	Facte ur 3	Facte ur 4	Facte ur 5	Facte ur 6	Facte ur 7	Facte ur 8	Facte ur 9	Facte ur 10	Facte ur 11
Log_Area_pxl	-0.161	-0.031	-0.940	-0.036	-0.154	-0.082	-0.143	0.148	0.080	-0.025	-0.064
Log_Border_ind	0.174	0.095	-0.104	0.155	0.951	0.018	0.053	0.103	-0.010	-0.017	0.038
Log_Border_len	-0.084	-0.022	-0.947	0.016	0.251	-0.073	-0.123	0.022	0.053	-0.025	-0.040
Log_GLCM_Contr	0.048	0.006	0.203	0.900	0.093	0.045	0.063	-0.030	-0.273	0.116	-0.004
GLCM_Corre	0.304	0.089	-0.270	-0.056	-0.086	-0.033	-0.071	0.140	0.846	0.070	0.062
GLCM_Entro	0.263	0.077	-0.212	0.883	0.179	-0.018	-0.027	0.071	0.064	-0.026	0.031
GLCM_Mean_	-0.034	0.014	0.065	0.484	-0.023	0.036	-0.044	-0.040	0.058	0.828	0.040
GLCM_StdDe	0.264	0.068	0.063	0.880	0.050	0.044	-0.004	0.018	0.153	0.262	-0.072
Log_Length_Pxl	-0.135	-0.073	-0.930	-0.049	0.085	-0.069	-0.108	-0.277	0.029	-0.002	-0.028
Log_LengthWidt	0.020	-0.116	0.116	-0.043	0.121	0.006	0.040	-0.969	-0.092	0.027	0.049
Log_Shape_inde	0.174	0.020	-0.040	0.118	0.926	0.019	0.042	-0.287	-0.059	-0.000	0.052
Log_Width_Pxl	-0.126	0.023	-0.854	-0.009	-0.016	-0.062	-0.118	0.465	0.090	-0.020	-0.059
MEAN_B1	-0.051	-0.712	-0.070	-0.086	-0.049	-0.169	-0.601	-0.025	-0.108	0.169	0.000
STD_B1	0.804	-0.187	0.199	0.027	0.105	0.192	0.012	-0.090	0.297	-0.023	0.132
MEAN_B2	-0.080	-0.635	-0.132	-0.052	-0.054	-0.232	-0.680	0.001	-0.088	0.149	0.001
STD_B2	0.869	-0.015	0.091	0.113	0.070	0.259	0.130	-0.052	0.282	-0.072	0.025
MEAN_B3	-0.133	-0.743	-0.023	-0.024	-0.002	-0.045	-0.623	-0.018	0.048	0.018	0.020
STD_B3	0.952	-0.030	0.066	0.161	0.093	-0.015	0.018	-0.004	0.010	0.090	0.032
MEAN_B4	-0.237	0.148	-0.213	-0.020	-0.036	-0.165	-0.898	0.080	0.015	0.107	0.021
STD_B4	0.686	0.100	0.113	0.105	0.071	0.001	0.353	0.033	0.362	-0.126	0.302
MEAN_GDVI	-0.179	0.915	-0.090	0.038	0.022	0.084	-0.235	0.090	0.121	-0.051	0.024
STD_GDVI	0.394	-0.282	0.184	-0.059	0.109	0.027	-0.015	-0.087	0.086	0.047	0.811
MEAN_GNDVI	-0.037	0.883	0.109	0.040	0.060	0.351	0.152	0.014	0.104	-0.067	-0.065
STD_GNDVI	0.473	0.032	0.082	0.022	0.035	0.636	0.395	-0.017	0.054	-0.005	0.332
MEAN_GRVI	0.046	0.413	0.155	0.030	0.013	0.841	0.172	-0.013	-0.019	0.010	-0.051
STD_GRVI	0.163	0.240	0.068	0.025	0.007	0.929	0.115	0.000	-0.023	0.027	-0.014
MEAN_HP	-0.155	-0.387	-0.185	0.105	-0.014	-0.128	-0.741	0.028	0.085	-0.168	0.044
STD_HP	0.817	0.189	0.085	0.299	0.147	-0.041	0.143	0.043	-0.228	-0.065	-0.053
MEAN_NDVI	0.148	0.932	-0.041	0.021	0.006	-0.035	0.079	0.044	-0.126	0.160	-0.101
STD_NDVI	0.669	0.193	0.166	0.076	0.079	0.060	0.602	0.023	-0.104	0.115	0.161
MEAN_NIR	0.079	0.949	0.046	0.032	0.035	0.199	0.144	0.029	-0.027	0.070	-0.075
STD_NIR	0.674	0.271	0.103	0.069	0.071	0.293	0.505	0.009	-0.084	0.088	0.199
Var. Expl.	5.287	5.571	3.874	2.826	2.011	2.499	3.730	1.416	1.302	0.962	1.001
Prp.Tot	0.165	0.174	0.121	0.088	0.063	0.078	0.117	0.044	0.041	0.030	0.031

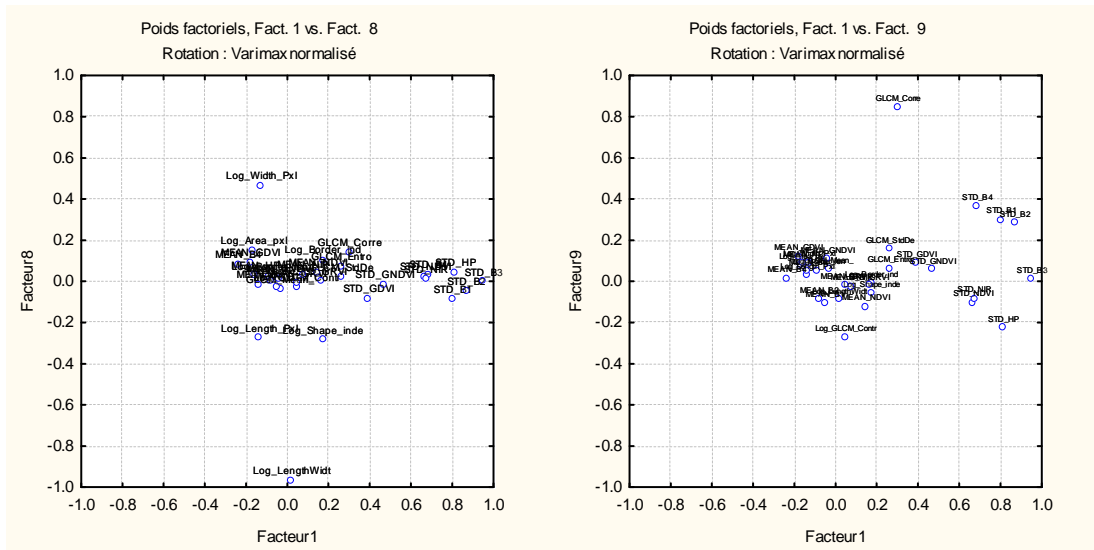


Figure 99. Position des variables attributaires des segments obtenus sous Definiens dans l'espace des facteurs obtenus par ACP appliquée à la matrice des corrélations. (suite)

Facteur 1

Il s'agit clairement d'un facteur mesurant **la dynamique** qui est clairement mise en évidence par les saturations positives écart-types zonaux sur B1 à 4, NIR, NDVI et HP.

Facteur 2

Le facteur 2 semble mesurer l'influence de la **végétation** sur les signatures spectrales moyennes zonales. Les saturations positives des moyennes zonales sur le GDVI, GNDVI, NIR et NDVI en attestent.

Les trois autres variables qui sont aussi clairement mises en évidence par ce facteur sont les moyennes zonales des bandes B1, B2, B3 dont les saturations sont négatives. Ceci montre clairement la redondance de ces variables attributaires avec les indices cités ci-dessus.

Une seule des variables synthétisées par ce facteur devrait donc suffire dans la caractérisation des segments. Un des indices semble être le plus adapté et le plus aisément interprétable.

Facteur 3

Le facteur trois traduit la **petitesse de taille** (corrélation négative) des segments. Il se caractérise par des saturations négatives des variables attributaires Length_Pxl, Area_pxl, Border_len et Width_Pxl. Une seule de ces variables serait donc suffisante pour caractériser les segments. La plus évidente à interpréter est Area_pxl.

Facteur 4

Le facteur 3 traduit la **texture**. Il synthétise les variables attributaires GLCM_contr, GLCM_entro et GLCM_StDe présentant toutes une saturation positive. Il est assez surprenant que ces variables texturales soient regroupées indépendamment des écart-types zonaux corrélés au facteur 1. Il existe donc deux mesures indépendantes de caractéristiques qui semblent liées, la dynamique et la texture. La seconde mesure prend certainement mieux en compte les agencements spatiaux alors que la première ne la fait pas du tout.

Facteur 5

Ce facteur synthétise des variables de **forme** et plus particulièrement la **compacité** et la **régularité de la forme** des segments telles que border_ind et shape_ind qui présentent une saturation positive.

Facteur 6

Les seules variables corrélées positivement avec ce facteur sont les moyennes et écart-types zonaux de l'indice de végétation GRVI (NIR/Green) et dans une moindre mesure du GNDVI qui mesurent des caractéristiques spectrales de la végétation tout à fait indépendantes des autres indices de végétation synthétisés par le facteur 2. Un autre constat concernant le premier de ces indices est la redondance entre sa moyenne et son écart-type zonaux due à un effet d'échelle très prononcé. Le calcul d'un coefficient de variation est donc peut-être justifié.

Facteur 7

De fortes valeurs positives de saturation sont constatées pour les moyennes zonales de B4 et HP, alors qu'elles sont négatives pour les écart-types zonaux des indices de végétations NDVI, GNDVI, NIR et de la bande B4. Ce facteur met en évidence les fortes intensités présentant de faibles dynamiques dans la bande 4 et inversement les faibles intensités présentant les plus fortes dynamiques. Ce cas de figure semble correspondre à l'influence des segments forestiers. L'interprétation de ce facteur n'est cependant pas claire !

Facteur 8 (facteur peu significatif sur base de sa valeur propre)

La variable attributaire morphologique Log_LengthWidth, mesurant l'**asymétrie** et l'**allongement**, est corrélée négativement avec ce facteur.

Facteur 9 (facteur peu significatif sur base de sa valeur propre)

Facteur mesurant la texture indépendamment des autres variables de texture synthétisées par le facteur 4 vu la valeur positive de la saturation de GLCM_Correlation. L'interprétation reste cependant ambiguë.

Les facteurs 10 et 11 ont une signification trop peu importante et n'ont pas été interprétés. Ils sont associés généralement à une seule variable de plus faible variance. Néanmoins, cette variable mesure une caractéristique indépendante et complémentaire des autres. Elle peut donc s'avérer discriminante dans la procédure de classification des segments.

La conclusion de cette analyse permet d'orienter le choix des variables nécessaires à la discrimination des segments dans une procédure de classification. Les variables qui garantissant la discrimination sans redondance doivent être choisies dans chaque groupe associé à un facteur.

Une analyse discriminante exploitant la vérité terrain a été envisagée en séparant les segments en deux groupes : Groupe 1, agricole, et groupe 2, résineux, feuillus, mixte R/F.

A la suite de ce travail, nous n'avons pas obtenu de résultat probant étant donné le nombre important d'individus et la signification ambiguë des classes de la vérité terrain. Dans celle-ci, on ne trouve que les groupes suivants : résineux, feuillus, mixtes f/r, agricole, mise à blanc, autre, vide. Il ne semble pas judicieux de réaliser une analyse discriminante sur base de ces groupes, étant donné qu'il existe probablement une plus grande variabilité de caractéristiques au sein même du groupe agricole qu'entre les feuillus et les résineux par exemple.

2.7.4. Analyse intégrée des attributs et classification (GRFMN)

2.7.4.1. Méthodologie

a) Arbres de décision

L'algorithme CART (« *Classification and Regression Trees* ») (Breiman *et al.*, 1984) a été utilisé dans le cadre de cette étude afin d'établir un schéma de classification. Cet algorithme est basé sur la recherche de partitions dichotomiques univariées au départ des attributs fournis (Brostaux, 2005). A chaque nœud de l'arbre, un test compare la valeur d'un attribut du segment à classer à une valeur seuil et, selon le résultat du test, le segment est soit classé (nœud terminal ou « feuille »), soit dirigé vers un autre nœud et ainsi de suite jusqu'à être classé (exemple d'un arbre en figure 100). On observe que le même attribut peut apparaître dans plusieurs nœuds et qu'une même classe peut constituer plusieurs « feuilles ». Cette double possibilité illustre la flexibilité de la méthode à traiter des données à grande variabilité intra-classe.

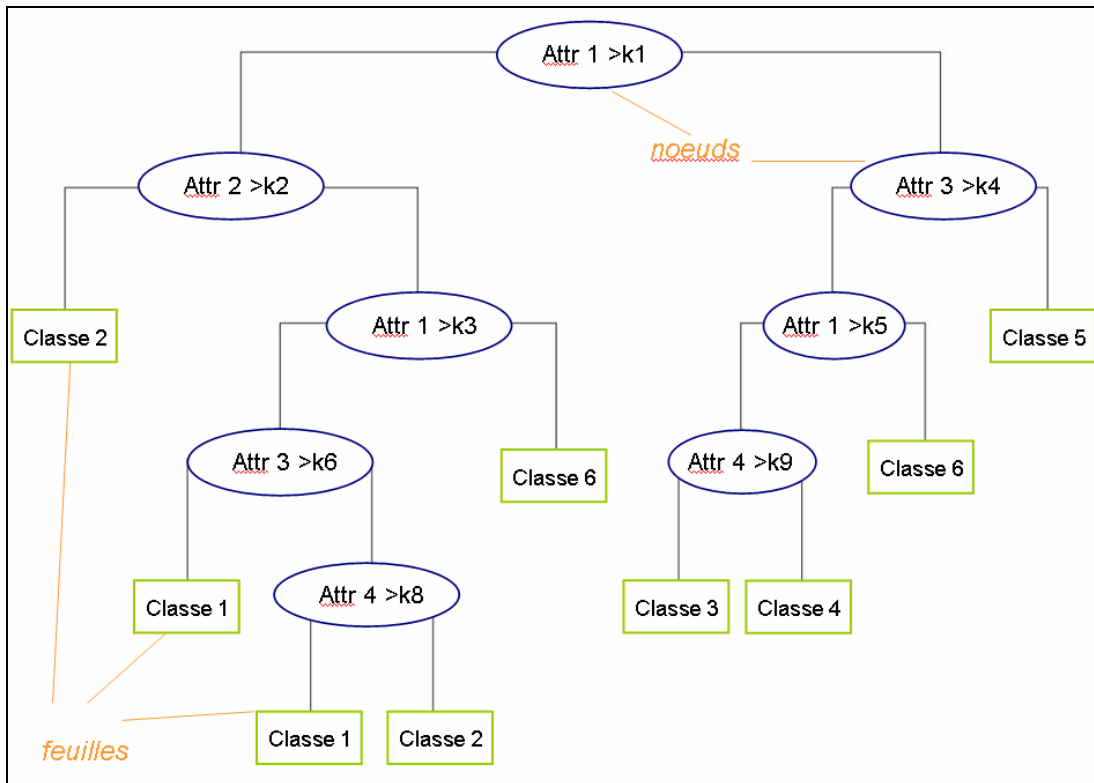


Figure 100. Exemple schématique d'un arbre de décision

Lors de l'élaboration de l'arbre, chaque partition candidate au statut de nœud – c'est à dire chaque condition portant sur un attribut – est évaluée par l'intermédiaire d'une valeur caractérisant son impureté, l'indice de Gini (cf. équation 3). La partition retenue est celle présentant une réduction maximale d'impureté entre le nœud père et ses deux fils.

Équation 3. Indice de Gini du nœud E, avec c : le nombre de classes, f_j : la fréquence de la classe j dans le nœud

$$Gini(E) = \sum_{j=1}^c f_j (1 - f_j)$$

La croissance de l'arbre est contrôlée via la mesure du taux d'erreur de classification à l'intérieur des nœuds ou par un nombre d'individus minimal par nœud. La classe attribuée à une feuille est celle dont le maximum d'individus a atteint celle-ci.

Pour diminuer la complexité de l'arbre, celui-ci peut être élagué. La phase d'élagage consiste à extraire au départ d'un arbre complètement développé, le meilleur sous-arbre, obtenu par suppression d'une ou plusieurs branches choisies par l'intermédiaire d'une recherche le plus souvent heuristique.

L'élagage a dans cette étude été basé sur la méthode de coût-complexité minimal (« *minimal cost complexity pruning* »), développée par Breiman et al. (1984) pour l'algorithme CART.

D'un point de vue opérationnel, les arbres de décision ont été construits dans le logiciel statistique *open source* « R » via la fonction « *rpart* » du *package* du même nom. L'élagage a été opéré via la fonction « *prune* » du même *package* et s'est basé sur la minimalisation de l'erreur en resubstitution calculée via une technique de validation croisée.

b) Forêts aléatoires

L'algorithme *Random Forests* (Breiman, 2001) est basé sur l'agrégation d'arbres de décision randomisés. Le processus de construction des arbres est mis en œuvre d'une façon légèrement différente par comparaison au CART. En effet, une randomisation est opérée à la fois au niveau de l'échantillon d'apprentissage et des partitions internes (c'est-à-dire des conditions sur les attributs).

Une forêt aléatoire peut être utilisée comme classificateur : l'individu à classer est propagé dans chaque arbre de la forêt et étiqueté en fonction des règles CART. La prédiction globale de la forêt est donnée par un vote à la majorité simple des attributions de classe des arbres individuels.

Dans le cadre de cette étude, cet algorithme n'a pas été utilisé en tant que tel pour classer chaque segment car cela alourdirait la chaîne de traitement, d'autant plus si elle est envisagée pour une production à grande échelle. En outre, ce classificateur n'est pas retranscriptible dans un logiciel de traitement d'images tandis qu'un arbre de décision l'est facilement dans Definiens par exemple.

Par contre, les forêts aléatoires ont été utiles à la quantification de la performance des différents attributs dans la classification. En effet, elles permettent d'évaluer l'importance des différents attributs via notamment un indice appelé la "MeanDecreaseGini" (MDG). Cet indice correspond à la diminution totale en impureté de nœuds suite à la prise en compte d'un attribut supplémentaire. Cette diminution est moyennée sur tous les arbres de la forêt (Liaw and Wiener, 2002). L'impureté de nœuds est mesurée par l'indice de Gini (cf. équation3).

Après la classification des attributs par MDG décroissante, les attributs sélectionnés sont injectés dans un arbre de décision produit par l'algorithme CART.

Ces forêts aléatoires ont été mises en œuvre dans le logiciel statistique « R » via la fonction « *randomForest* » du *package* du même nom.

c) Objets d'entraînement et analyse de la qualité

Lors de l'établissement des règles de classification, un set d'îlots dont l'occupation du sol de référence a été produite en tâche 4 a été utilisé. Ce set de 70 îlots (22 km²) – que nous appellerons **set n°1** – représente en superficie environ la moitié du set de 123 îlots disponibles dans la zone « SPOT ». Ce premier set d'îlots est constitué des 43 îlots déjà utilisés en phase de segmentation auxquels en ont été ajoutés 27 autres.

Le reste des îlots (53 pour une superficie totale de 21,8 km²) – le **set n°2** – sera utilisé pour l'évaluation de la **qualité de l'entièreté de la chaîne de traitement**. Ces derniers îlots sont complètement indépendants des processus d'élaboration de règles de segmentation et de classification.

Les superficies des différentes classes d'occupation du sol sont détaillées dans le tableau 43 pour les deux sets d'îlots. On peut y voir que la répartition des classes d'occupation du sol est équivalente entre les 2 sets de données.

Tableau 43. Répartition des différentes classes d'occupation du sol dans les sets d'établissement des règles de classification d'une part (set n°1) et de vérification de la chaîne de traitement d'autre part (set n°2), superficies et pourcentages.

Occupation	Surface totale (km ²)	Set n° 1		Set n° 2	
		Surface (km ²)	pourcentage	Surface (km ²)	pourcentage
Agricole	39.92	5.9	27	7.21	33
Autre	1.36	0.22	1	0.14	1
Feuillus	45.24	6.98	32	6.55	30
Mixte	7.39	1.16	5	0.96	4
Résineux	39.56	6.12	28	5.63	26
Vide	10.32	1.63	7	1.28	6
Total	143.78	22.01	100	21.77	100

Le nombre de segments par classe et par set d'îlots produits par les deux logiciels étudiés est renseigné au tableau 44.

Tableau 44. Nombre de segments produits par set d'îlots pour les deux logiciels étudiés : Definiens 8.1 et Envi Ex

Set n° 1		Set n° 2	
Definiens 8.1	Envi Ex	Definiens 8.1	Envi Ex
9 770	167 349	8 230	246 042

Nous observons, tel que déjà mentionné au paragraphe 2.5.3, que le volume de données à traiter est considérablement plus élevé pour Envi que pour Definiens.

Lors de l'établissement des arbres de décision, trois-quarts des segments du set n°1 est sélectionné aléatoirement et est utilisé pour servir à la construction de l'arbre de classification. Le quart restant est réservé pour caractériser la classification produite via une matrice de confusion et notamment la valeur de l'*overall accuracy* (Lillesand *et al.*, 2008). Afin d'entraîner le logiciel avec des objets représentatifs de la réalité, les objets d'entraînement ont été limités aux segments ayant 85% et plus de leur superficie sur la même occupation du sol.

Ensuite quand un arbre de décision est choisi, la qualité globale de l'ensemble du processus de segmentation/classification est évaluée sur les îlots du set n°2.

2.7.4.2. Résultats et discussion

a) Introduction

Les méthodes d'arbres de décision et de forêts aléatoires utilisées dans le cadre de cette étude ont un aspect empirique qu'il convient de noter. En effet, on ne produit pas « l'arbre » de décision parfait menant à la meilleure classification. Ces méthodes font intervenir des échantillons d'apprentissage et sont donc tributaires notamment de leur choix.

b) Definiens

Flow chart

Le *flow chart* en figure 101 illustre la succession d'étapes et les différents tests réalisés sur le set de données issus de Definiens.

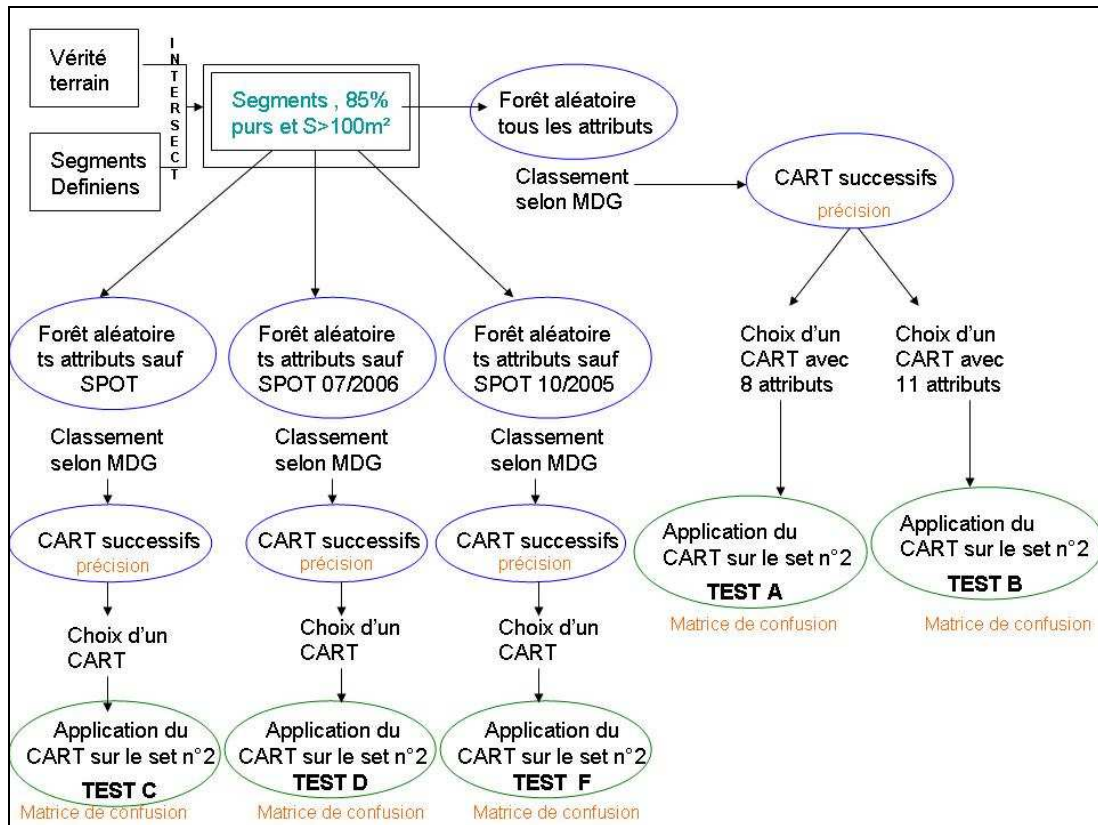


Figure 101. Flow chart des tests de classification réalisés à partir des résultats de segmentation de Definiens

avec : MDG : Mean Decrease Gini

Forêt aléatoire

Une forêt aléatoire a été réalisée sur tous les segments produits par Definiens sur le set d'îlots n°1. Celle-ci a permis de caractériser l'importance des attributs selon la « Mean Decrease Gini » (MDG) (cf. paragraphe b)). Les valeurs de ce paramètre pour les attributs testés sont renseignées au tableau 45.

Tableau 45. Mean Decrease Gini pour les attributs étudiés, classement selon la MDG décroissante

Attribut	MDG	Attribut	MDG	Attribut	MDG
S7b3_M	329.09	GLCM_Contr	51.97	Odb3_Sd	27.72
S7b4_M	281.18	Odgndvi_M	51.91	Ohp_M	27.50
S7b2_M	261.81	Odgrvi_M	47.86	S7b3_Sd	26.39
S7gndvi_M	226.11	S10ndvi_M	47.62	Area_Pxl	26.38
S10b3_M	211.65	S10gndvi_M	47.45	S7b2_Sd	25.94
S7nir_M	201.23	GLCM_StdDe	45.57	Ob3_M	25.10
S7ndvi_M	192.33	Ogrvi_M	44.79	Odgndvi_Sd	24.27
S10b4_M	188.71	S10nir_M	42.46	Length_Pxl	23.68
Hest_M	186.32	GLCM_Entro	42.40	Width_Pxl	23.21
Ohp_Sd	167.90	Ob4_M	41.74	GLCM_Corre	21.86
S7gdvi_M	159.09	Odb1_M	41.68	LengthWidt	21.51
S10b2_M	151.92	Ondvi_M	41.14	Odnir_Sd	20.86
S7b1_M	111.62	Odgdvi_Sd	40.77	S10b4_Sd	20.57
Ob4_Sd	100.33	Odb2_M	40.59	Border_len	20.31
S10b1_M	88.21	Ogdvi_Sd	40.15	Ondndvi_Sd	19.64
Odb4_Sd	78.44	Odnir_M	40.13	S7b4_Sd	19.53
S7grvi_M	72.15	Odb3_M	39.27	S7gdvi_Sd	18.97
Odbr_Sd	71.26	Ondndvi_M	38.34	S7gndvi_Sd	18.74
Odb2_Sd	66.82	Onir_Sd	37.28	S7nir_Sd	17.87
Obr_Sd	65.25	Ogrvi_Sd	35.97	S7b1_Sd	17.85
Odgdvi_M	64.27	Obr_M	35.10	Shape_inde	17.40
Odb1_Sd	63.92	S7grvi_Sd	35.01	Border_ind	17.03
Odbr_M	59.32	Onir_M	33.30	S7ndvi_Sd	16.90
Odb4_M	59.29	S10grvi_M	33.04	S10b3_Sd	16.46
Ob2_Sd	58.77	Ob1_M	33.00	S10b2_Sd	15.69
Ogdvi_M	56.24	Ob2_M	32.64	S10gdvi_Sd	14.27
S10gdvi_M	55.46	Ogndvi_M	31.52	S10b1_Sd	13.65
Hest_Sd	54.27	Odgrvi_Sd	29.43	S10ndvi_Sd	13.49
Ob3_Sd	52.61	GLCM_Mean_	28.97	S10grvi_Sd	13.21
Ondvi_Sd	52.17	Ob1_Sd	28.50	S10nir_Sd	12.85
		Ogndvi_Sd	28.22	S10gndvi_Sd	12.44

On remarque que les attributs des images SPOT occupent majoritairement les 15 premières places avec Hest_M (moyenne de la couche de hauteur HEST, MNH voir paragraphe 2.3.1.5) et le filtre passe-haut appliqué aux ORTHOS_DGA.

CART

Suite à la caractérisation par une forêt aléatoire de l'importance de chaque attribut selon la MDG associée, nous avons généré, avec 75 % des données du set d'îlots n°1, une succession d'arbres via la méthode CART, ces arbres intégrant un attribut supplémentaire à chaque lancement (les attributs étant injectés dans l'ordre décroissant de MDG). Pour chaque arbre, un indice de qualité est calculé sous la forme de l'*overall accuracy* (précision globale) extraite d'une matrice de confusion produite sur le quart du jeu de données restant. Cette démarche visait à produire un arbre présentant un rapport temps de calcul / précision intéressant dans une optique de production à grande échelle.

La procédure détaillée construite dans le logiciel R est la suivante :

- connexion à la base de données résultats ;
- chargement de la table d'attributs produite par l'intersection (geoprocessing) des segments avec la vérité terrain (sur le set d'îlots n°1). Cette table a été préalablement limitée aux objets ayant plus de 85% de leur surface sur la même occupation du sol et dont la superficie est supérieure à 100 m². Le nombre de segments/classe provenant de cette table est présenté au tableau 46.

Tableau 46. Nombres de segments d'apprentissage par classe (85% de pureté d'occupation du sol, superficie>100m²)

Occupation	n segments
Agricole	1 403
Autre	245
Feuillus	4 532
Mixte	669
Résineux	1 094
Vide	545

- constitution aléatoire d'un set de données contenant le même nombre de segments par classe d'occupation du sol (limitation à 245 segments, qui est le nombre de segments de la classe « Autre », cf. tableau 46) ;
- *split* du set de données en $\frac{3}{4}$, $\frac{1}{4}$ (toujours en respectant l'égalité du nombre de segments par classe);

- importation de la table produite lors de la forêt aléatoire décrite ci-dessus fournissant notamment la valeur de la MDG pour chaque attribut. Classement de ceux-ci par MDG décroissante ;
- boucle sur le nombre d'attributs disponibles avec production d'un CART sur base du set contenant $\frac{3}{4}$ des données d'apprentissage, élagage du CART, tracé et export de celui-ci, prédiction sur le dernier $\frac{1}{4}$ du set de données, export de la matrice de confusion de celui-ci (avec notamment l'*overall accuracy*)

La figure 102 présente l'évolution de l'*overall accuracy* en fonction du nombre d'attributs insérés dans le CART. La succession d'étapes décrites *supra* a été répétée cinq fois, menant à l'établissement des 5 courbes de la figure 102.

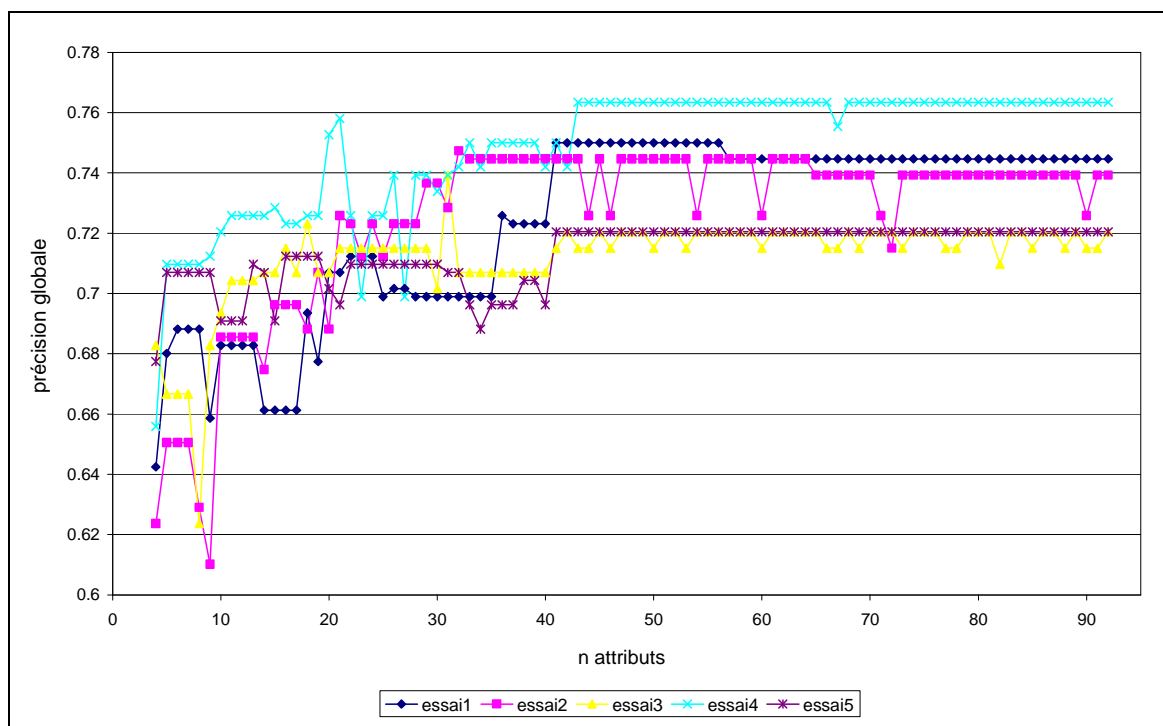


Figure 102. Evolution de la précision globale (*overall accuracy*) en fonction du nombre d'attributs insérés dans le CART, différents essais

On peut voir que le processus présente une composante aléatoire (courbes différentes d'un essai à l'autre). En effet, l'arbre produit et le résultat affiché dépendent des sets d'objets aléatoirement sélectionnés d'une part pour l'entraînement de l'arbre, d'autre part pour l'évaluation de celui-ci. Il est à noter que le nombre d'attributs finalement retenus dans le CART n'est pas égal à celui injecté. En effet, l'élaboration du CART en elle-même ne retient qu'une partie des attributs, certains pouvant ensuite disparaître avec le(s) nœud(s) associé(s) lors de l'étape d'élagage.

Tests

Nous avons décidé de tester sur le set d'îlots n°2, deux CART sélectionnés sur base de la précision globale obtenue (cf. figure 102). Ces tests seront appelés « test A » et « test B ».

Pour le CART du test A, 8 attributs sont retenus à l'issue du processus (cf. tableau 47). La figure 103 présente l'architecture de l'arbre produit pour ce même test. Notons que par convention d'illustration, les objets répondant à la condition définissant le nœud sont dirigés vers la partie gauche de l'arbre.

Tableau 47. Attributs utilisés dans le CART du test A

Attributs utilisés
S7b3_M
Obr_Sd
S10b1_M
Ohp_Sd
S10b3_M
S7b4_M
S7gdvi_M
Odgdvi_M

La première grande distinction créant une partie gauche et une partie droite dans l'arbre s'opère avec la moyenne de la bande 3 (Vert) de l'image SPOT de juillet 2006. Celle-ci sépare les classes arborées – « Feuillus », « Résineux », « Mixte » – des autres : « Agricole », « Autre » et « Vide » (classe en forêt mais ayant une végétation non arborée pour la plupart des exemples). Les autres nœuds opèrent leur séparation dichotomique sur base d'attributs uniquement spectraux sur base des ORTHOS_DGA, SPOT 5 d'octobre 2005 et de juillet 2006.

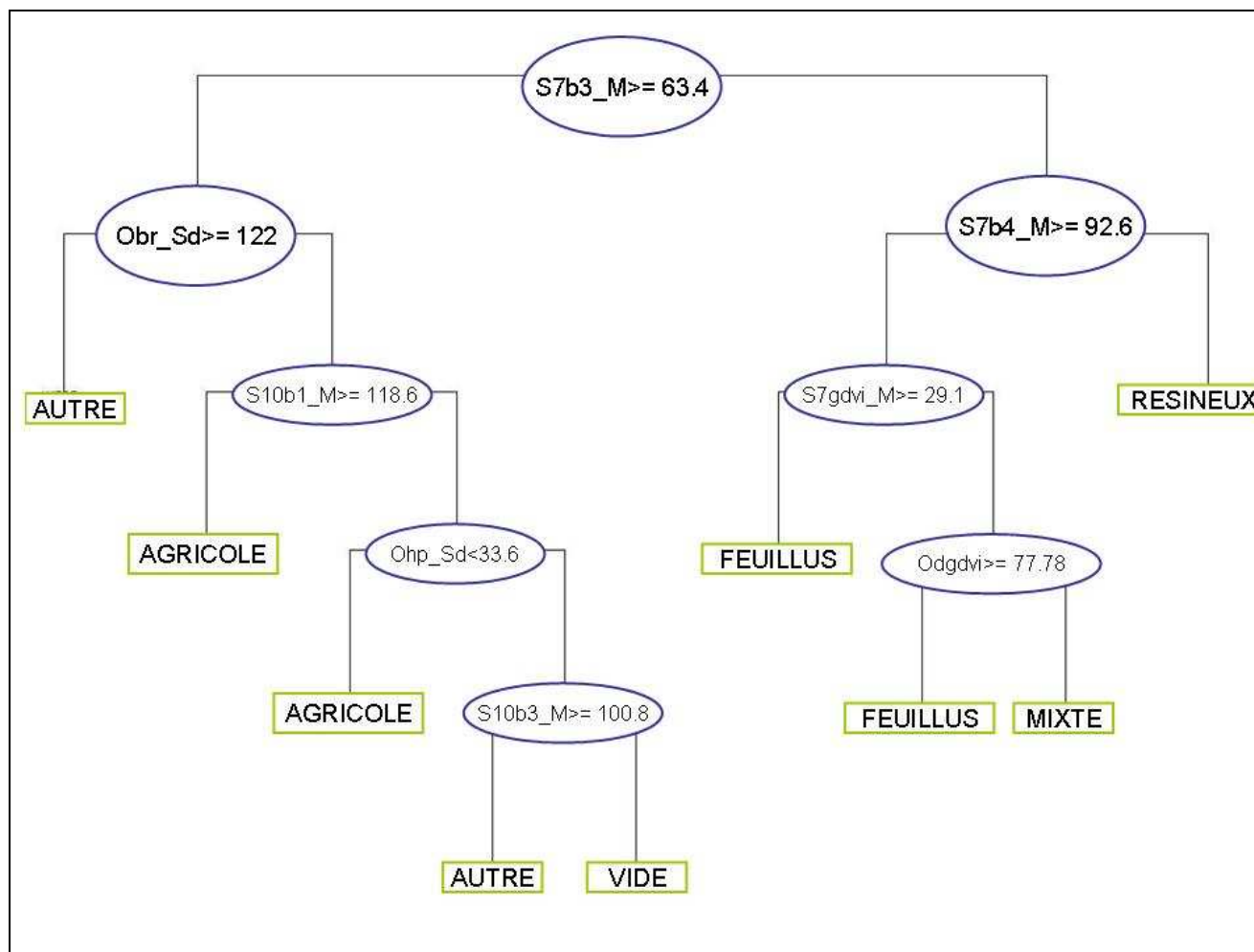


Figure 103. Architecture de l'arbre CART utilisé dans le test A

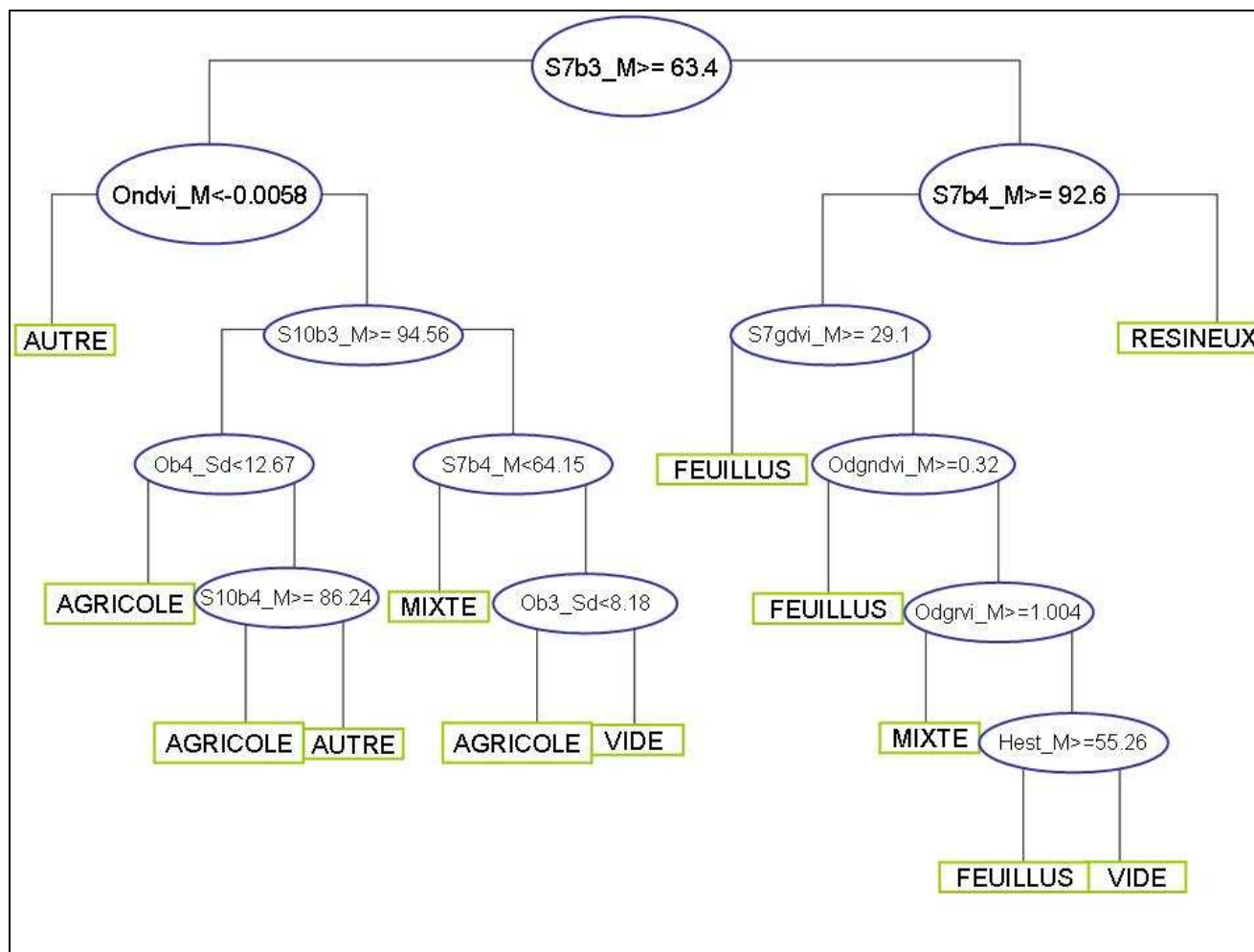


Figure 104. Architecture de l'arbre CART utilisé dans le test B

Pour le CART qui sera utilisé dans le test B (Figure 104), seuls 11 attributs de 43 présélectionnés sont finalement utilisés (cf. tableau 48).

Tableau 48. Attributs utilisés dans le CART du test B

Attributs utilisés
S7b3_M
Ondvi_M
S10b3_M
Ob4_Sd
S7b4_M
S10b4_M
Ob3_Sd
S7gdvi_M
Odgndvi_M
Odgndvi_M
Hest_M

La première grande distinction s'opère de nouveau avec la moyenne de la bande 3 (Vert) de l'image SPOT de juillet 2006. Celle-ci sépare les classes arborées des autres. Les autres nœuds opèrent leur séparation dichotomique sur base d'attributs spectraux et de Hest_M.

Pratiquement, la procédure mise en œuvre pour le test global (segmentation+classification) réalisé sur le set d'îlots n°2 est la suivante :

- production des segments dans Definiens (*scale parameter* :125, *shape* : 0.3 et *compactness* : 1.0)
- calcul des attributs sur ces segments (cf. paragraphe 2.7.2)
- intersection (geoprocessing) des segments avec la vérité terrain → table d'attributs stockée dans une base de données
- depuis le logiciel R, connexion cette base de données ;
- chargement de la table d'attributs produite par l'intersection (geoprocessing) des segments avec la vérité terrain (sur le set d'îlots n°2).
- importation de l'objet CART sélectionné
- prédiction sur chaque objet (issu de l'intersection des segments et de la vérité terrain) et export de celle-ci sous forme de table

La table produite par le logiciel R, contenant le résultat de la classification est ensuite jointe avec celle de la base de données qui contient l'information « vérité terrain » pour chaque segment. Une matrice de confusion est construite sur base d'une analyse croisée des deux informations (prédiction et vérité de terrain).

Ces matrices de confusion produites suite aux tests A et B sont illustrées aux tableaux suivants.

Tableau 49. Matrice de confusion construite sur base du test A, précisions globale, du producteur et de l'utilisateur

		CLASSIFICATION								
		Surface (ha)	AGRICOLE	AUTRE	FEUILLUS	MIXTE	RESINEUX	VIDE	Total	précision du producteur
VÉRITÉ TERRAIN	AGRICOLE		672	32	2	1	2	12	721	0.93
	AUTRE		3	9	0	0	0	1	14	0.64
	FEUILLUS		26	14	326	232	35	21	655	0.50
	MIXTE		2	1	38	11	40	4	97	0.12
	RESINEUX		2	2	11	15	518	16	563	0.92
	VIDE		38	12	11	3	7	57	128	0.45
	Total		743	70	388	263	602	112	2 178	
précision de l'utilisateur			0.90	0.13	0.84	0.04	0.86	0.51		0.73

Tableau 50. Matrice de confusion construite sur base du test B, précisions globale, du producteur et de l'utilisateur

		CLASSIFICATION								
		Surface (ha)	AGRICOLE	AUTRE	FEUILLUS	MIXTE	RESINEUX	VIDE	Total	précision du producteur
VÉRITÉ TERRAIN	AGRICOLE		644	51	1	3	2	20	721	0.89
	AUTRE		2	10	0	0	0	1	14	0.72
	FEUILLUS		27	18	469	91	35	14	655	0.72
	MIXTE		2	2	37	14	40	2	97	0.14
	RESINEUX		3	9	7	20	518	5	563	0.92
	VIDE		12	16	9	18	7	66	128	0.51
	Total		690	105	524	147	602	110	2 178	
précision de l'utilisateur			0.93	0.10	0.90	0.10	0.86	0.60		0.79

Les résultats obtenus pour les tests A et B sont très similaires. Le test B, utilisant 3 attributs de plus, donne des résultats sensiblement meilleurs pour la précision « utilisateur » (= % de surface classée en classe x et qui appartient réellement à cette classe x) mais aussi pour celle du producteur (= % de surface d'une classe prédite correctement). En effet, la précision du producteur des classes « Autre » et « Feuillus » sont augmentées.

Ce double test démontre qu'en multipliant les essais (différents CART), des améliorations peuvent être apportées tout en garantissant une certaine stabilité des résultats.

En ce qui concerne le test B, on constate que les grandes classes : « Agricole », « Feuillus » et « Résineux » présentent de bonnes précisions : respectivement 0.93, 0.90 et 0.86 en précision « utilisateur », et 0.89, 0.72 et 0.92 en ce qui concerne la précision « producteur ». La précision globale pour ces trois classes est de 0.90 pour l'utilisateur et de 0.84 pour le producteur. Ce qui signifie que pour ces trois classes, 84% de superficie référencée comme appartenant à l'une de ces classes a été bien classée et que 90 % de ce que l'utilisateur voit apparaître comme appartenant à l'une de ces classes est bien « étiqueté ».

On observe de la confusion entre la classe « Autres » et la classe « Agricole ». La classe « Autre » a été utilisée principalement pour gérer le cas de parcelles à vocation agricole, mais comportant du bâti. Il s'agit donc d'occupations mixtes. Le taux de confusion élevé avec la classe « agriculture » s'explique en partie par le fait que la procédure de segmentation parvient à créer au sein d'objet réputés « autres » (via la photo-interprétation) des segments que la procédure de classification va ranger dans la classe « Agricole », ce qui en soit n'est pas une confusion (cf. figure 105). Cette confusion serait également due à l'hétérogénéité de la classe « Autre » qui contient en plus de bâtiments, des parcelles recouvertes par des surfaces herbeuses et des espaces arborés (cf. figure 106). Cette dernière considération explique également la confusion survenant entre les classes « Autre » et « Feuillus ». La définition même de la classe « Autre » a créé de la confusion et notamment la difficulté d'identifier des cordons arborés en zone agricole.

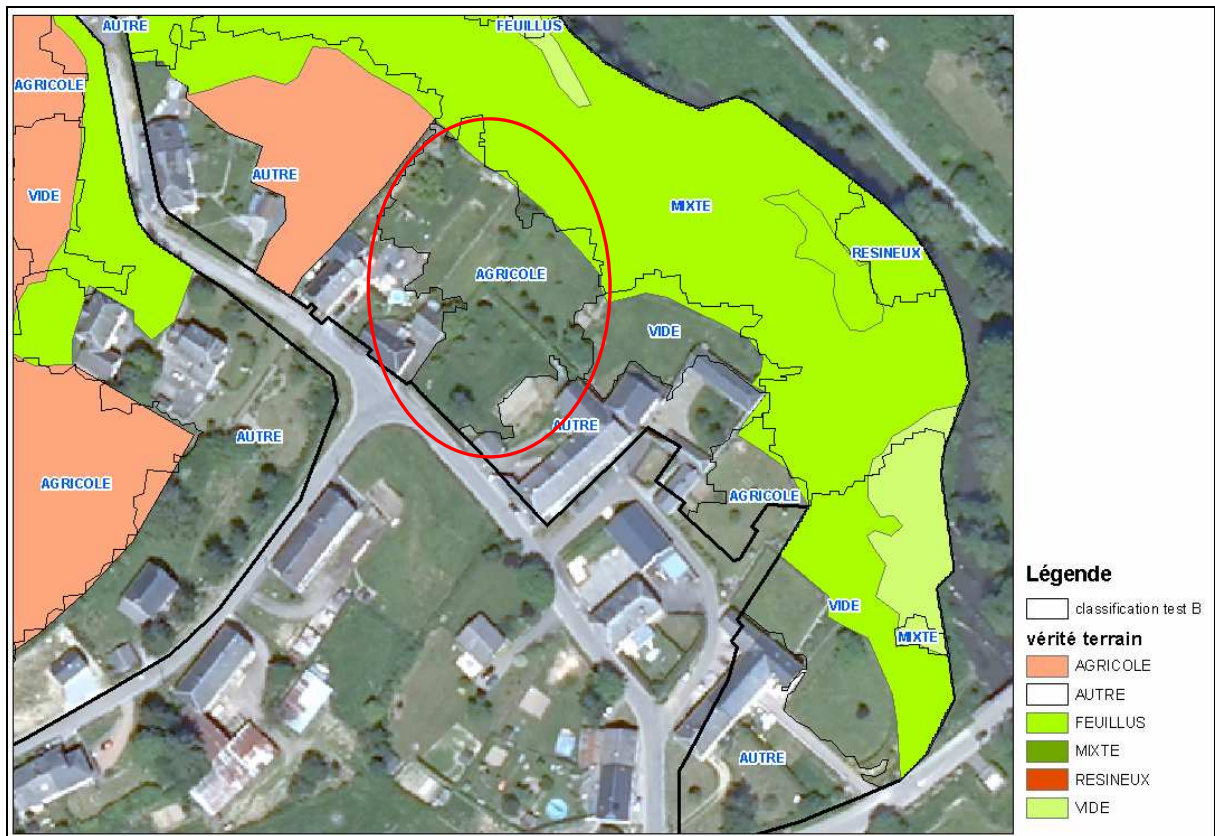


Figure 105. En « transparence » et dans les limites de l’ilot traité, classe « Autre ». En transparence également, bordure noire et étiqueté en bleu, le résultat de classification du test B. Exemple de confusion "Agricole" - "Autre" : segment classé en Agricole alors que la « vérité terrain » le classe en « Autre ».



Figure 106. Vérité de terrain, hétérogénéité de la classe "Autre" affichée en « transparence », limites noires : limites de l’ilot

On observe également une difficulté importante à classer correctement les objets de la classe « Mixte ». Cette difficulté est inhérente à la difficulté même de définir cette classe sur le terrain et la grande ressemblance des forêts mixtes avec les forêts feuillues. En effet, les forêts définies comme mixtes sont souvent des forêts majoritairement feuillues avec quelques îlots d'arbres résineux. Ce n'est donc pas un hasard si c'est avec la classe « Feuillus » que la confusion est la plus grande. On peut même trouver certains cas où le processus de segmentation est parvenu à extraire ces îlots résineux, alors que l'opérateur en charge de la vérité terrain avait défini l'ensemble comme peuplement « mixte ». La figure 107 illustre cette difficulté. Le grand polygone transparent (où on voit l'ortho-image) a été classé en « Mixte » par l'opérateur car il comprend à la fois des zones feuillues et résineuses. Surimposées en jaune et étiquetées, les limites et classes de la couche « prédiction, test B ». On peut observer que le polygone central (limites cyan) comprend en effet une majorité de résineux. On pourrait dire que le classement via le CART est donc aussi représentatif que la vérité-terrain tout en étant répertorié comme faux dans l'analyse. On observe également qu'une partie de ce grand polygone est classée par le CART en mixte mais également en feuillus. La partie étiquetée « Vide » à gauche, est une partie que le logiciel a bien délimitée et classée alors qu'elle a été intégrée à l'occupation « mixte » lors de la phase de digitalisation.

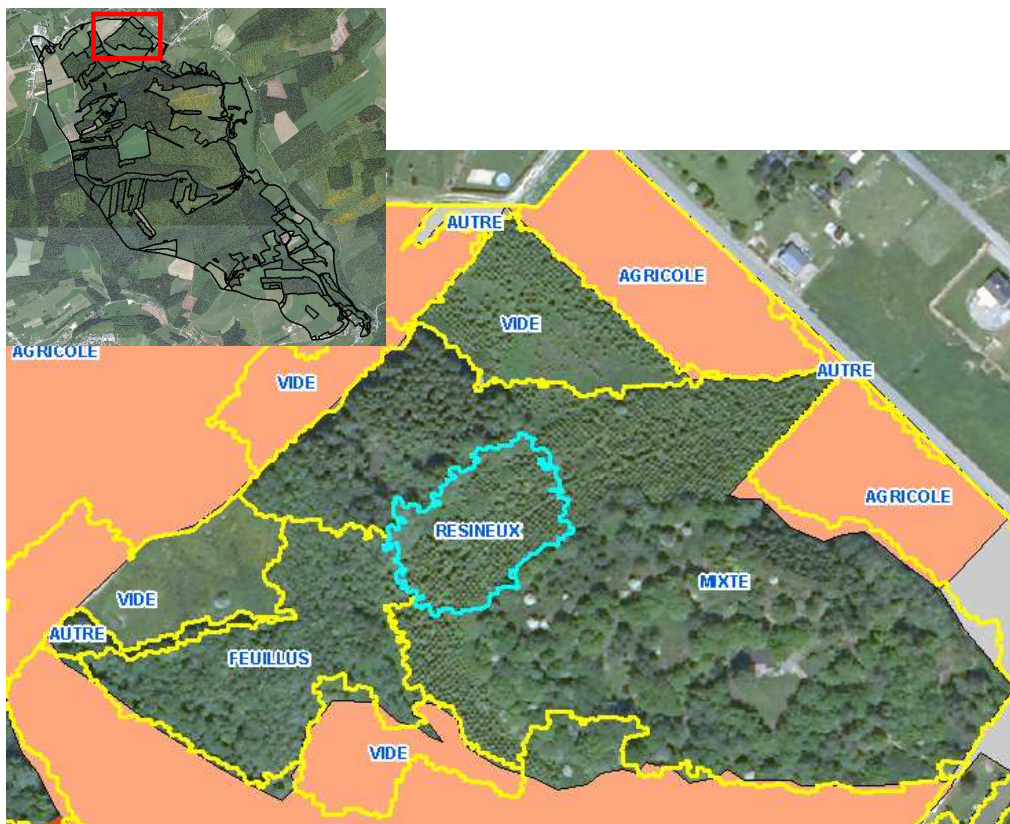


Figure 107. Exemples de confusions "mixte" - "résineux" ou "mixte" - "vide" qui se révèlent être en partie correctes. Polygone classé en « mixte » dans la vérité terrain, limites du classement prédit en jaune, étiqueté.

De même, dans le haut de l'image, un polygone classé en vide est en réalité constitué d'une jeune plantation. Il s'agit d'un autre exemple où le résultat fourni par l'algorithme présente une certaine pertinence, alors qu'il est catalogué comme « faux » dans la matrice de confusion.

Ces dernières considérations illustrent très bien la difficulté existant dans la définition des classes « vérité terrain » (qui d'ailleurs, ne sont pas toujours « vraies » puisque établies par photointerprétation).

La classe « Vide », regroupant les zones de végétation basse et de sol nu en zone forestière, présente également des résultats quelques peu mitigés, inhérents à la très grande variabilité interne à cette classe : coupes à blanc, gagnages, très jeunes plantations,

Le remplacement de la couche « vérité terrain » issue d'une digitalisation, par une photo-interprétation des segments (dans le cas de la construction des arbres de classification), voire d'une photo-interprétation par point devrait permettre de lever certaines ambiguïtés dans l'évaluation de la précision associée aux classes hétérogènes telles que « Autre » et « Forêt mixte ».

Une autre piste qui mérite d'être approfondie est la réalisation de traitements de post-classification. Ceux-ci se baseraient sur les relations topologiques entre objets et permettraient d'éliminer certaines confusions sur base de règles de voisinage.

Influence des attributs SPOT

Afin d'évaluer l'importance des images SPOT dans la performance du processus de classification, des tests supplémentaires, baptisés respectivement C, D et E, ont été réalisés selon le même schéma que pour les tests A et B. Dans ces tests, les attributs issus des 2 images SPOT (C), de l'image SPOT de juillet 2006 (D) ou de l'image SPOT de novembre 2005 (E) ont été retirés du jeu de données.

Les tables reprenant la classification des attributs selon la MDG sont présentées en annexe 2.

Sur base de ces classements et de la même manière que précédemment, des CART ont successivement été construits en injectant de plus en plus d'attributs. L'évolution de la précision globale obtenue en fonction du nombre d'attributs injectés dans le CART est illustrée aux figures suivantes (Figure 108 pour les CART sans aucun attribut SPOT, figure 109, pour les CART sans les attributs dérivés de la SPOT de juillet 2006 et figure 110, sans les attributs dérivés de la SPOT d'octobre 2005).

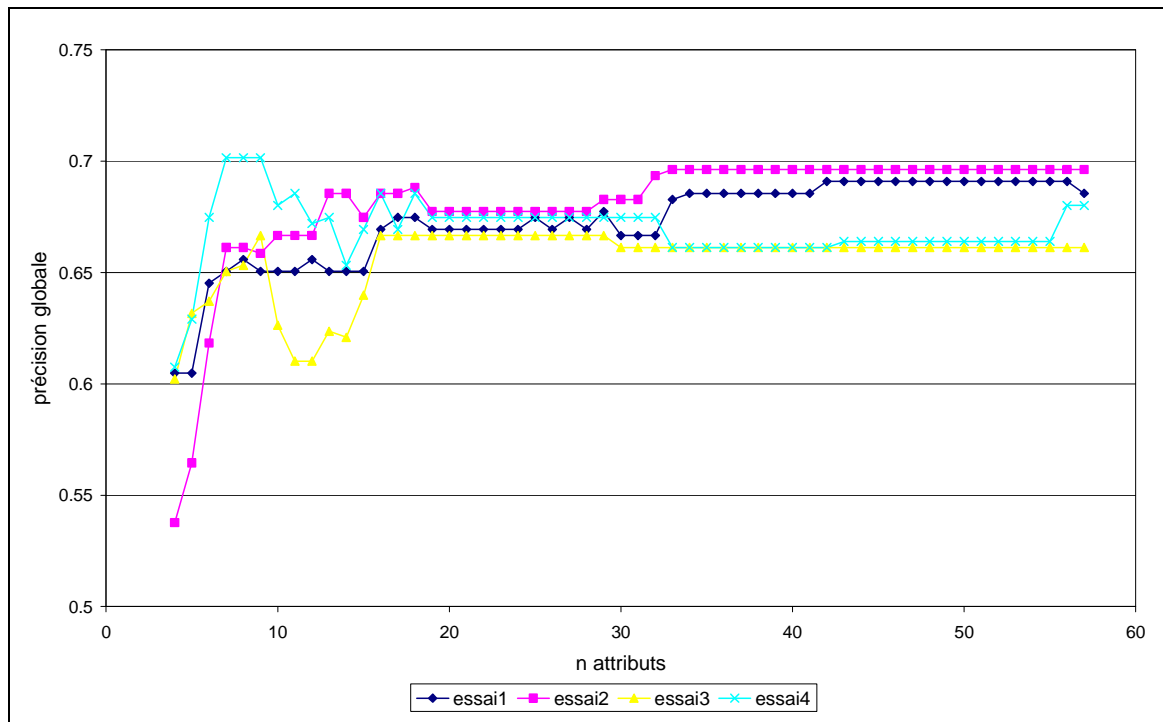


Figure 108. Evolution de la précision globale (overall accuracy) en fonction du nombre d'attributs insérés dans le CART (forêt aléatoire sans les attributs SPOT), différents essais

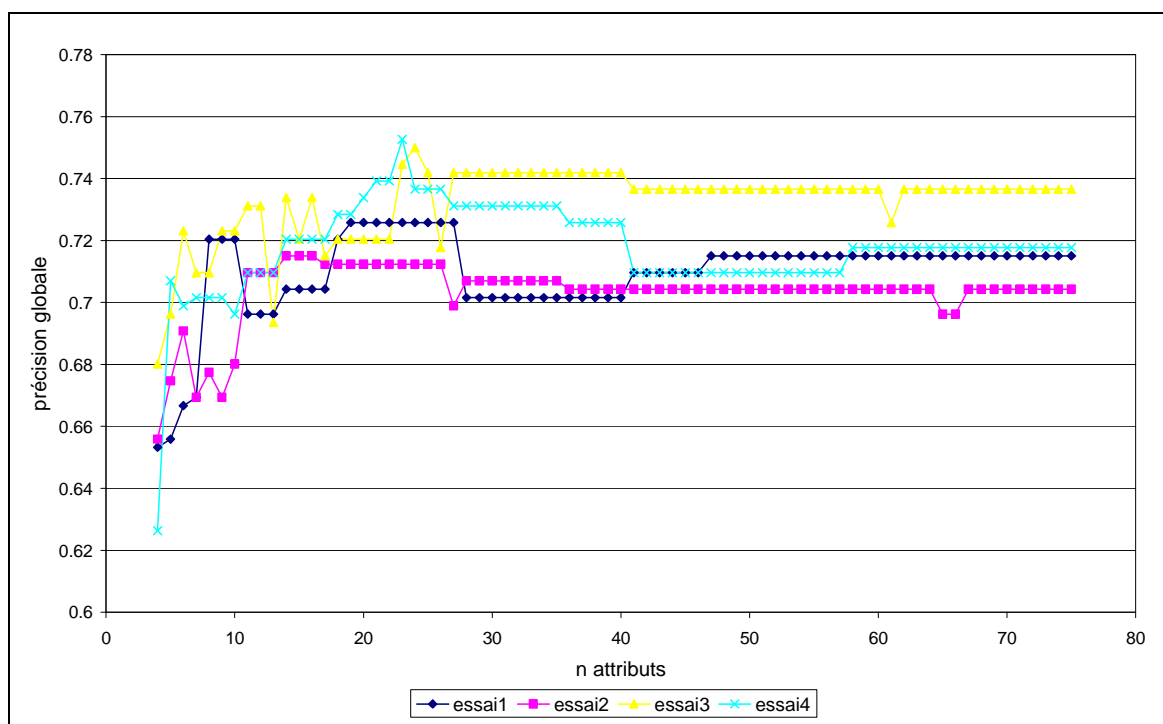


Figure 109. Evolution de la précision globale (overall accuracy) en fonction du nombre d'attributs insérés dans le CART (forêt aléatoire sans les attributs SPOT 07/2006), différents essais

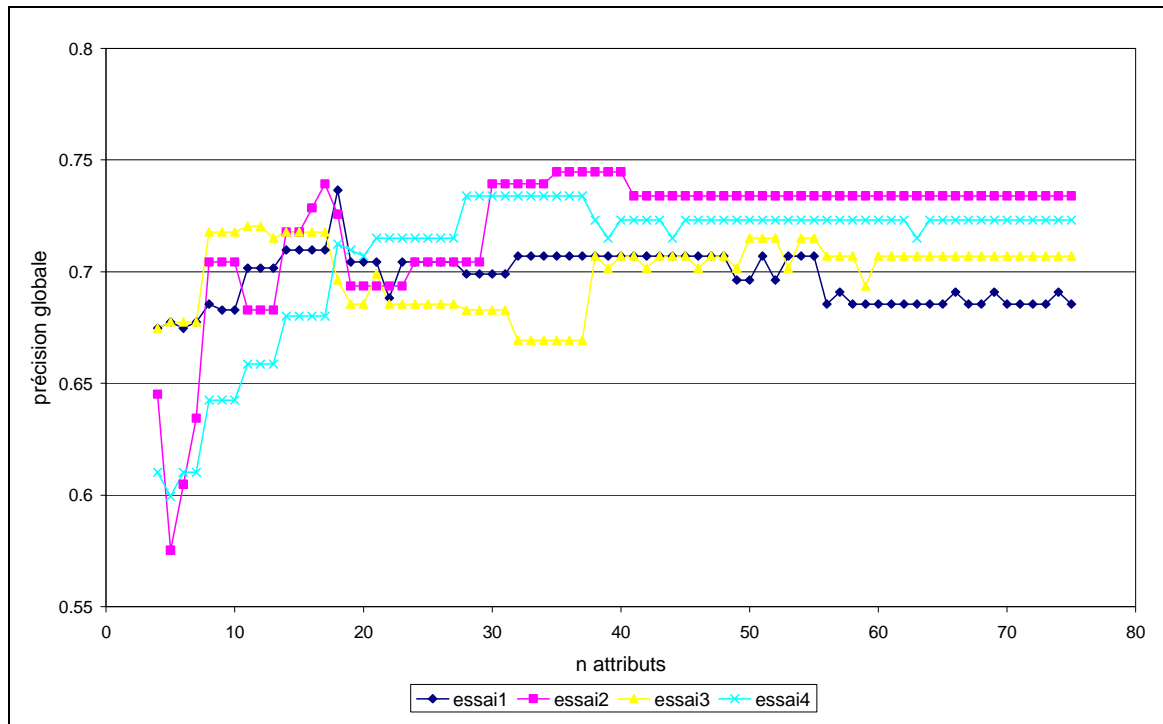


Figure 110. Evolution de la précision globale (overall accuracy) en fonction du nombre d'attributs insérés dans le CART (forêt aléatoire sans les attributs SPOT octobre 2005), différents essais

On observe une diminution de qualité lorsque tous les attributs SPOT sont retirés du jeu de données (précision globale majoritairement située entre 0.6 et 0.7 contre 0.66 à 0.76 pour les CART où tous les attributs peuvent intervenir cf. figure 102). Les tests D et E, où une seule image SPOT est retirée du jeu de données, semblent donner de précision similaires à celles obtenues pour les tests A et B issus des CART associés à la figure 102.

Les arbres de décision associés aux tests C, D et E n°2 sont présentés aux figures 111 à 113.

Les attributs intervenant dans ces trois CART sont renseignés au tableau 51.

Tableau 51. Attributs utilisés dans les CART des tests, de gauche à droite, C, D et E

Attributs - Test C	Attributs - Test D	Attributs - Test E
Hest_M	S10b3_M	S7b3_M
Ohp_Sd	Odbr_Sd	Ondvi_M
Ogdvi_M	S10b4_M	Ohp_Sd
Odbr_Sd	Hest_M	Ogdvi_Sd
Odgrvi_M	Odbr_M	S7b2_M
Ogdvi_M	Ohp_Sd	S7b4_M
Odbr_M	S10b2_M	Ogdvi_M
Ogrvi_M	Odgndvi_M	GLCM_Contr
Area_Pxl	Odb1_M	Odbr_Sd
GLCM_StdDe_		Odb1_Sd
Odb1_Sd		

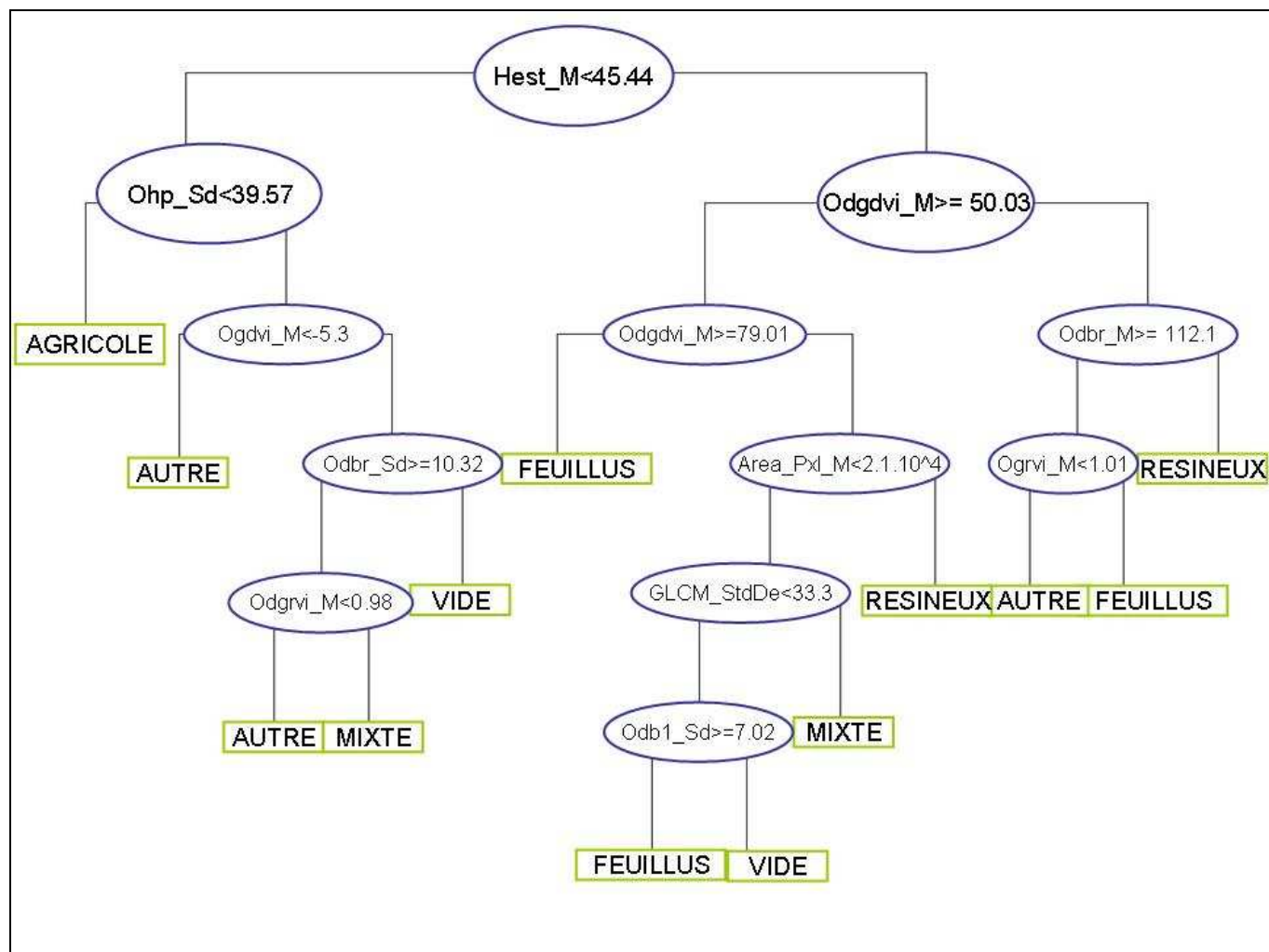


Figure 111. Architecture de l'arbre CART utilisé dans le test C

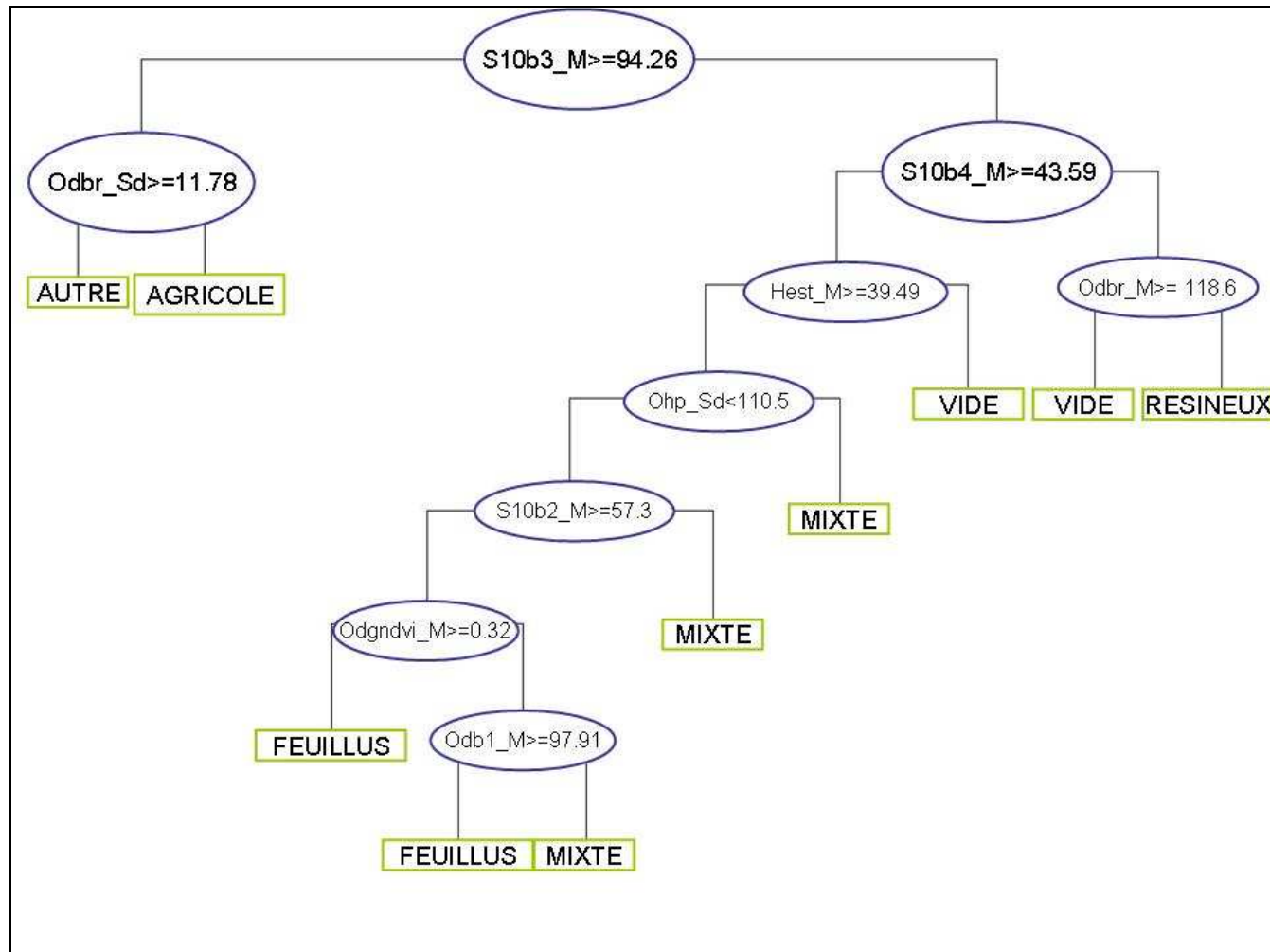


Figure 112. Architecture de l'arbre CART utilisé dans le test D

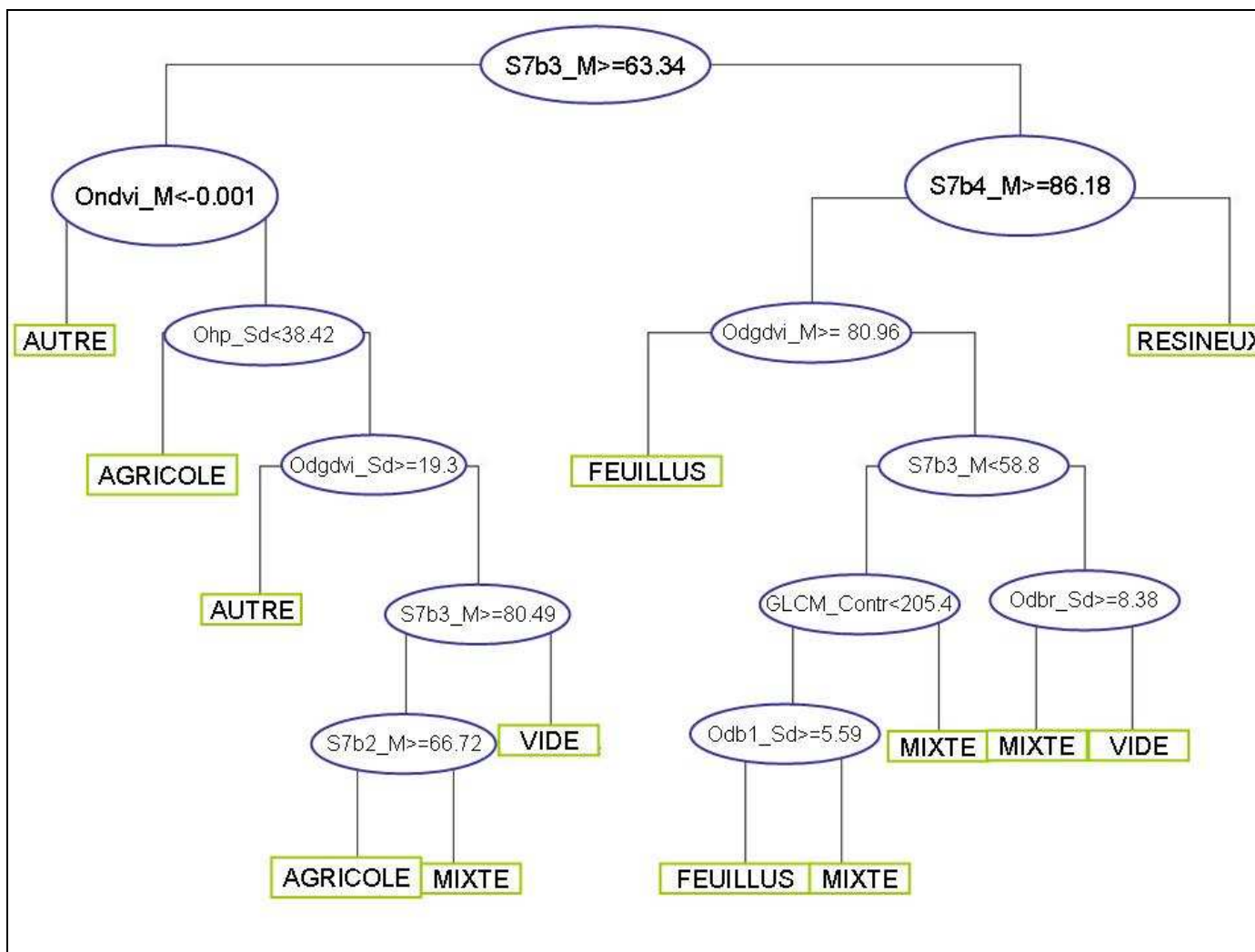


Figure 113. Architecture de l'arbre CART utilisé dans le test E

A la figure 111, correspondant au test C, on observe que l'arborescence fournit de nouveau une première grande séparation entre les classes arborées à droite et les classes non arborées (« Agricole », « Autre » et « Vide ») à gauche. Cette dichotomie est basée sur l'attribut Hest_M qui remplace l'attribut basé sur la bande 3 de Spot (juillet 2006) dans les arbres précédents. Les autres attributs intervenant sont basés sur les bandes *bright* et *high pass* ainsi que sur des indices de végétation associés à la couche ortho_DGA. Un indice de « taille » fait son apparition (Area_Pxl), ainsi qu'un indice de texture sous la forme de GLCM_StdDe.

Dans le cas du test D (retrait des attributs de l'image SPOT de juillet 2006), on constate que c'est l'attribut S10b3_M (moyenne de la bande 3 de l'image SPOT d'octobre 2005) qui occupe le premier nœud de l'arbre de décision. De même, lorsque les attributs de l'image SPOT d'octobre 2005 sont enlevés, c'est l'attribut S7b3_M (moyenne de la bande 3 de l'image SPOT de juillet 2006) qui sert de première distinction entre éléments arborés et les autres. Le nombre d'attributs utilisés reste similaire aux deux premiers tests : 11, 9 et 10 respectivement pour les tests C, D et E.

Les matrices de confusion qui découlent de ces tests sont illustrées aux tableaux suivants.

Tableau 52. Matrice de confusion construite sur base du test C, précisions globale, du producteur et de l'utilisateur

		CLASSIFICATION						Total	précision du producteur
		Surface (ha)	AGRICOLE	AUTRE	FEUILLUS	MIXTE	RESINEUX		
V E R I F I C A T I O N	AGRICOLE	619	20	13	10	5	54	721	0.86
	AUTRE	1	9	0	2	0	2	14	0.61
	FEUILLUS	19	5	527	40	34	30	655	0.80
	MIXTE	1	1	42	11	37	5	97	0.12
	RESINEUX	0	11	34	40	448	30	563	0.80
	VIDE	16	22	8	7	8	68	128	0.53
	Total	657	66	624	111	531	189	2 178	
	précision de l'utilisateur	0.94	0.13	0.85	0.10	0.84	0.36		0.77

Tableau 53. Matrice de confusion construite sur base du test D, précisions globale, du producteur et de l'utilisateur

		CLASSIFICATION						Total	précision du producteur
		Surface (ha)	AGRICOLE	AUTRE	FEUILLUS	MIXTE	RESINEUX		
V E R R I E T E	AGRICOLE	673	13	7	4	3	21	721	0.93
	AUTRE	4	8	0	0	0	1	14	0.56
	FEUILLUS	56	13	457	80	22	27	655	0.70
	MIXTE	3	0	38	21	30	5	97	0.21
	RESINEUX	3	1	21	46	473	19	563	0.84
	VIDE	46	2	10	5	8	58	128	0.45
	Total	784	38	533	156	536	131	2 178	
précision de l'utilisateur		0.86	0.21	0.86	0.13	0.88	0.44		0.78

Tableau 54. Matrice de confusion construite sur base du test E, précisions globale, du producteur et de l'utilisateur

		CLASSIFICATION						Total	précision du producteur
		Surface (ha)	AGRICOLE	AUTRE	FEUILLUS	MIXTE	RESINEUX		
V E R R I E T E	AGRICOLE	649	16	1	3	1	51	721	0.90
	AUTRE	3	8	0	0	0	3	14	0.57
	FEUILLUS	19	5	466	88	21	56	655	0.71
	MIXTE	2	1	38	14	31	11	97	0.15
	RESINEUX	0	1	16	31	484	31	563	0.86
	VIDE	17	7	7	8	5	86	128	0.67
	Total	689	38	527	144	542	238	2 178	
précision de l'utilisateur		0.94	0.22	0.88	0.10	0.89	0.36		0.78

En termes de précision globale, on obtient des valeurs très légèrement inférieures aux résultats du test B (prenant en compte des attributs de toutes les sources de données) : 0.77 pour le test C, 0.78 pour les tests D et E contre 0.79 pour le test B.

En termes de précisions globales liées aux trois classes majeures (« Agricole », « Feuillus » et « Résineux ») on obtient les résultats présentés au tableau 55.

Tableau 55. Précisions globales de l'utilisateur et du producteur pour les classes (« Agricole », « Feuillus » et « Résineux »), comparaison des tests B à D

	précision de l'utilisateur	précision du producteur
Test B	0.90	0.84
Test C	0.88	0.82
Test D	0.86	0.83
Test E	0.91	0.82

Le test B donne des résultats très légèrement supérieurs en ce qui concerne la précision du producteur et est presque en tête en ce qui concerne la précision de l'utilisateur.

Suite à cette analyse, on peut noter que les résultats obtenus en enlevant tous les attributs SPOT sont très légèrement inférieurs au test qui utilise les 2 images SPOT (test B). Les tests D et E qui enlèvent les attributs d'une image SPOT n'ont pas montré de diminution des performances dans la configuration des classes telle qu'elle est définie actuellement. En termes de précision de l'utilisateur, on note cependant une confusion beaucoup plus importante entre les occupations « Agricole », « Feuillus » et « Vide » dans le cas du test D (suppression de l'image SPOT de juillet 2006).

On peut également noter le remplacement, dans le nœud principal de séparation entre les classes arborées et les autres, de l'attribut lié à la bande 3 de l'une ou l'autre SPOT par la couche de hauteur Hest.

c) Envi Ex

De la même manière que pour les segments issus de Definiens, une forêt aléatoire a été générée sur base des segments dont l'occupation du sol était à plus de 85% pure et dont la superficie dépassait 5m² (notons qu'Envi a tendance à créer de très petits segments dans certaines configurations de l'image). Toutefois, il a été impossible de générer cette forêt sur l'entièreté des objets issus de l'intersection entre la couche segments et la couche « vérité terrain » (cf. tableau 56). En effet, la capacité de la machine de travail utilisée¹ a été dépassée. Pour pouvoir arriver à un résultat, il a fallu diviser le jeu de données par 10, tout en ayant un temps de travail considérablement supérieur au traitement des données Definiens sur 100% de la même zone.

Tableau 56. Nombres de segments d'apprentissage par classe (85% de pureté d'occupation du sol, superficie>5m²)

Occupation	n de segments
Agricole	9 970
Autre	2 499
Feuillus	35 558
Mixte	5 960
Résineux	58 284
Vide	3 574
Total	115 845

Les mêmes étapes que décrites précédemment pour Definiens ont été conduites sur base du tableau de classement des attributs obtenus via la forêt aléatoire. Un test (F) a été conduit sur base d'un CART dont la précision globale était la meilleure.

¹ Machine utilisée : Intel Dual Core 2.13 GHz, 3 Gb de RAM

Toutefois, la production de la matrice de confusion a mis en lumière la disparition de certaines classes ayant été trop désavantagées par le très grand déséquilibre du nombre de segment produit par classe. Celle-ci est présentée au tableau 57.

Tableau 57. Matrice de confusion construite sur base du test F

		CLASSIFICATION				
		Surface (ha)	AGRICOLE	FEUILLUS	RESINEUX	Total
V E R I T E	T E R R I T O I R E	AGRICOLE	602	93	26	721
		AUTRE	8	2	3	14
		FEUILLUS	31	574	50	655
		MIXTE	5	44	47	97
		RESINEUX	3	79	481	563
		VIDE	17	56	55	128
Total		666	849	663	2 178	

Pour remédier à ce phénomène, une forêt aléatoire a été générée sur un nombre constant de segments par classe (2488, classe limitante « Autre »). Celui-ci a permis de caractériser les différents attributs par la MDG (cf. annexe 3).

Une succession de CART a été produite de la même manière que précédemment. L'évolution des précisions globales est illustrée en figure 114.

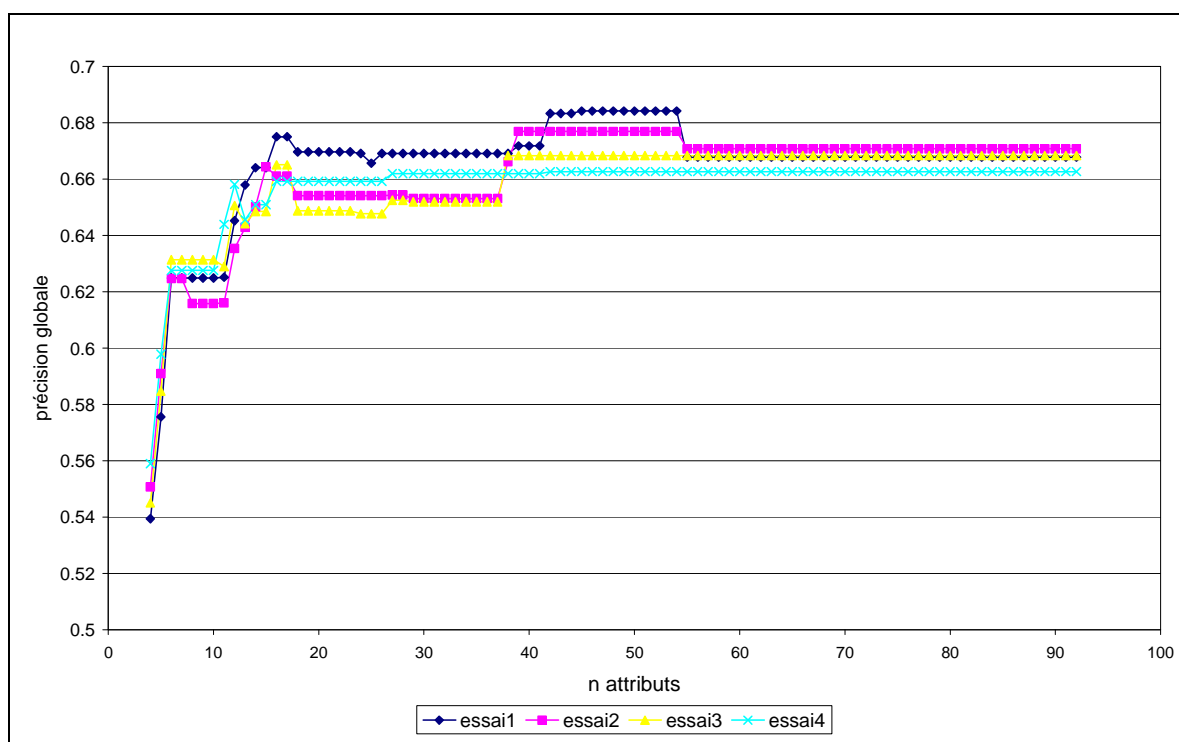


Figure 114. Evolution de la précision globale (overall accuracy) en fonction du nombre d'attributs insérés dans le CART, différents essais

Le test (G) de classification sur les îlots du set n°2 a été opéré avec un CART parmi ceux associés aux courbes en figure 114. En effet, celui-ci semble donner des résultats acceptables en termes de nombre de segments classés.

Après application de cet arbre au set n°2, la matrice de confusion est extraite (cf. tableau 58).

Tableau 58. Matrice de confusion construite sur base du test G, précisions du producteur et du consommateur

		CLASSIFICATION						Total	précision du producteur
		Surface (ha)	AGRICOLE	AUTRE	FEUILLUS	MIXTE	RESINEUX		
V E R R I F I C A T I O N	AGRICOLE	342	291	0.000		16	72	721	0.47
	AUTRE	2	9	0.000		1	3	14	0.65
	FEUILLUS	19	16	0.011	0	25	594	655	0.00
	MIXTE	1	9	0.000	0	29	58	97	0.00
	RESINEUX	1	7	0.001	2	400	152	563	0.71
	VIDE	7	40		1	19	60	128	0.47
	Total	371	373	0.012	3	491	939	2 178	
précision du consommateur		0.92	0.02	0.89	0.00	0.82	0.06		0.37

On observe une précision globale assez mauvaise : 0.37. Cette grande disparité entre la précision globale lors de l'établissement du CART (cf. figure 114) et la précision globale d'application au set de données n°2, non rencontrée lors du traitement des segments Definiens, est due aux déséquilibres de taille (et donc de nombre) de segments entre classes. En effet, le nombre important de petits segments bien classés tire la précision globale vers le haut en termes de nombre de segments bien classés mais ne le fait pas en termes de superficie bien classée.

Pour le reste, on observe des valeurs similaires de précision de l'utilisateur pour les grandes classes : « Agricole », « Feuillus » et « Résineux » avec respectivement des valeurs de 0.92, 0.89 et 0.82 tandis que les précisions associées aux trois autres classes sont quasiment nulles. Toutefois, ces valeurs doivent absolument être mises en parallèle avec les précisions du producteur. En effet, ces dernières annihilent l'apparente bonne précision de l'utilisateur dans le cas de la classe «Feuillus». En effet, presque aucun segment n'est classé comme tel (ils sont majoritairement classés en Vide). De même aucun segment n'a été classé en 'Mixte'. Pour la classe «Agricole», seule 47% de la superficie référencée comme telle a été prédite comme appartenant à cette classe. Pour les résineux, c'est un peu moins mauvais (72% de précision du producteur) tout en gardant des valeurs inférieures à celles obtenues pour Definiens.

Pour les trois grandes classes («Agricole», «Feuillus» et «Résineux»), la précision globale de l'utilisateur est en moyenne bonne (0.86) ; tandis que la précision du producteur équivalente est médiocre (0.38).

Toutes ces considérations nous mènent à la conclusion que l'utilisation d'Envi dans un tel schéma de production d'information thématique sur l'occupation du sol est de loin moins optimale que celle de Definiens. Toutefois, une analyse plus approfondie permettant de cerner les raisons de telles disparités n'a pas pu être menée. L'extraction d'information concernant de segments de très petite taille a certainement dû contribuer à la confusion générale observée dans les résultats. On peut néanmoins affirmer que le processus de segmentation d'Envi a engendré une lourdeur notable dans ce processus d'étude exploratoire

2.7.4.3. Conclusions

Dans un premier temps, l'analyse statistique des attributs utilisant la technique d'analyse en composantes principales a permis de mieux comprendre ceux-ci, leur importance et la redondance qui les lie. Toutefois, à la suite de cette analyse, aucun résultat probant en termes de classification n'a été obtenu, étant donné le nombre important d'individus et la signification ambiguë des classes de la vérité terrain. L'utilisation de l'analyse discriminante ne s'est pas avérée pertinente dans la mesure où les classes ciblées comportent une grande variabilité interne.

Une analyse intégrée des attributs a été menée à l'aide de méthodes statistiques permettant l'extraction des attributs les plus discriminants dans le cadre de l'établissement d'un schéma de classification. Dans un premier temps, la technique des forêts aléatoires a permis d'établir un classement des attributs selon leur aptitude à créer des partitions optimales au sein d'un ensemble d'arbres de décision (forêt). Ensuite, sur base de cette liste d'attributs, des arbres de décisions ont été produits via la méthode CART. Une sélection d'entre eux a permis de tester l'apport des différentes sources de données dans la classification de l'occupation du sol. Les précisions globales liées aux trois classes majeures que sont les classes « Agricole », « Feuillus » et « Résineux » pour les différents tests réalisés sont renseignées au tableau 59. Toutefois, il faut noter que les précisions liées aux classes « secondaires » (« Autres », « Mixte » et « Vide ») sont moindres. Ceci a notamment été expliqué par la difficulté inhérente à la définition même des classes ainsi que la constitution de la couche vérité de terrain ayant pour conséquence directe l'introduction de confusions dans le processus d'évaluation du processus de classification.

Tableau 59. Précisions globales de l'utilisateur et du producteur pour les classes («Agricole», «Feuillus» et «Résineux»), comparaison des différents tests menés (dans l'ordre, B, C, D et E)

Logiciel	sources de données	précision de l'utilisateur	précision du producteur
D E F I N I E N S	toutes	0.90	0.84
	toutes sauf SPOT	0.88	0.82
	juillet 2006 et SPOT octobre 2005		
	toutes sauf SPOT juillet 2006	0.86	0.83
	toutes sauf SPOT octobre 2005	0.91	0.82
ENVI	toutes	0.86	0.38

Au tableau 59, une première distinction peut être faite entre les résultats obtenus par la chaîne de traitement impliquant d'une part Definiens 8.1, d'autre part Envi Ex. En effet, la précision du producteur obtenue via l'utilisation du logiciel Envi Ex est bien moindre (0.38) que celle obtenue dans les différents tests impliquant Definiens.

Malgré le fait que l'analyse des causes de ces disparités n'a pas pu être approfondie dans le temps imparti, on peut affirmer sur base de l'expérience acquise que le processus de segmentation d'Envi engendre une lourdeur notable dans ce processus d'étude exploratoire, lourdeur qui serait démultipliée dans l'optique d'un procédé de production à plus grande échelle.

Les meilleures performances de classification au regard à la fois du producteur et de l'utilisateur ont été obtenues par l'arbre utilisant des attributs provenant de toutes les sources de données (cf. test B, arbre en figure 104). La suppression des attributs provenant des images SPOT a légèrement diminué la précision globale, tout en maintenant une bonne précision.

Si l'on souhaite ne garder qu'une seule image SPOT, on peut considérer, sur base des précisions obtenues pour les classes dominantes (Tableau 59), qu'il est préférable de conserver l'image correspondant à la période estivale. Il faut cependant noter que l'image automnale qui a été utilisée est celle de l'année « t-1 » par rapport à l'année de référence de la couche ortho.

En cas de suppression des 2 images SPOT, le nœud principal de l'arbre de classification qui opère la séparation entre les classes arborées et les classes non arborées est occupé par l'attribut HEST provenant de la couche de hauteur produite en tâche 2.

2.8. Intégration des données produites dans la base de données géographique de la Région wallonne (Unité de Géomatique - GRFMN)

Etant donné le caractère exploratoire des travaux menés dans ce projet, la grande majorité des couches d'informations produites ne présentent pas de réel intérêt pour l'utilisateur final. Ces couches ont en effet des couvertures spatiales souvent limitées et ont un contenu à finalité technique très spécifique.

La seule couche qui semble pouvoir être utilement mise à disposition des utilisateurs finaux est la couche ORTHO_DGA dégradée à 2,5 m de résolution. Pour cette dernière, les modalités de transfert et d'intégration dans la base de données existante restent à préciser.

Néanmoins, les principales couches de données produites dans le cadre de cette recherche ont été archivées et documentées et seront remises au Département du Développement Rural de la DGO3. De la même manière, la base de données des objets de référence constitue un précieux outil pour des études ultérieures. Cette base de données sera également remise au bailleur de fonds.

2.9. Recommandations pour les réalisations des futures couches « image satellitale HR » de la Wallonie (Unité de Géomatique - GRFMN)

Les résultats obtenus dans le paragraphe 2.7 ont démontré l'intérêt de disposer pour la phase de classification, de couches images à haute résolution en complément de la couche ortho. Ces images présentent en effet une plus grande cohérence radiométrique que les ortho-images.

Les tests réalisés dans le paragraphe 2.2 sur les niveaux de prétraitement nous orientent vers le choix d'images SPOT orthorectifiées de niveau 3.

Les résultats obtenus en n'utilisant qu'une seule image (été ou automne) montrent un léger avantage à l'utilisation de l'image « été ». La littérature semble confirmer cette préférence, mais il convient de rappeler que l'image « automne » qui a été utilisée dans l'étude était celle de l'année « t-1 » par rapport à l'année de référence de la couche ortho.

Dans le contexte spécifique de la production/mise à jour de la carte d'occupation du sol, la création d'une couche « continue », couvrant sur une période de temps limitée, l'ensemble du territoire de la Wallonie ne semble pas devoir être recommandée. En effet, l'acquisition et le renouvellement de la couche ortho au niveau régional sont actuellement organisés sur base d'un découpage de la région en plusieurs blocs distincts programmés sur 2 années successives.

En outre, les contraintes techniques (notamment de corrections radiométriques) liées à l'utilisation simultanée de plusieurs images dans une démarche de classification conduisent à penser qu'il sera plus simple d'exploiter les images SPOT sur une base individuelle, en fonction de leur disponibilité et de la combinaison de leur découpage avec le découpage des blocs de la couche ortho de la DGA.

Ce double découpage donnerait lieu à un ensemble de sous-blocs susceptibles de donner lieu à la construction d'un arbre de classification spécifique. Il conviendrait pour cela de disposer au sein de chaque sous-bloc d'un nombre d'objets de référence (« vérité terrain ») suffisamment important que pour construire un arbre de classification de qualité suffisante.

Cette approche aurait en outre le mérite de prendre en compte les variations sous-régionales en matière de phénologie de la végétation (importante à considérer lors de l'utilisation d'ortho-image printanières), ou de dominance de certains types d'occupations (cultures vs prairies, résineux vs feuillus).

2.10. Journée d'étude sur l'occupation du sol en Wallonie

Une journée d'étude sur l' « Occupation du Sol en Wallonie » a été organisée le 20 octobre 2010 à la Gembloux Agro-Bio Tech. Le programme de cette journée d'étude est annexé au présent rapport (Annexe 4).

3. Conclusions générales

Les lignes qui suivent synthétisent l'ensemble des conclusions intermédiaires des différentes parties de cette étude. Celles-ci sont complétées d'un ensemble de recommandations et de perspectives.

- Malgré la difficulté d'analyser la qualité radiométrique des ortho-images due à l'influence de nombreux facteurs sur la variation spatiale de la radiométrie, des effets du mosaïquage ont été identifiés sur les ortho-images de 2006. Par contre, sur le bloc d'ortho-images de 2007, l'effet « mosaïquage » est très peu visible. L'importance dans ce genre d'étude de connaître la disposition des seamlines a été notifiée. La fourniture des seamlines a d'ailleurs été intégrée au cahier des charges de production de la couverture ORTHO de 2009/2010.
- La variabilité radiométrique de la couche ORTHO_DGA identifiée dans une étude préalable ainsi que les effets du mosaïquage identifiés lors de la présente étude incitent au recours à des données satellitales haute résolution, plus « stables » radiométriquement parlant. L'exploitation des valeurs radiométriques est donc à proscrire pour la réalisation de la classification des zones forestières.
- L'analyse d'images SPOT de niveaux de prétraitement différents a permis d'identifier l'image SPOT orthorectifiée de niveau 3 comme étant celle qui convient le mieux pour être exploitée en combinaison avec la couche ORTHOS_DGA.
- La période de prise de vue recommandée pour les images haute résolution est l'été. Si une seconde image peut être ajoutée au jeu de données, il serait préférable de choisir une prise de vue automnale de la même année que l'image printanière pour éviter l'interférence de modification de limites de parcelles toujours possible en zone agricole entre 2 saisons de végétation.
- Considérant le niveau de détail de classification de l'occupation du sol pris en compte dans cette étude, le gain de précision apporté par l'utilisation de données diachronique, bien que significatif ne s'est pas avéré très important. Il est fort probable qu'il en serait autrement, notamment si la distinction entre occupations agricoles (cultures et prairies) avait été prise en compte.

- La qualité des MNH produit par différence du MNS et d'un MNT est surtout influencée par la qualité de ce dernier qui s'est avérée très variable, en fonction des données de base utilisées (cf. paragraphe 2.3.1.3). S'il n'est pas envisageable d'utiliser ce MNH dans le processus de segmentation, il peut cependant intervenir dans l'étape de classification des segments. Cette dernière utilisation doit cependant rester prudente sachant que ce MNH contient certaines incohérences. L'analyse de arbres de classification réalisés en supprimant les données SPOT à permis de confirmer l'intérêt potentiel de ce type de données.
- La base de données d'objets de référence comportant à la fois des îlots dont l'occupation du sol a été digitalisée et des objets spectraux constitue une source d'informations intéressante qui pourra encore être exploitée et valorisée par la suite, surtout pour les occupations forestières qui présentent un taux de changement au cours du temps moins important que pour les occupations agricoles.
- L'approche de découpage du processus de traitement des images par îlot a été privilégiée dès le début des travaux. Ce choix s'inscrivait dans une logique d'intégration opérationnelle à la démarche de mise à jour de la carte d'occupation du sol. Le temps imparti à l'étude n'a pas permis de mener une analyse comparative de cette approche avec un autre type de découpage du travail. Le maintien de l'approche par îlot ne remet cependant pas en cause la pertinence de la démarche de segmentation-classification utilisée.
- La comparaison des 2 logiciels de traitement d'images a porté essentiellement sur leur capacité à segmenter de manière efficiente les ortho-images qui sont les seules sources de données utilisées dans cette phase du processus. Les différences de comportement, et principalement la tendance de ENVI à sursegmenter les objets en zones forestières ont cependant des répercussions y compris dans la mise en œuvre de la phase de classification, puisque le très grand nombre d'objets générés rend plus difficile la procédure de construction des arbres de classification. L'expérience acquise au cours de ces différents essais plaide donc en faveur de l'utilisation de Définiens. La seule réserve vis-à-vis de ce logiciel réside dans le coût très élevé de la version « serveur » qui s'avère nécessaire dans la mise en œuvre de la procédure segmentation/classification à une échelle opérationnelle (recours à un langage de programmation pour automatiser les opérations).

- Le choix des paramètres de segmentation retenu dans le cas du logiciel Definiens (*scale parameter* :125, *shape* 0.3 et *compactness* : 1.0) s'est avéré pertinent en regard du niveau de détail et des éléments de la légende pris en considération (Zones « Agricole », « Autre (essentiellement bâti en zone agricole et annexes, « Feuillus », « Résineux », « Mixte » et « Vide » (composée de sous-classes : mises à blanc, vide et gagnage)).
- En ce qui concerne ces derniers, plusieurs remarques méritent d'être avancées :
 - o La différenciation des classes « Agricole », « Résineux » et « Feuillus » s'opère avec des niveaux de précision « utilisateur » et « producteur » globalement très élevés (cf. tableau 60)

Tableau 60. Précisions de l'utilisateur et du producteur pour les classes («Agricole», «Feuillus» et «Résineux»), comparaison des différents tests menés

Logiciel	sources de données	précision de l'utilisateur			précision du producteur		
		Agricole	Feuillus	Résineux	Agricole	Feuillus	Résineux
D E F I N I E N S	toutes	0.93	0.90	0.86	0.89	0.72	0.92
	toutes sauf SPOT juillet 2006 et SPOT octobre 2005	0.94	0.85	0.84	0.86	0.80	0.80
	toutes sauf SPOT juillet 2006	0.86	0.86	0.88	0.93	0.70	0.84
	toutes sauf SPOT octobre 2005	0.94	0.88	0.89	0.90	0.71	0.86

- o Les classes « Autre », « Mixte » et « Vides » s'avèrent plus problématiques.
- o La classe « Autre » a été utilisée principalement pour gérer le cas de parcelles à vocation agricole, mais comportant du bâti ainsi que de petits îlots ou cordons arborés disséminés dans une pairie annexe au bâtiment. Il s'agit donc d'occupations du sol mixtes. C'est cette grande variabilité qui crée de la confusion entre ces classes et qui pousse le logiciel à classer des morceaux de prairies situés à proximité du bâti en classe « Agricole » (ce qui n'est pas vraiment faux) et inversement des parcelles agricoles en « Autre ». De même, les cordons arborés sont occasionnellement classés en « Autre » ou en « Feuillus ».
- o Le même phénomène est observé pour la classe « Mixte » (forêts mixtes) qui se présente généralement sous la forme d'une matrice feuillue comportant des îlots résineux. La combinaison du processus de segmentation/classification parvient à certains moments à discriminer au sein d'un objet considéré par le photo-interprète comme une forêt mixte, des éléments relevant des classes « Feuillus » et « Résineux ». La confusion entre forêts mixtes d'une part et forêts feuillues ou résineuses d'autre part doit donc être relativisée.

- L'utilisation combinée des méthodes « forêts aléatoire » et « CART », respectivement pour sélectionner les attributs et pour réaliser la classification semblent donner des résultats concluants. Un des principaux attraits de cette approche est la capacité à gérer efficacement les classes présentant une variabilité interne importante (cas de la classe « agriculture » par exemple).
- Les précisions globales liées aux trois classes majeures que sont les classes «Agricole», «Feuillus» et «Résineux» pour les différents tests réalisés sont renseignées ci-dessous.

Rappel du Tableau 59. Précisions globales de l'utilisateur et du producteur pour les classes («Agricole», «Feuillus» et «Résineux»), comparaison des différents tests menés

Logiciel	sources de données	précision de l'utilisateur	précision du producteur
D E F I N I E N S	toutes	0.90	0.84
	toutes sauf SPOT	0.88	0.82
	juillet 2006 et SPOT		
	octobre 2005		
toutes sauf SPOT	0.86	0.83	
juillet 2006			
toutes sauf SPOT	0.91	0.82	
octobre 2005			
ENVI	toutes	0.86	0.38

On remarque que la précision du producteur obtenue via l'utilisation du logiciel Envi Ex est bien moindre (0.38) que celle obtenue dans les différents tests impliquant Definiens. Cette faible précision du producteur relativise donc l'apparente bonne précision de l'utilisateur. Même si l'analyse des causes de ces disparités n'a pas pu être approfondie, on peut d'ores et déjà affirmer sur base de l'expérience acquise lors de cette étude exploratoire que le processus de segmentation d'Envi engendre une lourdeur notable qui serait démultipliée dans l'optique d'un procédé de production à plus grande échelle.

La suppression des attributs provenant des images SPOT a légèrement diminué la précision globale, tout en maintenant un bon niveau (les précisions concernant les classes mineures étant similaires). L'attribut SPOT lié au nœud principal qui opère la grande séparation entre classes arborées et les autres a été remplacé par la moyenne de l'attribut HEST provenant de la couche de hauteur produite en tâche 2.

- L'utilisation des arbres de classification devrait pouvoir être complétée d'une étape de post-classification comportant ensemble de géotraitements exploitant les relations topologiques entre segments et permettant d'éliminer certaines confusions sur base de règles de voisinage.
- La production d'une couche image « haute résolution » réalisée à la même période sur l'ensemble de la Wallonie ne semble pas constituer une priorité. En effet, cette démarche ne serait pas en phase avec le découpage et le calendrier d'acquisition des ortho-images sur la région. En outre cette manière de procéder induirait sans doute plus de contraintes qu'elle n'apporterait d'avantages. Il est plutôt recommandé d'opérer un découpage du territoire sur base du découpage du plan d'acquisition des ortho-images et de la disponibilité des images SPOT. Chacun de ces sous-blocs pourrait être traité sur base d'un arbre de classification spécifique, dès lors que les objets de référence soient disponibles en quantité suffisante.
- Dans une perspective de mise en production du processus de classification, il pourrait s'avérer intéressant d'appliquer les deux étapes du processus de cartographie, non plus sur un îlot, mais sur un bloc constitué de plusieurs ortho-images.

4. Références bibliographiques

Baatz M., Schäpe A., 2000. Multiresolution Segmentation – an optimization approach for high quality multi-scale image segmentation. In: *Angewandte Geographische Informationsverarbeitung XII. Beiträge zum AGIT-Symposium*, Salzburg, 12-23.

Blaschke T., 2010, Object based image analysis for remote sensing, *ISPRS Journal of Photogrammetry and Remote Sensing*, Volume 65, Issue 1, Pages 2-16.

Binard M., 2005. Usage des données 3D IGN, PICC et SETHY, pour l'orthorectification d'une image QuickBird en région liégeoise. *AM/FM GIS BELUX News*, n° 32, 20-24.

Bonnet S., 2009. Un modèle numérique de canopée pour l'estimation de la hauteur dominante des peuplements résineux en Région Wallonne. *Forêt Wallonne* n°98, 53-59.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone, 1984. *Classification and Regression Trees*, Wadsworth & Brooks/Cole Advanced, Books & Software, Pacific Grove, CA, USA.

Brostaux Y., 2005. Etude du classement par forêts aléatoires d'échantillons perturbés à forte structure d'interaction, *Faculté Universitaire des Sciences Agronomiques de Gembloux*, pp. 168.

Bunting P. and R. Lucas, 2006. The delineation of tree crowns in Australian mixed species forests using hyperspectral Compact Airborne Spectrographic Imager (CASI) data. *Remote Sensing of Environment* 101(2): 230-248.

Carleer A.P. and E. Wolff, 2006. Region-based classification potential for land-cover classification with very high spatial resolution satellite data. In: Stefan Lang T. B. and Elisabeth S. (Eds.), *1st International Conference on Object-based Image Analysis.*, Salzburg University, Austria, July 4-5, 2006.

Chavez, P.S., 1996. Image-based atmospheric corrections – revisited and improved. *Photogrammetric Engineering and Remote Sensing*, 62,1025-1036.

Clinton N., Holt A., Yan L., Gong P., 2008. An accuracy Assessment Measure For Object Based Image Segmentation. In: *GEOBIA 2008 - Pixels, Objects, Intelligence: GEOgraphic Object Based Image Analysis for the 21st Century*, The International Archives of The Photogrammetry, Remote Sensing and Spatial Information Sciences, Calgary, Alberta, Canada.

Cornet Y., Binard M., Laplanche F. and Lambotte J-M., à soumettre. 2D and 3D data in mobility and planning application.

Dagnelie P., 2006, *Statistique théorique et appliquée. Tome 2. Inférence statistique à une et à deux dimensions*. Bruxelles, De Boeck et Larcier, 734 p.

Demarcin P., Smoos A., Dautrebande S., Mokadem A.I., Degré A., 2009. 'Projet Erruissol ou la cartographie des zones à risque de ruissellement et d'érosion en Région Wallonne (Belgique)', *10èmes journées d'Etude des sols*, Strasbourg.

Demarcin P., Degré A., Smoos A., Dautrebande S., 2009. Projet ERRUISSOL. Cartographie numérique des zones à risques de ruissellement et d'érosion des sols en Région wallonne. Rapport final de convention DGO3-FUSAGx. Unité d'hydrologie et hydraulique agricole. Faculté universitaire des Sciences agronomiques de Gembloux. 55p+annexes.

Demoulin A., Bovy B., Rixhon G., and Cornet Y., 2007. An automated method to extract fluvial terraces from digital elevation models: The Vesdre valley, a case study in eastern Belgium. *Geomorphology*, Vol. 92, Is. 3-4, 51-64.

DGA et Unité de Géomatique, 2008. Etude du contrôle de la qualité de l'ortho-image et du Modèle Numérique de Surface. Vols 2006-2007. Avril 2008, 39 p.

FUSAGx, Hydrologie et Hydraulique, 2008. *Projet Erruissol*, Rapport final, Gembloux.

Geneletti D. and B.G.H. Gorte, 2003. A method for object-oriented land cover classification combining Landsat TM data and aerial photographs. *International Journal of Remote Sensing*. 24(6): 1273-1286.

Gorham B. 1998. 'Mapping multi-temporal agricultural landuse in the Mississippi alluvial valley of Arkansas', Paper presented at the Spatial technologies in agriculture and natural resources, SRIEG-10 annual conference.

Hajek F., 2006. Object-oriented classification of Ikonos satellite data for the identification of tree species composition. *J. For. Sci.* 52(4): 181-187

Haralick R.M., K. Shanmugam, et al., 1973. Textural Features for Image Classification. *IEEE Transactions on Systems, Man and Cybernetics*. 3(6).

Liaw A. and M. Wiener, 2002. Classification and Regression by randomForest. *R News* 2(3).

Lucas R., Rowlands A., Brown A., Keyworth S. & Bunting P. 2007. Rule-based classification of multi-temporal satellite imagery for habitat and agricultural land cover mapping, *ISPRS Journal of Photogrammetry & Remote Sensing*, 62, pp. 165-168.

Lillesand T., Kiefer R.W., Chipman J., 2008, *Remote Sensing and Image Interpretation*, Wiley Ed. 6th ed., 756 p.

Mather P.M., 1999. *Computer Processing of remotely-sensed Images. An Introduction*. Second edition. Wiley, 292 p.

User's Guide

ENVI, 2008a: Feature Extraction Module Version 4.6 December, 2008 Edition, ITT Visual Information Solutions

ENVI, 2008b: Getting started with ENVI Version 4.6 December, 2008 Edition, ITT Visual Information Solutions

ENVI, 2009: ENVI EX Tutorial: Working with ENVI EX

Definiens, 2007a: Definiens Developer 7.0 Reference Book. Definiens AG, 2007.

Definiens, 2007b: Definiens Developer 7.0 User's Guide. Definiens AG, 2007.

Sites Internet

Site 1 : Programme ORFEO sur le site du CNES : <http://smc.cnes.fr/PLEIADES/Fr/>

Site 2 : Site de l'ORFEO Toolbox : <http://www.orfeo-toolbox.org>

Site 3 : Imagery and Sources for Vegetation Mapping, California Department of Fish and Game, Vegetation Classification and Mapping Program and California Native Plant Society Vegetation Program, Sacramento :
http://www.cnps.org/cnps/vegetation/pdf/mapping_images_maps.pdf

Site 4 : Portail cartographique de la Région Wallonne (métadonnées) :
<http://cartographie.wallonie.be>

Annexe 1 :

Signification des abréviations d'attributs utilisées lors de la tâche 7 (analyses thématiques), renvoi aux différents tableaux les définissant.

Dénomination	Signification	définition
Area_Pxl	Area	Tableau 38, p. 138
Border_ind	Border Index	
Border_len	Border Length	
Length_Pxl	Length	
LengthWidt	Length-to-Width ratio	
Shape_inde	Shape Index	
Width_Pxl	Width	
Hest	Couche de hauteur (MNS_DGA - MNT_ERRUISSOL)	
GLCM_Contr	GLCM Contrast	Tableau 37, p. 137
GLCM_Corre	GLCM Correlation	
GLCM_Entro	GLCM Entropy	
GLCM_Mean_	GLCM Mean	
GLCM_StdDe	GLCM StDev	

Dénomination	Signification	définition
Ob1	ORTHO_DGA bande 1 (Rouge)	Tableaux 35 et 36, pp. 135 et 136
Ob2	ORTHO_DGA bande 2 (Vert)	
Ob3	ORTHO_DGA bande 3 (Bleu)	
Ob4	ORTHO_DGA bande 4 (proche IR)	
Obr	ORTHO_DGA bright	
Ogdvi	ORTHO_DGA GDVI	
Ogndvi	ORTHO_DGA GNDVI	
Ogrvi	ORTHO_DGA GRVI	
Ohp	ORTHO_DGA High-pass	
Ondvi	ORTHO_DGA NDVI	
Onir	ORTHO_DGA NIR	
Odb1	idem ci-dessus , ORTHO_DGA dégradées à 2,5 m	
Odb2		
Odb3		
Odb4		
Odbr		
Ogdvi		
Ogndvi		
Odgrvi		
Ondvi		
Odnir		

Dénomination	Signification	définition
S7b1	SPOT5 - juillet 2006 bande 1 (proche IR)	Tableaux 35 et 36, pp. 135 et 136
S7b2	SPOT5 - juillet 2006 bande 2 (Rouge)	
S7b3	SPOT5 - juillet 2006 bande 3 (Vert)	
S7b4	SPOT5 - juillet 2006 bande 4 (IR moyen)	
S7gdvi	SPOT5 - juillet 2006 GDVI	
S7gndvi	SPOT5 - juillet 2006 GNDVI	
S7grvi	SPOT5 - juillet 2006 GRVI	
S7ndvi	SPOT5 - juillet 2006 NDVI	
S7nir	SPOT5 - juillet 2006 NIR	
S10b1	SPOT5 - octobre 2005 bande 1 (proche IR)	
S10b2	SPOT5 - octobre 2005 bande 2 (Rouge)	
S10b3	SPOT5 - octobre 2005 bande 3 (Vert)	
S10b4	SPOT5 - octobre 2005 bande 4 (IR moyen)	
S10gdvi	SPOT5 - octobre 2005 GDVI	
S10gndvi	SPOT5 - octobre 2005 GNDVI	
S10grvi	SPOT5 - octobre 2005 GRVI	
S10ndvi	SPOT5 - octobre 2005 NDVI	
S10nir	SPOT5 - octobre 2005 NIR	

Annexe 2 :

Résultats de 3 forêts aléatoires réalisées sur le set d'îlots numéro 1 (résultats de Definiens) en supprimant tous les attributs SPOT du jeu de données (Tableau 1), les attributs issus de la SPOT 5 de juillet 2006 (Tableau 2) et enfin , les attributs issus de la SPOT 5 d'octobre 2005

Tableau 1 :

Attribut	MDG	Attribut	MDG
Hest_M	380.93	GLCM_StdDe	77.94
Ohp_Sd	331.27	Odb3_M	74.50
Odb4_M	223.98	Area_Pxl	71.42
Odb4_Sd	202.13	Odb1_Sd	70.90
Ob4_Sd	182.77	Ogndvi_M	67.32
Ogdvvi_M	174.08	Ogndvi_Sd	65.54
Ogdvi_M	153.25	Surf_seg	65.26
Hest_Sd	150.79	Obr_M	64.69
Odbr_Sd	147.51	GLCM_Mean_	64.03
Odbr_M	133.08	Odb3_Sd	62.68
Odb2_Sd	131.23	Ondvi_M	62.15
Obr_Sd	127.27	Onir_M	58.55
Ogrvi_M	124.88	Border_len	58.00
Ob2_Sd	116.09	Ogrvi_Sd	57.55
Odgndvi_M	114.39	Length_Pxl	56.52
Ogdvvi_Sd	114.35	Ob1_M	55.44
Ob4_M	110.46	Width_Pxl	53.81
Odgrvi_M	109.05	Ob2_M	53.68
Odb1_M	104.17	Ob1_Sd	53.23
Ondvi_M	95.50	LengthWidt	50.92
GLCM_Entro	95.19	Odgrvi_Sd	48.31
Ogdvi_Sd	95.13	Ob3_M	47.40
Onir_Sd	91.98	Odgndvi_Sd	43.28
Ob3_Sd	91.12	GLCM_Corre	42.05
Ondvi_Sd	89.78	Ondvi_Sd	40.69
Odnir_M	89.29	Odnir_Sd	39.29
Odb2_M	81.42	Shape_inde	38.80
Ohp_M	81.05	Border_ind	37.80
GLCM_Contr	80.64		

Tableau 2 :

Attribut	MDG	Attribut	MDG
S10b3_M	404.60	S1Ogndvi_M	49.30
S10b2_M	357.28	Ohp_M	47.13
S10b4_M	352.77	Obr_M	45.61
Hest_M	311.69	Ogndvi_Sd	45.08
Ohp_Sd	244.44	S1Ogrvi_M	45.07
Odb4_Sd	155.70	S1Onir_M	44.98
Ob4_Sd	141.41	Odb3_Sd	43.46
Hest_Sd	123.31	Ogrvi_Sd	43.35
Odb4_M	118.65	GLCM_Mean_	42.23
Ogdvi_M	114.46	Ondvi_M	40.94
S10b1_M	114.27	Surf_seg	40.38
Obr_Sd	108.61	Onir_M	39.05
Ogdvi_M	105.84	Ob1_M	38.25
Odbr_Sd	102.54	Area_Pxl	38.03
Ob3_Sd	86.85	Ob2_M	37.70
Odb2_Sd	86.42	Width_Pxl	36.04
Odgrvi_M	83.89	Length_Pxl	35.81
Odgndvi_M	80.79	Ob1_Sd	35.28
Ondvi_Sd	78.94	Odgrvi_Sd	35.16
Ogrvi_M	78.11	Ob3_M	33.38
Ob2_Sd	73.35	S10b4_Sd	32.90
GLCM_Contr	72.56	Border_len	32.89
Odb1_M	70.82	S10b3_Sd	32.61
Odbr_M	70.69	S10b2_Sd	30.31
Ogdvi_Sd	69.70	LengthWidt	30.29
Ondvi_M	63.27	Odgndvi_Sd	27.47
S10gdvi_M	61.74	GLCM_Corre	26.85
GLCM_Entro	61.46	Odnir_Sd	25.21
Onir_Sd	59.97	Ondvi_Sd	24.68
Ob4_M	58.83	Shape_inde	23.88
Odb3_M	58.77	Border_ind	22.55
Odb1_Sd	58.62	S1Ondvi_Sd	22.00
Odnir_M	58.04	S10gdvi_Sd	20.32
GLCM_StdDe	58.03	S1Onir_Sd	19.16
Odb2_M	57.61	S10gndvi_Sd	19.12
S1Ondvi_M	53.38	S10b1_Sd	18.95
Ogdvi_Sd	53.05	S1Ogrvi_Sd	17.83
Ogndvi_M	51.48		

Tableau 3 :

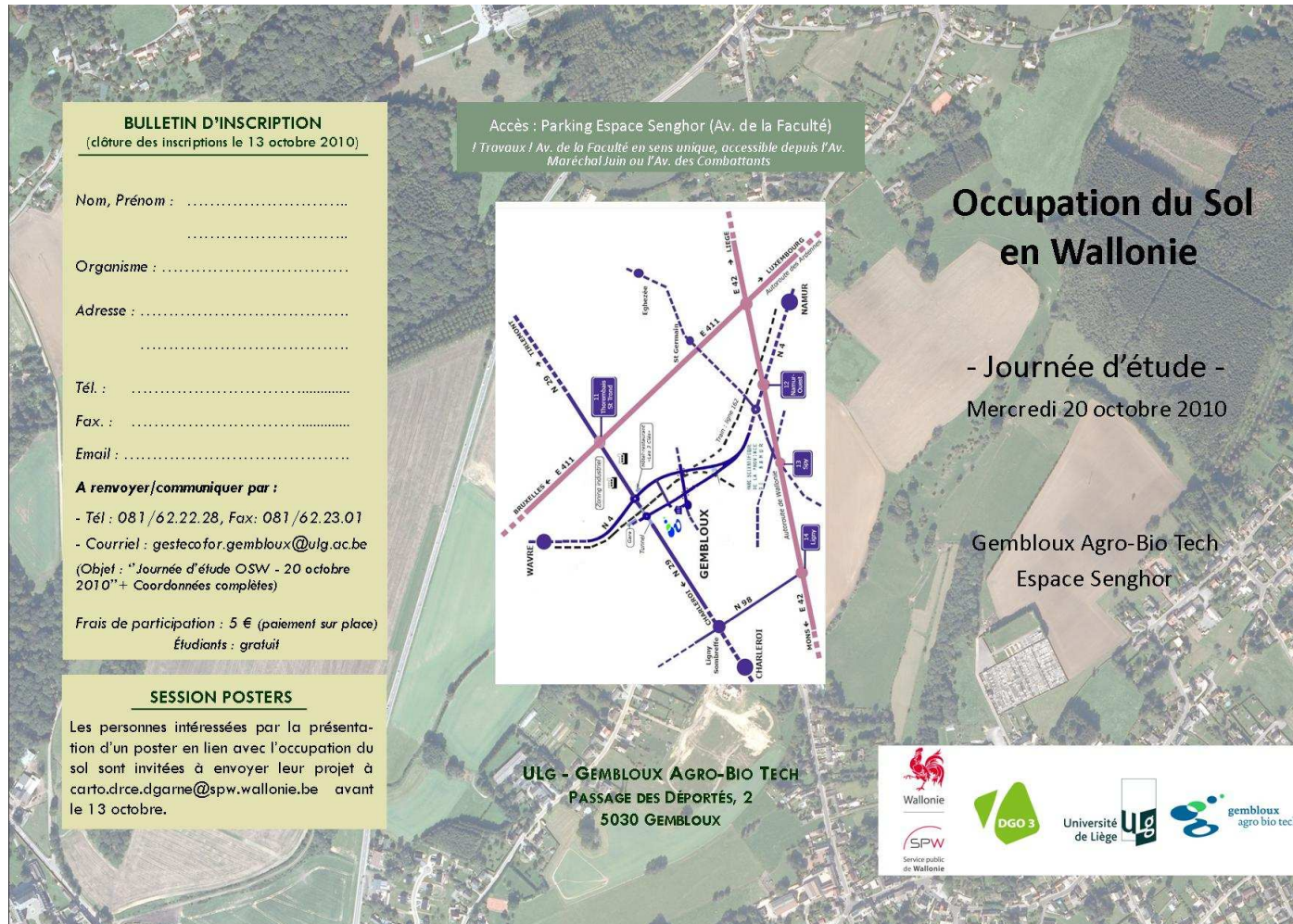
Attribut	MDG	Attribut	MDG
S7b3_M	401.15	Obr_M	47.80
S7b4_M	325.07	Odnir_M	47.72
S7b2_M	269.44	GLCM_StdDe	47.10
S7nir_M	246.46	Odb3_M	43.33
S7gndvi_M	239.59	Ogrvi_Sd	43.11
S7ndvi_M	227.86	S7grvi_Sd	41.67
Hest_M	209.98	Ogndvi_Sd	41.02
Ohp_Sd	204.36	Odb3_Sd	39.08
S7gdvi_M	200.39	Ogndvi_M	38.68
S7b1_M	155.66	Onir_M	38.54
Ob4_Sd	125.04	GLCM_Mean_	37.62
Odb4_Sd	118.74	Ob2_M	37.09
Odb4_M	90.03	Ob1_Sd	36.84
Obr_Sd	87.39	Odgrvi_Sd	36.49
S7grvi_M	84.37	Ob1_M	35.60
Odbr_Sd	83.43	S7b2_Sd	34.77
Odb2_Sd	73.64	Ohp_M	34.60
Ogdvi_M	71.99	Area_Pxl	33.11
Odb1_Sd	71.20	Surf_seg	32.89
Ob2_Sd	70.53	Ob3_M	31.02
Ogdvi_M	68.36	S7b3_Sd	30.98
Ob3_Sd	68.08	Ogndvi_Sd	29.94
Odbr_M	67.85	Length_Pxl	29.87
GLCM_Contr	63.23	Odnir_Sd	29.18
Hest_Sd	61.05	GLCM_Corre	29.06
Ob4_M	60.60	Width_Pxl	28.94
Ondvi_Sd	58.12	Odndvi_Sd	28.36
Ogndvi_M	57.11	LengthWidt	27.76
Ogdvi_Sd	55.16	Border_len	27.62
Ogrvi_M	54.56	S7gndvi_Sd	25.68
Ogdvi_Sd	53.19	S7b4_Sd	25.21
Odgrvi_M	52.60	S7gdvi_Sd	25.04
Odb2_M	52.21	Shape_inde	23.64
GLCM_Entro	51.56	S7nir_Sd	23.48
Ondvi_M	49.64	S7b1_Sd	23.44
Onir_Sd	49.41	S7ndvi_Sd	22.83
Odndvi_M	48.42	Border_ind	21.68
Odb1_M	47.83		

Annexe 3 :

Résultats d'une forêt aléatoire réalisée sur le set d'îlots numéro 1 (résultats d'Envi) avec tous les attributs SPOT du jeu de données

Attribut	MDG	Attribut	MDG	Attribut	MDG
S7grvi_Sd	14.77	Odndvi_Sd	51.38	Odb3_M	159.45
S10grvi_Sd	17.44	Odgrvi_Sd	56.17	Odndvi_M	166.75
S10b3_Sd	21.94	Ob3_M	59.89	Ob2_Sd	171.85
S10b2_Sd	22.06	Odbr_Sd	62.93	Odnir_M	178.55
S10gndvi_Sd	22.17	Hest_Sd	64.51	GLCM_Mean_	179.05
S10nir_Sd	22.55	Odb3_Sd	73.31	S10gndvi_M	184.99
Border_len	22.96	Ob1_Sd	77.45	Ogndvi_M	187.88
S10ndvi_Sd	23.53	Ondvi_M	79.04	Odbr_M	207.24
Border_ind	24.60	Ob4_M	83.75	Ob4_Sd	207.56
S7b2_Sd	25.08	Onir_Sd	83.90	S10gdvi_M	207.71
S10gdvi_Sd	25.83	Obr_M	84.75	S10b1_M	223.33
Length_Pxl	26.12	Ogdvi_Sd	86.85	Ogdvi_M	225.86
S7nir_Sd	26.37	Odgdvi_Sd	90.89	S10nir_M	229.39
S7gdvi_Sd	26.41	Odb4_Sd	95.90	Odgrvi_M	234.04
S7b1_Sd	26.66	S10grvi_M	100.62	S7ndvi_M	236.98
S7ndvi_Sd	27.11	Onir_M	106.14	S7b1_M	255.91
Area_Pxl	27.38	S7grvi_M	110.25	Odgdvi_M	256.00
S7b4_Sd	27.60	Odb2_Sd	110.51	S7gdvi_M	267.21
S7b3_Sd	27.69	Ob2_M	111.71	S7nir_M	277.59
S10b1_Sd	27.82	Ondvi_Sd	112.09	Ogrvi_M	281.62
Width_Pxl	28.05	Ogrvi_Sd	116.28	S7gndvi_M	291.25
S7gndvi_Sd	28.13	Odb1_M	116.51	GLCM_Contr	291.61
Surf_seg	28.25	Ogndvi_Sd	121.29	S10ndvi_M	331.70
S10b4_Sd	29.30	Ohp_Sd	121.50	GLCM_StdDe	375.79
Shape_inde	29.83	Odndvi_M	127.21	S10b2_M	375.97
LengthWidt	32.33	Odb2_M	132.87	Hest_M	436.04
GLCM_Corre	33.89	Obr_Sd	134.35	S7b2_M	528.01
GLCM_Entro	35.77	Odb1_Sd	135.14	S10b4_M	534.57
Odndvi_Sd	41.27	Ob3_Sd	151.49	S10b3_M	536.41
Odnir_Sd	41.41	Odb4_M	156.51	S7b4_M	578.81
Ohp_M	41.44	Ob1_M	156.91	S7b3_M	635.35

Annexe 4 : Journée d'étude sur l'occupation du sol en Wallonie : programme



BULLETIN D'INSCRIPTION
(clôture des inscriptions le 13 octobre 2010)

Nom, Prénom :

Organisme :

Adresse :

Tél. :

Fax :

Email :

A renvoyer/communiquer par :

- Tél : 081/62.22.28, Fax: 081/62.23.01
- Courriel : gesteco@ulg.ac.be

(Objet : "Journée d'étude OSW - 20 octobre 2010" + Coordonnées complètes)

Frais de participation : 5 € (paiement sur place)
Étudiants : gratuit

SESSION POSTERS

Les personnes intéressées par la présentation d'un poster en lien avec l'occupation du sol sont invitées à envoyer leur projet à carto.drce.dgame@spw.wallonie.be avant le 13 octobre.



Accès : Parking Espace Senghor (Av. de la Faculté)
Travaux l'Av. de la Faculté en sens unique, accessible depuis l'Av. Marechal Juin ou l'Av. des Combattants

Occupation du Sol en Wallonie

- Journée d'étude -
Mercredi 20 octobre 2010

Gembloux Agro-Bio Tech
Espace Senghor

ULG - GEMBLoux AGRO-BIO TECH
PASSAGE DES DÉPORTÉS, 2
5030 GEMBLoux



PROGRAMME	PROGRAMME	RÉSUMÉ DE LA JOURNÉE
<p>9h00 - 9h30 : Accueil</p> <p>9h30 - 9h40 : Allocation de M. Benoît Lutgen, Ministre des Travaux publics, de l'Agriculture, de la Ruralité, de la Nature, de la Forêt et du Patrimoine.</p> <p>9h40 - 9h50 : Intervention de la Direction du Développement Rural, SPW-DGARNE (A.I. Mokadem)</p> <p>9h50 - 10h20 : La carte d'occupation du sol de Wallonie, version 2.07 (P. Lejeune, GxABT - ULg)</p> <p>10h20 - 10h50 : <i>Pause-café</i></p> <p>10h50 - 11h20 : Utilisation conjointe d'orthoimages et d'images SPOT pour la cartographie de l'occupation du sol (Y. Cornet, ULg et D. Brogna, GxABT - ULg)</p> <p>11h20 - 11h40 : Mise à jour de cartes d'occupation des sols via l'analyse de données de télédétection (J. Radoux, UCL)</p> <p>11h40 - 12h00 : Questions - réponses</p> <p>12h00 - 13h30 : <i>Repas et session « Posters »</i></p>	<p>13h30 - 13h50 : Réalisation d'une cartographie d'occupation du sol transfrontalière, (C. Tesnière, M. Haine et I. Chambodut, Parc naturel transfrontalier du Hainaut)</p> <p>13h50 - 14h10 : COSW et mesures de densité pour l'évaluation de l'impact de la structure du territoire wallon sur les émissions de GES (F. Pirart, Lepur - ULg)</p> <p>14h10 - 14h30 : L'évolution de la superficie résidentielle par habitant : un indicateur de développement territorial durable? (J. Charlier et I. Reginster, IWEPS)</p> <p>14h30 - 14h50 : Reconstruction de la dynamique des paysages à partir des documents historiques : la dynamique complexe des milieux semi-naturels (J. Piqueray, GxABT - ULg)</p> <p>14h50 - 15h10 : Questions - réponses</p>	<p>Cette journée d'étude sur la thématique de l'Occupation du Sol est organisée dans le cadre du « Projet de Cartographie Numérique de l'Occupation du Sol de Wallonie » initié par la DGARNE en 2005.</p> <p>Le projet vise à produire une cartographie de l'utilisation du sol à grande échelle sur base de données raster, vectorielles et alphanumériques gérées par les services du Gouvernement wallon, à l'intégrer dans un SIG et à la mettre à jour.</p> <p>Lors de cette journée, différentes interventions s'articuleront autour de cette thématique allant de la présentation de techniques de production et d'amélioration de cartes d'occupation du sol jusqu'à l'utilisation de celles-ci. Une nouvelle version de la carte d'occupation du sol de Wallonie sera également présentée.</p> <p>Organisation :</p> <ul style="list-style-type: none"> - Unité de Gestion des Ressources forestières et des Milieux naturels, GxABT - ULg - Unité de Géomatique, ULg