

Université de Toulouse

MASTER 2 GEOMATIQUE

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

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

RAPPORT DE STAGE

Expertise qualité de la production d'une couche d'occupation du sol amenant à des « indicateurs aménagement des territoires urbains »

Galvani Aurélie

Centre d'études et d'expertise sur les risques, l'environnement, la mobilité
et l'aménagement



Maître de stage : Dominique HEBRARD
Tuteur-enseignant : David SHEEREN

Septembre 2018

Résumé

Le CEREMA développe un outil permettant d'obtenir une carte d'occupation du sol à partir d'images satellites. Celui-ci amène à la mise en place d'indicateurs métiers abordant des problématiques liés aux grands enjeux urbains tels que la consommation de l'espace, l'imperméabilisation des surfaces, la climatologie urbaine et les îlots de chaleur, ou encore la composante verte en ville. L'objectif de ce travail est la mise en place d'une méthodologie permettant de définir la qualité de ces résultats de classification.

Dans une première étape, la méthode de classification du CEREMA a été appliquée sur la zone de Bordeaux métropole afin d'obtenir une carte d'occupation du sol. Cette étape a permis de prendre en main la méthode ainsi que d'apporter des améliorations quand cela était possible.

Dans une seconde étape, la qualité de la donnée de classification en sortie a été étudiée pour la zone Bordeaux métropole. Une méthode prenant en compte 6 critères de qualité a été proposée, dont 5 sont cités dans la norme ISO 19157 et le dernier développé pour cette étude.

On obtient en résultats une notation de la qualité pour chaque critère permettant de définir si la donnée obtenue est en accord avec la qualité nécessaire pour un usage donné (définie comme la qualité externe).

Afin de pouvoir reproduire cette méthode sur de multiples cas d'études, des fiches méthodologiques sont créés pour accompagner l'utilisateur dans l'application de cette méthodologie de qualité.

Abstract

The CEREMA develops a method to map land cover from satellite images. This method defines technical indicators that address problematics linked with majors urban issues, such as land consumption, waterproofed areas, urban climatology and urban "heat island" effect and green areas in cities. This work aims at developing a methodology to define the quality of these classification's results.

In a first phase, the CEREMA's classification method is applied on Bordeaux metropole to obtain a land covermap. It allowed me to take control of the method and make improvements when it is possible.

In a second phase, the classification data's quality for Bordeaux metropole has been studied. A methodology using 6 quality criteria is proposed. 5 of these criteria are defined following the ISO 19157 norm, and the last one is developed especially for this study.

In results is obtained a quality notation for each criteria, which allows us to define if the data's quality is in accordance with the quality required for specific uses (defined as the external quality). To replicate this method on multiple case studies, methodological sheets are proposed to helps end-users to apply these quality definition methods.

Remerciements :

Je tenais tout d'abord à remercier Dominique HEBRARD pour sa disponibilité et son encadrement au cours de ce stage ainsi que pour les nombreux conseils qu'il a pu me donner.

Je remercie également Gilles FOUVET pour son aide qui m'a permis d'avancer de nombreuses fois dans mon travail ainsi que Guillaume SAINTPIERRE pour son aide et ses conseils en statistique.

Je remercie Benjamin PICCINNI pour ses échanges mais également sa patience lors de l'interprétation de nombreux points d'échantillonnages.

Je tenais également à remercier David SHEEREN pour ses conseils sur le sujet de la qualité.

Enfin je souhaite remercier toute les personnes travaillant au DALETT ainsi que les stagiaires présents lors de ces 4 mois mais surtout Syrine KMIHA pour toutes les discussions que nous avons pu avoir lors des pauses du midi.

Sommaire

Sommaire	1
Table des figures.....	3
Table des tableaux.....	4
Table des annexes	5
I. Introduction.....	6
II. Contexte	6
1. <i>Structure d'accueil</i>	6
2. <i>Projet Urban View</i>	6
3. <i>Objectif et enjeux</i>	7
III. Etat de l'art	8
1. <i>Classification du sol</i>	8
2. <i>Qualité</i>	9
a. <i>Enjeux</i>	9
b. <i>Métadonnées</i>	9
c. <i>Qualité interne/externe</i>	10
IV. Présentation de la zone d'étude et des données utilisées.....	11
1. <i>Zone d'étude</i>	11
2. <i>Les données</i>	12
a. <i>Pléiade</i>	12
b. <i>MNS</i>	12
c. <i>MNT 1m</i>	12
d. <i>BDtopo® bâtiments et routes</i>	13
V. Production de la classification	13
1. <i>Méthodologie générale</i>	13
2. <i>Etape 1: La préparation des fichiers</i>	16
3. <i>Etape 2 et 3: Préparation des échantillons d'apprentissages et classification supervisée</i>	19
4. <i>Etape 4: Post-traitements</i>	20
a. <i>Micro-classes</i>	22
b. <i>Macro-classes</i>	24
5. <i>Résultats de la classification finale et apport de chaque post-traitements</i>	26
6. <i>Bilan</i>	27
7. <i>Plan d'assurance qualité</i>	27
VI. Production qualité	27
1. <i>La cohérence logique</i>	28

a.	Méthodologie	28
b.	Résultats pour Bordeaux Métropole zone Est et Ouest	29
c.	Conclusions.....	29
2.	<i>L'exhaustivité</i>	29
a.	Méthodologie	30
b.	Résultats pour Bordeaux Métropole zone Est et Ouest	30
c.	Conclusions.....	33
3.	<i>La précision thématique/sémantique</i>	34
a.	Méthodologie	34
b.	Résultats pour Bordeaux Métropole zone Est et Ouest	36
c.	Conclusions.....	40
4.	<i>La précision de position</i>	40
a.	Méthodologie	41
b.	Résultats pour Bordeaux Métropole zone Est et Ouest	41
c.	Conclusions.....	42
5.	<i>La qualité temporelle</i>	43
a.	Méthodologie	43
b.	Résultats	43
c.	Conclusions.....	43
6.	<i>L'écart des dynamiques</i>	44
a.	Méthodologie	44
b.	Résultats pour Bordeaux Métropole zone Est	44
c.	Conclusions.....	45
VII.	Résultats	46
1.	<i>Qualité interne</i>	46
2.	<i>Qualité externe</i>	46
VIII.	Etude de dynamique 2014-2016	47
IX.	Conclusion	49
	<i>SOURCES CONSULTEES</i>	51
	Bibliographie:	51
	Webographie:.....	52

Table des figures

Figure 1: Critères permettant de décrire les métadonnées d'une donnée.....	10
Figure 2: Description de la qualité interne et externe d'un jeu de donnée géographique.....	10
Figure 3: Critère permettant de décrire la qualité d'une donnée.....	11
Figure 4: Zone d'étude de Bordeaux Métropole.....	12
Figure 5: Qualité d'acquisition des données du RGE Alti® 1m sur la zone de Bordeaux métropole, réalisé à partir du tableau d'assemblage RGE Alti®.....	13
Figure 6: Description de la chaîne de traitement comprenant les différentes tâches utilisées (Celle-ci sont décrites en annexe 1).....	15
Figure 7:Création du MNH à partir du MNT et du MNS (source: Pôle de Compétences et d'Innovation Applications Satellitaires du CEREMA).....	16
Figure 8: Choix des outils pour l'étape consistant à boucher les zones non corrélées. Pour cela il faut (1) définir redéfinir les zones en no_data (qui correspondent à 0 sur le MNS), puis (2) combler ces zones par interpolation.....	17
Figure 9: Diagramme de flot pour la création du MNH.....	19
Figure 10: Diagramme de flot des post-traitements.....	21
Figure 11: Exemple de résultat pour le post-traitement 1. a) Capture d'une image satellite Pléiade sur la zone Bordeaux Ouest, b) Capture de la classification brute sur la même zone, c) Capture de la classification après avoir appliqué le post-traitement 1. En rouge la classe bâtie, en vert la végétation, en jaune le sol nu et en violet les routes.....	22
Figure 12: Exemple de résultat pour le post-traitement 2. a) Capture d'une image satellite Pléiade sur la zone Bordeaux Ouest, b) Capture de la classification brute sur la même zone, c) Capture de la classification après avoir appliqué le post-traitement 2. En rouge la classe bâtie, en vert la végétation, en jaune le sol nu, en violet les routes et en bleu les surface en eau.....	23
Figure 13: Exemple de résultat pour le post-traitement 3. a) Capture d'une image satellite Pléiade sur la zone Bordeaux Ouest, b) Capture de la classification brute sur la même zone, c) Capture de la classification après avoir appliqué le post-traitement 3. En rouge la classe bâtie, en vert la végétation, en violet les routes et en bleu les surface en eau.....	23
Figure 14: Exemple de résultat pour le post-traitement 4. a) Capture d'une image satellite Pléiade sur la zone Bordeaux Ouest, b) Capture de la classification brute sur la même zone, c) Capture de la classification après avoir appliqué le post-traitement 4. En rouge la classe bâtie, en vert la végétation, en violet les routes et en bleu les surface en eau.....	24
Figure 15: Exemple de résultat pour le post-traitement 5. a) Capture d'une image satellite Pléiade sur la zone Bordeaux Ouest, b) Capture de la classification brute sur la même zone, c) Capture de la classification après avoir appliqué le post-traitement 5. En rouge la classe bâtie, en vert la végétation, en violet les routes, en jaune les sols nus et en bleu les surface en eau.....	24
Figure 16: Exemple de résultat pour le post-traitement 6. a) Capture d'une image satellite Pléiade sur la zone Bordeaux Ouest, b) Capture de la classification brute sur la même zone, c) Capture de la classification après avoir appliqué le post-traitement 6. En rouge la classe bâtie, en vert la végétation, en violet les routes, en jaune les sols nus et en bleu les surface en eau.....	25
Figure 17: Exemple de résultat pour le post-traitement 7. a) Capture d'une image satellite Pléiade sur la zone Bordeaux Ouest, b) Capture de la classification brute sur la même zone, c) Capture de la classification après avoir appliqué le post-traitement 7. En rouge la classe bâtie, en vert la végétation, en violet les routes et en jaune le sol nu.....	25
Figure 18: Exemple de résultat pour le post-traitement 8. a) Capture d'une image satellite Pléiade sur la zone Bordeaux Ouest, b) Capture de la classification brute sur la même zone, c) Capture de la classification après avoir appliqué le post-traitement 8. En rouge la classe bâtie, en vert la végétation, en violet les routes et en jaune le sol nu.....	26

Figure 19: Choix effectués pour faire un échantillonnage (modifié à partir de la norme ISO19157).....	30
Figure 20: Exemple de comparaison entre l'image Pléiade et la classification (ici bâti présent sur l'image Pléiade et absent de la classification).	31
Figure 21: Résultats obtenu sur des classification entre de 2014 et 2016 lors de la construction d'un bâtiment.	34
Figure 22: Exemple de difficulté d'interprétation de la classe présente sur l'image Pléiade. Avec ici (a) l'image Pléiade comparée à (b) une image tirée de google satellite permettant d'obtenir en (c) une idée plus précise de la forme des bâtiments et ainsi de les visualiser sur l'image Pléiade.	35
Figure 23: Représentation du nombre d'éléments en déficits, en excédents ou corrects pour chaque classe de la classification finale, calculé à partir de le matrice de confusion.	38
Figure 24: Représentation du nombre d'éléments en déficits, en excédents, corrects et qui se compensent pour chaque classe de la classification finale, calculé à partir de le matrice de confusion.	38
Figure 25 : Pourcentage de bâti sur Bordeaux métropole découpé en des zones de 500m2 réalisé à partir de la classification brute, finale et de la BDtopo.	39
Figure 26: Comparaison entre les bâtiments de la BDtopo et ceux de l'image pléiade.	40
Figure 27: visualisation de l'écart de position entre la classification et l'image Pléiade pour chacun des points d'échantillonnage (les points orange correspondent à des échantillons situés sur des bâtiments présentant des ombres).	42
Figure 28: Ecart d'estimation des dynamiques urbaine entre la classification et la référence.	45
Figure 29: Représentation possible de la qualité externe dans le cas où l'objectif est la mesure de taux de végétation en ville.	46
Figure 30: Représentation possible de la qualité externe dans le cas où l'objectif est la mesure des surfaces imperméabilisées.....	46
Figure 31 : Représentation possible de la qualité externe dans le cas où l'objectif est la mesure des dynamiques urbaines qu'elle soit considérées comme faibles ou fortes.	47
Figure 32: Etude de l'évolution du bâti sur Bordeaux Métropole Est à partir d'un quadrillage de 500m2 (partie1)	48
Figure 33: Etude de l'évolution du bâti sur Bordeaux Métropole Est à partir d'un quadrillage de 500m2 (partie2)	48

Table des tableaux

Tableau 1: Différentes occupation du sol obtenues à partir d'images satellites et leurs caractéristiques.	8
Tableau 2: Formule utilisée sur Bandmath (OTB) afin de connaitre l'apport et la localisation de chaque post-traitement.	26
Tableau 3: Résultats de l'apport des différents post-traitements sur la classification.	26
Tableau 4: Résultats obtenus pour la cohérence logique sur la zone de Bordeaux métropole.....	29
Tableau 5: Notation du critère de cohérence logique.....	29
Tableau 6: Résultats obtenus lors d'une première comparaison des bâtiments entre l'image pléiade et la classification.....	31
Tableau 7 : Eléments en excès et omis obtenu lors d'une première comparaison des bâtiments entre l'image pléiade et la classification.	31
Tableau 8: Résultats obtenus lors d'une deuxième comparaison des bâtiments entre l'image pléiade et la classification.....	32
Tableau 9: Eléments en excès et omis obtenus lors d'une deuxième comparaison des bâtiments entre l'image pléiade et la classification.	32
Tableau 10 : Résultats obtenus lors de l'ajout des deux comparaisons des bâtiments entre l'image pléiade et la classification.....	32
Tableau 11 : Eléments en excès et omis obtenus lors de l'ajout des deux comparaisons des bâtiments entre l'image pléiade et la classification.	33

Tableau 12: Notation du critère d'exhaustivité.	33
Tableau 13: Matrice de confusion entre la classification finale et les résultats de l'interprétation sur l'image Pléiade.	37
Tableau 14 : Pourcentage d'éléments en excès et omis pour chaque classe de la classification finale, calculés à partir de la matrice de confusion.	37
Tableau 15: Comparaison du nombre d'élément bien classé et du Kappa entre la classification finale et la classification brute.:	39
Tableau 16: Notation du critère de précision thématique.	40
Tableau 17: Notation du critère de précision de position.	42
Tableau 18: Millésime utilisé pour chaque donnée en entrée de la chaîne de traitement.	43
Tableau 19: Notation du critère de qualité temporelle.	43
Tableau 20: Pourcentage de pixels omis sur Bordeaux métropole zone Est pour le millésime 2014.	44
Tableau 21: Pourcentage de pixels en excès sur Bordeaux métropole zone Est pour le millésime 2016.	45
Tableau 22: Notation du critère de dynamique.	45

Table des annexes

Annexe 1 : Les tâches utilisées dans la chaîne de traitement.	54
Annexe 2 : Exemple d'un fichier XML associé à une tâche de la chaîne de traitement.	56
Annexe 3 : Exemple du monitoring des tâches.	57
Annexe 4 : Résultats brut de la classification sur la zone de Bordeaux Métropole Ouest: Millésime 2016.	58
Annexe 5 : Résultats de la classification après les post-traitements sur la zone de Bordeaux Métropole Ouest : Millésime 2016.	59
Annexe 6 : Carte des modifications apportées par les post-traitements.	60
Annexe 7 : Fiche critère de cohérence logique.	61
Annexe 8 : Fiche critère d'Exhaustivité.	63
Annexe 9 : Fiche critère de précision thématique : justesse de classement.	70
Annexe 10 : Fiche critère de précision thématique : Précision des attributs quantitatifs.	80
Annexe 11 : Fiche critère de précision de position.	83
Annexe 12 : Fiche critère de qualité temporelle.	89
Annexe 13 : Fiche critère d'écart des dynamiques.	91
Annexe 14 : Choix de l'échantillonnage.	94
Annexe 15 : Etude de qualité de Bordeaux métropole.	99
Annexe 16 : Diagramme de Gantt.	107

I. Introduction

Ce stage d'une durée de 4 mois porte sur la mise en place d'une méthodologie pour définir la qualité d'une méthode mise en place par le CEREMA permettant d'obtenir des cartes d'occupation du sol. Celles-ci amènent à la mise en place d'indicateurs métiers abordant des problématiques liées aux grands enjeux urbains tels que la consommation de l'espace, l'imperméabilisation des surfaces, la climatologie urbaine et les îlots de chaleur, ou encore la composante verte en ville.

Cette étude s'est divisée en deux grandes parties. Il s'agit tout d'abord d'appliquer le traitement développé par la CEREMA pour obtenir une carte d'occupation du sol à partir d'image Pléiade pour la zone de Bordeaux métropole. Cette étape m'a permis de prendre en main la méthode mais également d'y apporter des améliorations quand cela était possible. Lors d'une deuxième étape, une méthodologie a été mise en place pour qualifier la qualité interne et externe des données produites par la chaîne de traitement avec pour objectif de se rapprocher le plus possible de la norme ISO 19157 sur la qualité de la donnée géographique.

II. Contexte

1. Structure d'accueil

Le CEREMA (Centre d'étude et d'expertises sur les Risques, l'Environnement, la Mobilité) est un établissement public rattaché depuis mai 2017 au ministère de la transition écologique et solidaire et au ministère de la cohésion des territoires.

Le CEREMA mène des activités d'expertises, de contrôles, de suivis, de conceptions et de recherches dans l'ensemble des domaines d'activité du ministère. Il accompagne les territoires pour la réalisation de leurs projets dans 9 champs d'action qui sont:

- l'aménagement et la cohésion des territoires
- la transition énergétique et le climat
- la mobilité et le transport
- les infrastructures de transport
- les stratégies urbaines
- l'environnement et les ressources naturelles
- la prévention des risques
- le bien-être et la réduction des nuisances
- l'habitat

Le CEREMA sud-ouest dont la direction se situe à Saint-Médard-en-Jalles (agglomération Bordelaise), est implanté sur les sites de Bordeaux et de Toulouse, ce dernier se nommant la DALETT (Délégation Aménagement Laboratoire Expertise Transport de Toulouse).

Le groupe Satellite, Climat, Gestion et Système d'information (SCGSI) dans lequel j'ai effectué mon stage fait partie de la délégation de Toulouse. Celui-ci traite de différentes thématiques telles que le littoral, les risques, l'occupation du sol ou encore le climat urbain.

Le CEREMA sud-ouest héberge le pôle "Applications Satellitaires et Télécommunication" qui a pour but d'intégrer l'utilisation d'images satellitaires dans la réalisation des missions du ministère, qu'elles soient en lien avec le littoral, les risques ou l'aménagement du territoire (CEREMA, 2018a).

2. Projet Urban View

Le CEREMA a mis en place une chaîne de traitement à partir d'images satellitaires afin de répondre aux grands enjeux urbains avec des problématiques liées à l'aménagement du territoire. En effet les

collectivités demandent aux organismes compétents de mettre en place une solution pour mesurer l'étalement urbain et suivre l'évolution de la consommation des terres.

L'étalement urbain constitue une forme d'urbanisation qui s'est généralisée depuis de nombreuses années dans les villes françaises. Elle correspond à une consommation toujours plus importante des espaces par les zones urbaines liée à l'augmentation de la population dans les grandes villes et donc l'expansion spatiale de celles-ci en couronnes périurbaines (Antoni & Youssoufi, 2007; Laugier, 2012).

Ces questions de consommation raisonnée de l'espace sont de plus en plus présentes dans la réglementation en matière d'urbanisme et d'environnement avec une volonté de mieux maîtriser l'aménagement du territoire. La législation sur ce sujet débute en 2000 avec la loi SRU puis le Grenelle de l'environnement en 2010 avec la loi Grenelle II (Laugier, 2012) et plus récemment, la loi ALUR en 2014 (Ministère du logement et de l'égalité des territoires, loi ALUR, 2014).

Cependant, pour bien définir la consommation de l'espace, il est nécessaire d'avoir des outils permettant de calculer ces évolutions de manière précise, ce qui n'est pas encore le cas de nos jours. En effet, des méthodes tel que Corine Land Cover ou Teruti-Lucas permettent d'observer une évolution de l'artificialisation, mais pas de la quantifier de manière précise, et encore moins pour des périodes rapprochées (Bechet et al, 2017).

La recherche de nouveaux outils pour répondre à ces problématiques est toujours d'actualité, et il est possible de se demander si l'apport de l'imagerie spatiale de très haute résolution peut apporter des réponses aux diverses problématiques urbaines. Dans cette optique, le projet Urban-View a été lancé au CEREMA dans le but d'apporter des réponses aux problématiques tel que l'imperméabilisation des sols, la densité urbaine, la climatologie urbaine et les îlots de chaleur (CEREMA, 2016).

3. Objectif et enjeux

L'objectif de ce stage est de mettre en place une méthode permettant d'analyser la qualité de la donnée en sortie de la chaîne de traitement mise en place par le CEREMA, tout en restant le plus proche possible des normes et directives en place actuellement.

Le CEREMA est en contact avec TerraNIS, une société spécialisée dans la conception, le développement et la commercialisation de services de géo-information dans les domaines de l'agriculture et de l'environnement. Cette société a pour but de commercialiser les indicateurs territoriaux basés sur la méthode mise en place par le CEREMA. En cela, l'objectif majeur de mon stage est de définir la qualité du produit en sortie de la chaîne de traitement mais également de mettre en place d'une méthodologie afin de savoir si le produit obtenu répond à une demande client en terme de qualité.

La première étape de ce stage consiste à effectuer la chaîne de traitement et obtenir une carte d'occupation du sol sur la zone de Bordeaux métropole, afin tout d'abord de prendre en main la méthode, mais également d'avoir un nouveau regard sur:

- la fluidité de la chaîne, en y apportant des améliorations quand cela est possible,
- améliorer la documentation déjà existante ou la créer,
- créer un fichier de paramétrage de référence pour lancer la chaîne de traitement pour des classifications dans l'urbain, utilisable par d'autres utilisateurs pour les prochaines études.

La deuxième étape de ce stage consiste, comme expliqué précédemment, à mettre en place une méthodologie à partir de normes et de directives existantes. Pour cela j'ai tout d'abord effectué un état de l'art sur la qualité des données géographiques, les différents critères existants et leurs significations. J'ai ensuite défini comment il était possible d'utiliser ces critères sur la méthode de classification du CEREMA avec le plus d'objectivité possible (par exemple concernant l'échantillonnage). Ces critères ont été appliqués sur ma zone d'étude, ce qui a permis de définir la qualité de la classification du sol effectué précédemment sur la zone Bordeaux métropole, avec la création d'une fiche de qualité pour cette zone

(annexe 15). Enfin, j'ai réalisé des fiches méthodologiques pour pouvoir reproduire ce travail sur la qualité dans d'autres zones d'études (annexe 7 à 13).

III. Etat de l'art

1. Classification du sol

Il existe plusieurs méthodes amenant à des données d'occupation du sol provenant d'images satellitaires. Les principales méthodes qui existent dans la littérature ainsi que leurs caractéristiques sont résumées dans le tableau 1 (COPERNICUS, 2013; European Commission, 2011; Ministère de la transition écologique et solidaire, 2018; IGN, 2016; THEIA, 2018; CESBIO, 2015).

Nom	Producteur	Données utilisées	Résolution	type	version	Nombre de classes	Zones
<i>Urban Atlas</i>	GMES/COPERNICUS	SPOT 5&6 et Formosat-2	2,5m	vecteur	2006, 2012	28	Européen: 800 zones urbaines
CORINE Land Cover	l'Agence européenne de l'environnement	Landsat, SPOT, IRS	20m	Rasters, vecteur	1990, 2000, 2006, 2012	44	Européen: 39 états
<i>OCS GE (Occupation du sol à grande échelle nationale)</i>	IGN	BD Ortho®, BD Topo®, Photo-interprétation	métrique	vecteur	2016	32	France métropolitaine et départements et régions d'outre-mer
<i>CES Occupation des sols (OSO)</i>	CESBIO, l'UMR Ispa, Dynafor, CNRM, UMR Tetis, IGN-Matis, Costel et Sertit	Sentinelle 2A et 2B	Entre 10m et 20m	Rasters, vecteur	Mise à jour annuelle	15 à 20 classes	France métropolitaine

Tableau 1: Différentes occupation du sol obtenues à partir d'images satellites et leurs caractéristiques.

Cela permet d'avoir une idée de l'enjeu du projet d'urban-view. En effet, le CEREMA a voulu développer un produit permettant d'obtenir une occupation du sol ayant à la fois:

- Une bonne résolution spatiale, permettant d'obtenir des informations plus fines et donc de travailler à des échelles moins importantes que celles déjà existantes.
- La possibilité d'obtenir une image quand on le souhaite en connaissant la date de prise de vue avec la programmation des images satellites Pléiade et donc ne pas être limité à des millésimes comme cela arrive pour certaines méthodes pour certaines méthodes (colonne version, tableau 1).
- Un outil peu excessif niveau prix permettant de financer des millésimes plus fréquents.

C'est donc pour cela que le CEREMA a mis en place ce produit, les produits déjà existant ne permettant pas forcément de répondre aux problématiques sur lequel travaille le CEREMA.

2. Qualité

a. Enjeux

Avec l'ouverture de l'open data et l'échange des données géo-localisées, de plus en plus de données sont accessibles par tous. Il est cependant important de bien les utiliser, ce qui nécessite de connaître leurs particularités (Peraldi, 2007).

Un des enjeux clés pour faciliter l'usage de ces données par un utilisateur est de bien définir la qualité de la donnée, que ce soit par les métadonnées ou bien par des critères de qualité permettant de vérifier si le niveau de qualité du produit est en adéquation avec l'utilisation qu'il est prévu d'en faire.

Plusieurs enjeux entourent cette notion de qualité (CERTU, 2008):

- Un enjeu économique, afin de ne pas acquérir des données qui ne correspondent pas à l'usage voulu.
- Un enjeu juridique car le diffuseur de la donnée est responsable de la qualité de celle-ci.
- Un enjeu organisationnel; il est en effet plus facile de trouver une donnée utile pour un projet donné si ses métadonnées bien définies. De même, cela facilite la vérification de la correspondance de la qualité de la donnée par rapport à l'usage voulu.
- Un enjeu de professionnalisation, car manipuler des données en prenant en compte leur qualité ainsi que bien définir la qualité d'un produit doit se faire avec une grande rigueur. Il faut également mieux cerner les utilisations possibles pour ces données et la relation entre la qualité et l'usage voulue.

Pour confirmer la prise de conscience lié aux problématiques sur la qualité des données géographique, de plus en plus de groupe de travail et de séminaires sont mis en place totalement dédiés à ces problématiques, comme le séminaire QuaDoGeo qui s'est déroulé à Lyon en février 2018 (CEREMA, 2018b).

b. Métadonnées

Une première étape pour bien définir la qualité de la donnée est de bien renseigner les métadonnées correspondantes.

Les informations sur la métadonnée ont été récupérées à partir de la norme ISO 19115 (Organisation internationale de normalisation, 2014; Ministère du budget et de la réforme de l'état, ministère du tourisme de l'équipement du transport et de la mer, 2006; NCDDC, NODC, NOAA, 2012; BRGM, 2016), la directive INSPIRE (CNIG, 2018 ; CNIG, 2013) et les standards COVADIS (Commission de validation des données pour l'information spatialisée) (Ministère de la transition écologique et solidaire, 2016; Ministère de l'alimentation de l'agriculture et de la pêche, 2010). La figure 1 regroupe les différents critères décrivant une métadonnée en fonction des différentes sources (ISO 19115, INSPIRE, COVADIS). On peut observer que ces critères sont pratiquement les mêmes entre ces trois références. Certains se regroupent en un seul critère alors qu'ils correspondent à plusieurs critères sur une autre source, comme c'est le cas pour les mots-clés dans la norme ISO 19115, qui correspondent à trois sous critères dans la directive INSPIRE (figure 1). Cependant il existe une cohérence entre les différentes approches, que se soit internationale avec la norme ISO, européenne avec INSPIRE ou bien nationale avec la COVADIS.

Qualité de la Métadonnée

ISO 19115	INSPIRE	COVADIS
Identificateur de la ressource unique	Identificateur de la ressource unique	Identificateur de la ressource unique
Intitulé de la ressource	Intitulé de la ressource	Intitulé de la ressource
Résumé de la ressource	Résumé de la ressource	Résumé de la ressource
Langue de la ressource	Langue de la ressource	Langue de la ressource
Autre langages utilisés	Autre langages utilisés	Autre langages utilisés
Langue des métadonnées	Langue des métadonnées	Langue des métadonnées
Catégorie thématique	Catégorie thématique	Catégorie thématique
Thème INSPIRE	Thème INSPIRE	Thème INSPIRE
Mots clés	Mots clés obligatoire INSPIRE Mots clés recommandés Facultatifs Mots clés libres Facultatifs	Mots clés INSPIRE Autres mots clés
Référence temporelle	Dates de référence (publication, création, dernière révision)	Date des métadonnées
Version (nombre, date)		Version (nombre, date)
Encodage de la ressource	Encodage de la ressource	Encodage de la ressource
Encodage des caractères	Encodage des caractères	Encodage des caractères
Auteur : (nom, pays, adresse, telephone...)	Localisateur de la ressource	Fournisseur
Contact de la métadonnées		
Type de la ressource	Type de la ressource	Type d'objet géométrique
Généalogie	Généalogie	Généalogie
Lot de données parent (code)	Lot de données parent (code)	Source des données
Lot de données parent (nom)	Lot de données parent (nom)	
Type de représentation géographique	Type de représentation géographique	Type de représentation spatiale
Référentiel de coordonnées	Référentiel de coordonnées	Système de Référence géodésique
Information des représentation spatiale	Projection	Projection
Résolution spatiale	Résolution spatiale	Résolution spatiale
Rectangle de délimitation géographique	Rectangle de délimitation géographique	Rectangle de délimitation géographique
Conformité INSPIRE	Conformité INSPIRE	Conformité INSPIRE
Conformité COVADIS	Conformité COVADIS	Conformité COVADIS
Etendue temporelle	Etendue temporelle	Etendue temporelle
Cohérence topologique	Cohérence topologique	Cohérence topologique
Système de référence temporel	Système de référence temporel	Système de référence temporel
Conditions applicables à L'utilisation dans le service et à l'accès, à la diffusion, à la réutilisation	Conditions applicables à L'utilisation dans le service et à l'accès, à la diffusion, à la réutilisation	Conditions applicables à L'utilisation dans le service et à l'accès, à la diffusion, à la réutilisation
Contraintes (usage, légal, sécurité)	Restrictions sur l'accès public	Restrictions sur l'accès public
Information sur les mise à jours des métadonnées	Information sur les mise à jours des métadonnées	Information sur les mise à jours des métadonnées
spécifications de produit	spécifications de produit	spécifications de produit
Limite d'utilisation de la ressource	Limite d'utilisation de la ressource	Limite d'utilisation de la ressource

Figure 1: Critères permettant de décrire les métadonnées d'une donnée.

c. Qualité interne/externe

En plus des métadonnées, on peut parler de la qualité de la donnée. Celle-ci est décrite et spécifiée au sein de la norme iso 19157 (ORGANISATION INTERNATIONALE DE NORMALISATION, 2013; Bonin & Gallais, 2017) et de la directive INSPIRE. On peut noter qu'il n'est pas obligatoire de définir la qualité de la donnée (CERTU, 2010). Pourtant il est indispensable de la connaître car une même donnée peut être utilisée pour un usage mais ne pas être d'assez bonne qualité pour un autre usage. Il est donc important de définir les différents types de qualité de la donnée (CERTU, 2010; Bonin & Gallais, 2017):

- La qualité interne: exprime la qualité d'un lot de donnée produit au regard de spécifications, rédigées au préalable (figure 2).
- La qualité externe: se définit comme l'adéquation à un besoin exprimé. Aptitude d'un jeu de donnée à satisfaire un usage donné (figure 2).

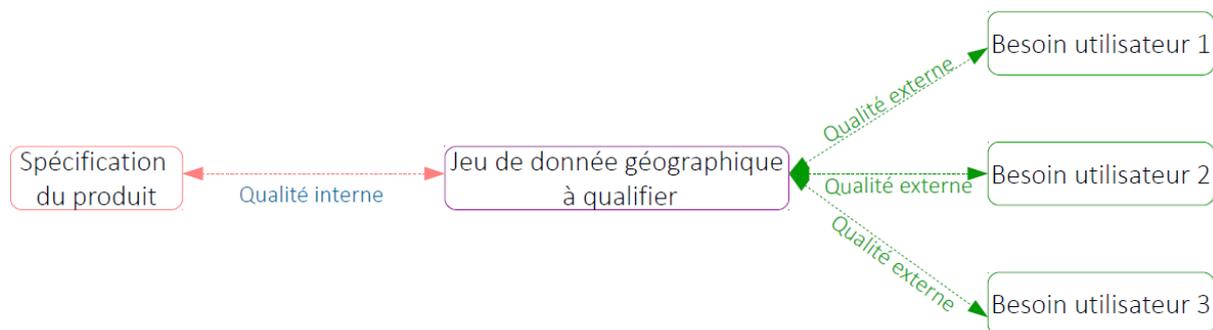


Figure 2: Description de la qualité interne et externe d'un jeu de donnée géographique.

Afin de définir la qualité du produit, plusieurs critères ont été mis en place et sont définis dans la norme ISO 19157 et dans INSPIRE. Ces différents critères sont présents sur la figure 3.

Qualité de la Donnée

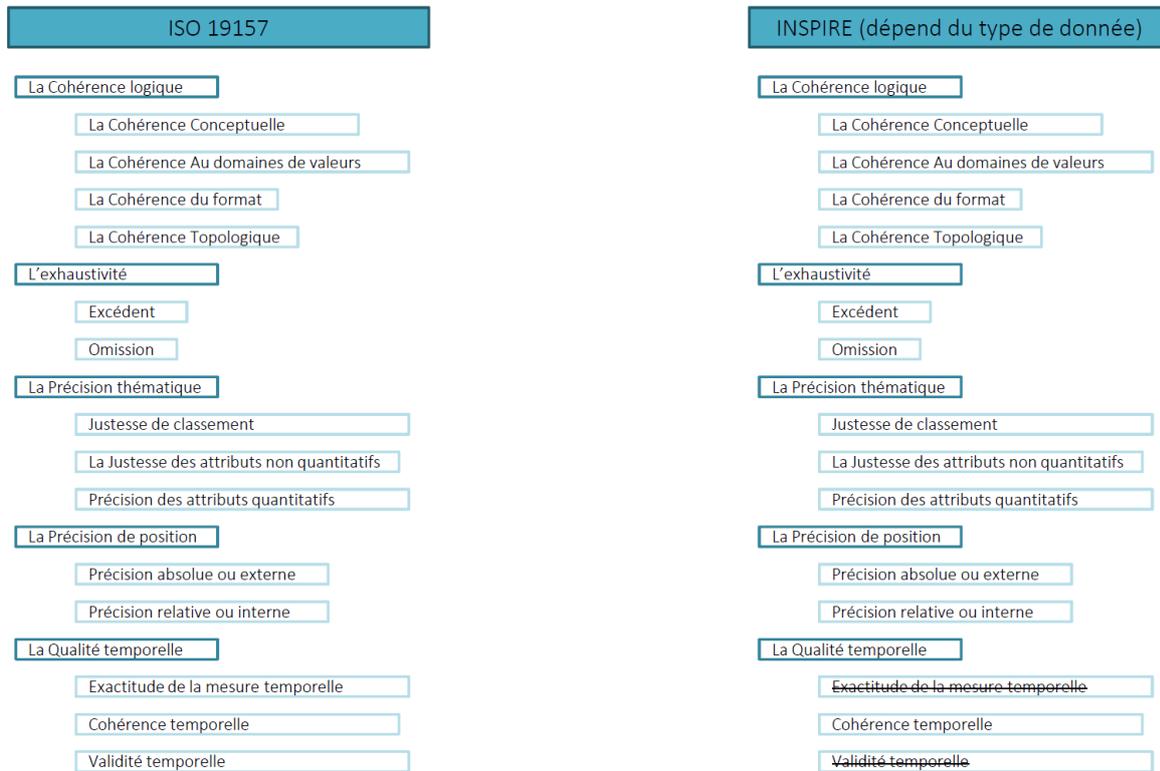


Figure 3: Critère permettant de décrire la qualité d'une donnée.

Ce rapport aura pour but de mettre en place une méthodologie pour chacun de ces critères afin de définir la qualité finale de la classification du sol réalisée lors de cette étude (Paragraphe VI. Production qualité).

IV. Présentation de la zone d'étude et des données utilisées

1. Zone d'étude

La zone d'étude se situe sur Bordeaux métropole. Ce choix provient du fait que de nombreuses données sont disponibles sur cette zone.

En fonction des études que j'ai réalisées, j'ai travaillé sur différentes zones de Bordeaux (figure 4).

- J'ai effectué la première étude, correspondant à la production de la classification (Paragraphe V. Production de la classification), sur la partie Ouest de Bordeaux.
- J'ai effectué la seconde étude, correspondant à l'étude de la qualité, sur la zone Est et Ouest de Bordeaux (Paragraphe VI. Production qualité).
- La troisième étude comprend toutes les comparaisons que j'ai réalisé entre les millésimes, que ce soit la partie qualité/écart des dynamiques (Paragraphe VI. Production qualité), ou encore la partie étude des dynamiques 2014-2016 (Paragraphe VIII. Etude de dynamique 2014-2016). Cela provient du fait que pour le millésime 2014, il existe une classification correcte que pour la zone Est de Bordeaux.

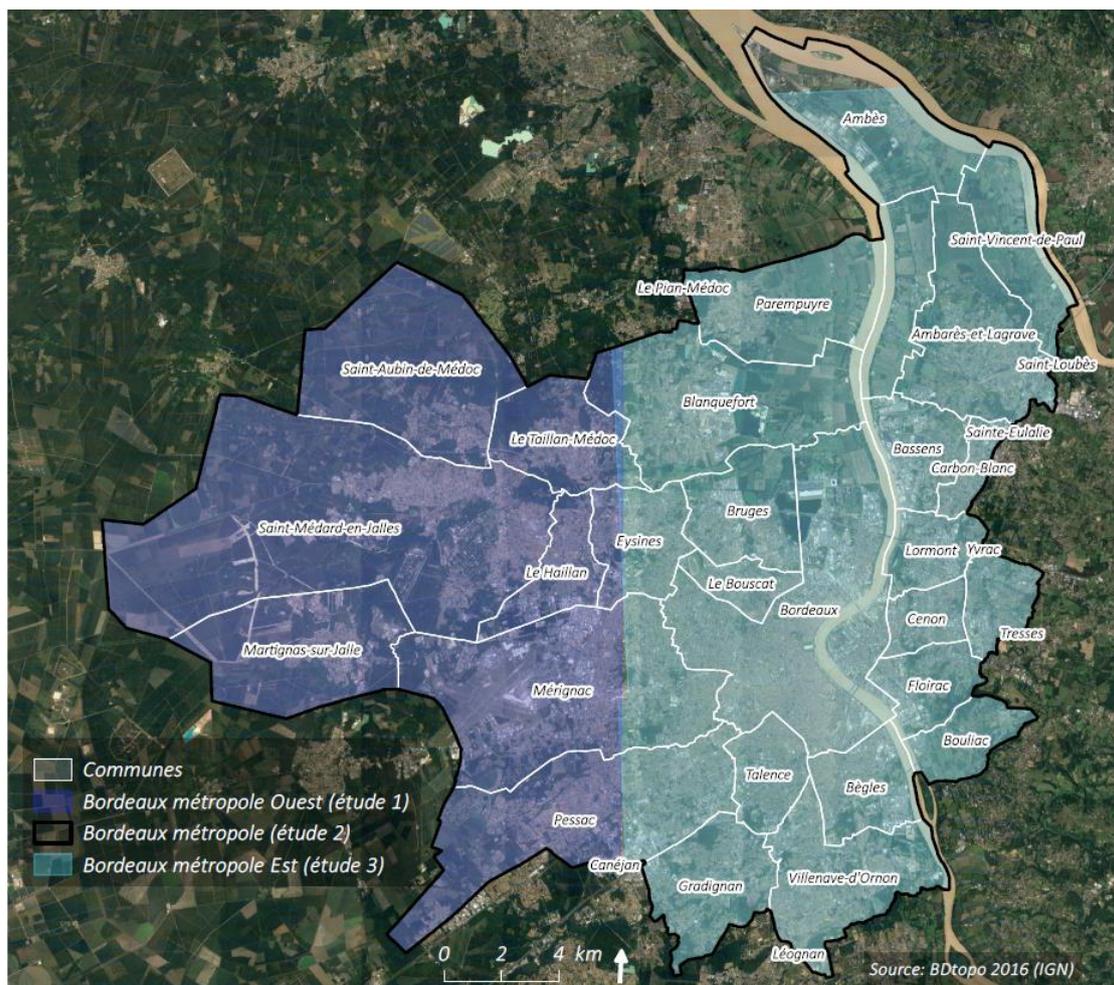


Figure 4: Zone d'étude de Bordeaux Métropole.

2. Les données

a. Pléiade

Les images utilisées pour effectuer ces études sont des images satellitaires de très haute résolution spatiale provenant des satellites Pléiades 1A et 1B. Sa résolution spatiale est de 70cm avec un ré-échantillonnage de 50cm. Il est important de noter que les satellites Pléiades fonctionnent en mode "programmation". En effet l'utilisateur peut choisir sa zone d'intérêt et la date à laquelle il souhaite obtenir des images (IGN, 2015; spécificités sur l'annexe 15, 1. Métadonnées).

b. MNS

Le modèle numérique de surface est obtenu à partir de couples d'images Pléiade. En effet, les satellites Pléiade ont la possibilité de changer d'angle de visée, ce qui permet d'obtenir des acquisitions stéréoscopiques (prise de vue d'une même zone par deux points de vue légèrement distant mais sur un même plan). Grâce à cela il est possible d'obtenir un Modèle Numérique de Surface représentant l'altitude de chaque pixel (spécificités sur l'annexe 15, partie 1. Métadonnées).

c. MNT 1m

Une donnée utilisée lors de ces traitements est le modèle numérique de terrain provenant du volet Inondation du service RGE Alti® de l'IGN constitués de dalles kilométriques couvrant la zone de Bordeaux métropole. Dans cette zone d'étude, la technologie Lidar est déployée ainsi que de la photographie

aérienne, comme on peut l'observer en figure 5 (IGN, 2017; spécificités sur l'annexe 15, partie 1. Métadonnées).

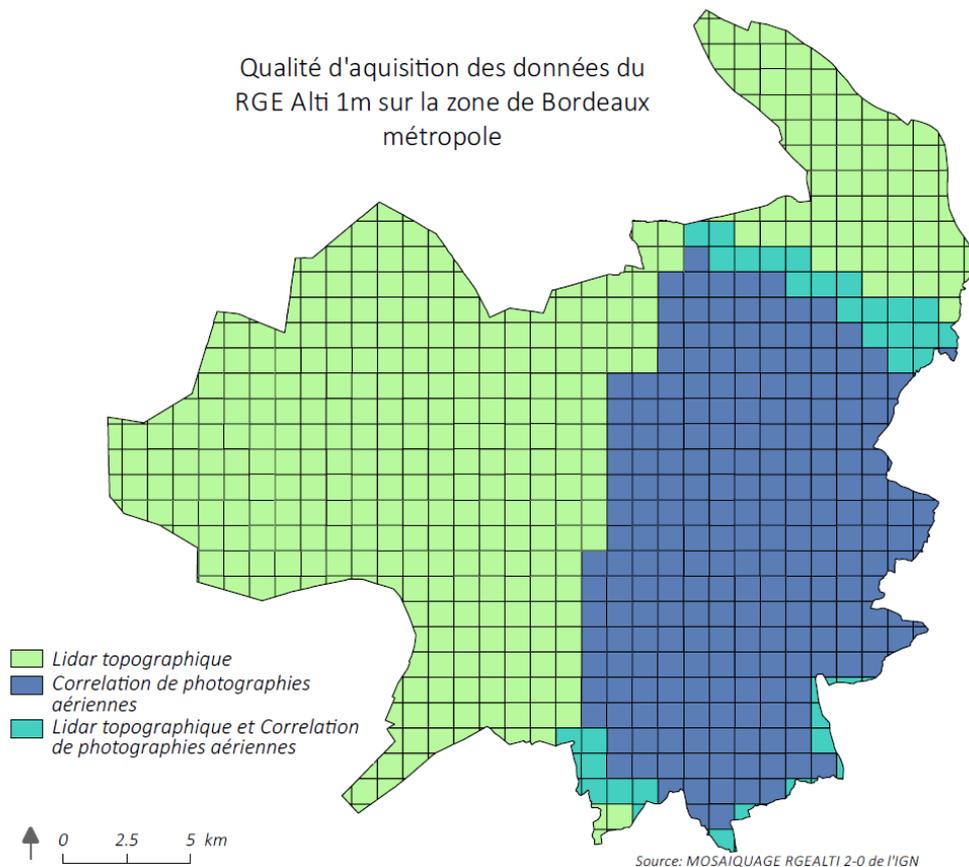


Figure 5: Qualité d'acquisition des données du RGE Alti® 1m sur la zone de Bordeaux métropole, réalisé à partir du tableau d'assemblage RGE Alti®.

Les données sont de très bonne qualité avec une exactitude altimétrique entre 0,2 et 0,5m pour le LIDAR et 0,5 à 0,7m pour la corrélation de photographie aériennes.

d. BDtopo® bâtiments et routes

L'utilisation de la BDtopo® de l'IGN concerne principalement les bâtiments et les routes. Ces données ont une précision métrique et ont été livrées en avril 2016. Cette période est en accord avec les millésimes des images Pléiade utilisées (IGN, 2018; spécificités sur l'annexe 15, partie 1. Métadonnées).

V. Production de la classification

1. Méthodologie générale

La méthodologie utilisée pour établir une classification du sol à partir des images satellitaires Pléiade a déjà été mise en place par le CEREMA et est en phase de production. Par ce fait, il n'est possible de mettre en place que des améliorations mineures.

Cette méthode comporte quatre grandes étapes, la première correspond à la préparation des données en amont de la classification, la deuxième étant la préparation des échantillons d'apprentissages, la troisième est la classification supervisée et enfin la dernière correspond au post-traitements.

Une nomenclature simple de la classification des types de sols a été mise en place, pour un compromis entre la technique et la thématique. Ces classes ont été choisies pour une reproductivité plus facile et pour rendre la méthode plus stable. Le choix des échantillons d'apprentissages est également plus simple à mettre en place.

Ces classes sont au nombre de 5 et ont chacune un code associé:

- Les surfaces bâties : 11100.
- La végétation : 20000.
- Les surfaces en eau : 12200
- Les surfaces de sol nu : 13000
- Les surfaces minéralisées (route, parking, trottoirs, terrasses) : 11200

La chaîne de traitement mise en place par le CEREMA fonctionne de la manière suivante:

- Il faut tout d'abord paramétrer un fichier XML permettant de simplifier l'utilisation de la chaîne ainsi que d'archiver les paramètres utilisés. Toutes les applications et leurs paramètres sont ainsi présents dans le fichier XML. L'utilisateur entre les différents paramètres (liés aux applications) dans le XML, qui fait le lien avec les applications codées en python développées par le CEREMA. Ces différentes applications sont nommées "tâches" dans la suite du document.
- Pour optimiser le temps de calcul, les tâches présentes dans le XML sont dirigées dans un séquenceur qui les lancera selon un ordre que l'on aura prédéfini dans le fichier XML. Certaines tâches peuvent alors se faire en parallèle ce qui permet d'optimiser le temps de traitements.

Les différentes tâches que j'ai utilisé dans le XML sont expliquées dans l'annexe 1.

Dans le cadre de cette étude, j'ai appliqué cette méthodologie à un jeu de donnée disponible sur Bordeaux métropole Ouest afin de :

- Prendre en main et comprendre le fonctionnement de la chaîne de traitement,
- Améliorer si cela est possible la chaîne de traitement (paragraphe V.2. La préparation des fichiers),
- Créer un fichier XML de référence pour des classifications dans l'urbain, utilisable par d'autres utilisateurs pour les prochaines études (exemple du fichier XML en annexe 2).

Cette étude m'a également permis de vérifier, modifier, compléter et créer les documents composant un guide d'utilisation de la chaîne de traitement.

En parallèle, j'ai réalisé un monitoring des tâches afin d'évaluer le temps de calcul de chaque tâche et l'espace de stockage utilisé par les fichiers temporaires et les fichiers de sorties (annexe 3). Dans l'optique de l'industrialisation de l'outil, ces informations seront utiles pour prévoir la durée nécessaire au traitement et l'espace de stockage qui doit être disponible pour la réalisation de chaque tâche. On note que si le nombre de pixels de l'image satellite diffère grandement de celui utilisé pour cette étude, ces valeurs ne représentent qu'un ordre de grandeur, puisque l'évolution du temps de calcul n'est pas forcément linéaire avec la taille des fichiers d'entrées.

La méthodologie étant déjà validée, et les calculs étant suffisamment rapide, nous avons choisi de l'appliquer tout de suite sur l'ensemble de Bordeaux Ouest, afin de gagner du temps. Lorsque des problèmes ont été rencontrés, comme lors de la création du Modèle Numérique de Hauteur (MNH), j'ai réalisé des études sur des zones tests plus petites afin de valider la méthodologie.

Pour effectuer la classification, plusieurs tâches sont réalisées. Pour mieux les visualiser, celles-ci sont synthétisées dans le diagramme de flow ci-dessous (Figure 6). Chacune de ces tâches est expliquée plus en détails en annexe 1.

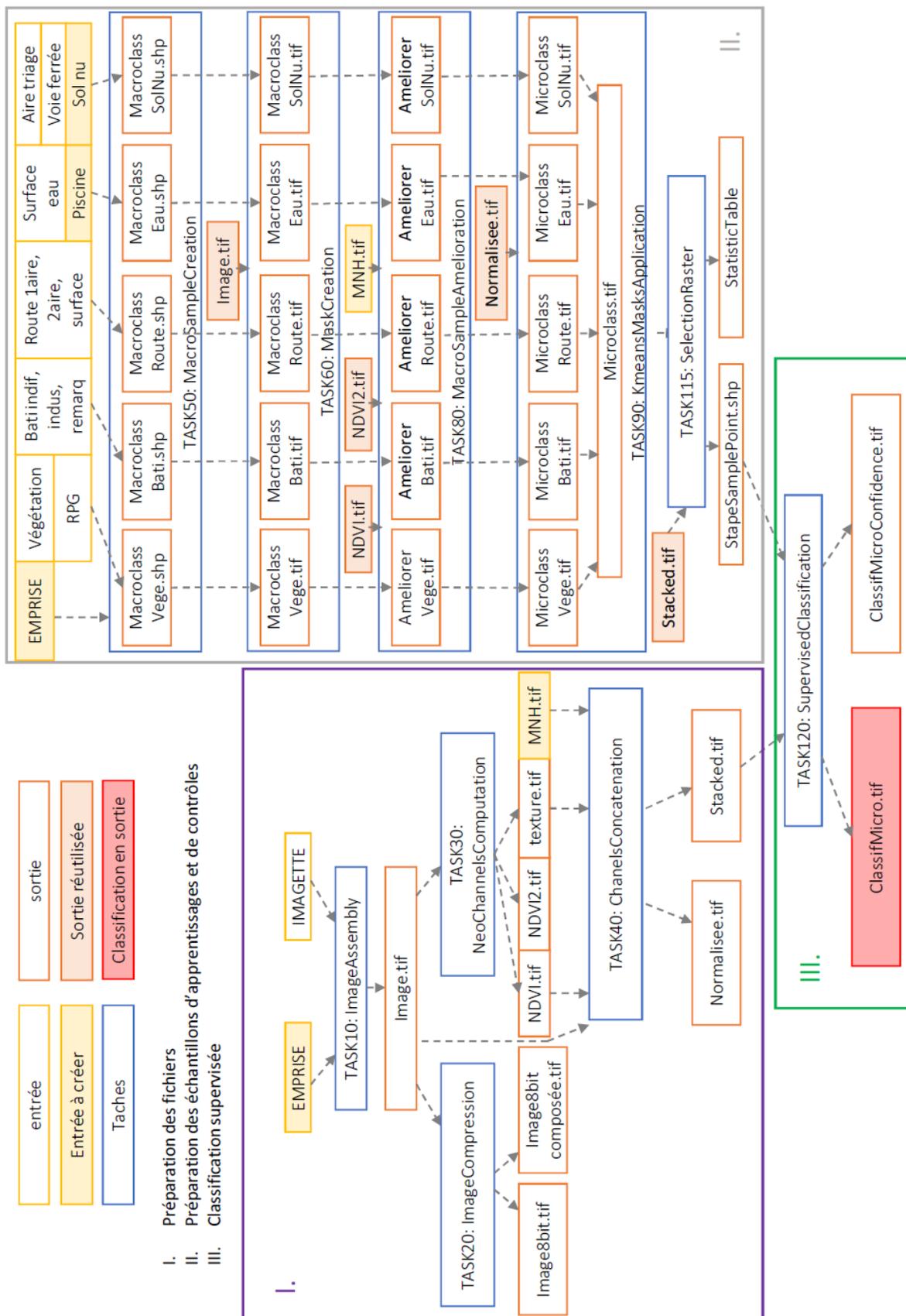


Figure 6: Description de la chaîne de traitement comprenant les différentes tâches utilisées (Celle-ci sont décrites en annexe 1).

L'exécution de la méthodologie de classification se divise en trois étapes principales (figure 6):

1. La création et la préparation des fichiers d'entrées, dont notamment l'emprise définissant la zone d'étude, le MNH et les calculs du NDVI, NDVI2 et de la texture. Ces différentes informations sont ensuite concaténées en une seule image.
2. La préparation des échantillons d'apprentissages (paragraphe V.3. Etape 2 et 3: Préparation des échantillons d'apprentissages et classification supervisée).
3. La classification supervisée, effectuée à partir de l'image concaténée ainsi que des points d'échantillonnages.

Pour effectuer toutes les tâches de la chaîne de traitement, il est nécessaire de créer en amont plusieurs fichiers d'entrées (figure 6):

- L'emprise de la zone d'étude,
- Une zone tampon de 1km autour de la zone d'étude, afin de limiter les effets de bords sur la classification,
- Le MNH, qui a nécessité des améliorations qui sont détaillées plus bas.
- Les différents shapefiles de la « Tâche50 », qui serviront à composer les différentes zones où seront par la suite sélectionnées les points d'échantillonnages.

2. Etape 1: La préparation des fichiers

La première étape consiste à assembler les bandes de l'image Pléiade avec différents néo-canaux dans un seul fichier tif afin d'obtenir un fichier tif concaténé, et un second fichier tif concaténé et normalisé. Les néo-canaux choisis par le CEREMA ont été confirmés lors de nombreuses études paramétriques (Joao, 2017) et sont (1) le NDVI et (2) la texture calculée par la tâche 30 à partir de l'image Pléiade (figure 6 et annexe 1). Pour obtenir ces fichiers de sortie, il faut également créer le modèle numérique de hauteur (MNH) qui sera également concaténé et normalisé avec les bandes de l'image et les néo-canaux.

Le Modèle Numérique de Hauteur (MNH) est calculé en soustrayant le Modèle Numérique de Terrain (MNT) au Modèle Numérique de Surface (MNS) (figure 7).

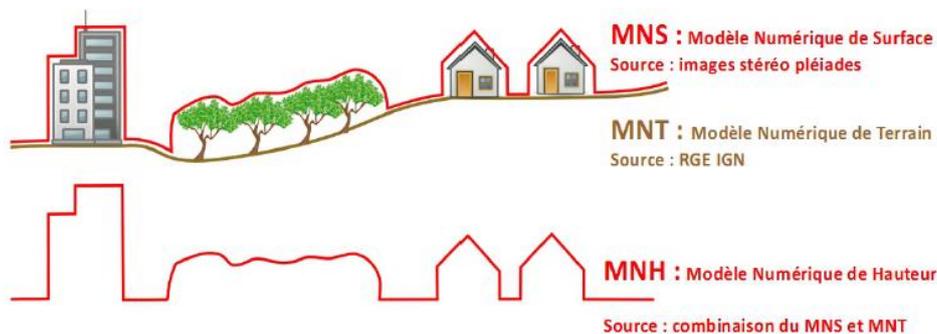


Figure 7:Création du MNH à partir du MNT et du MNS (source: Pôle de Compétences et d'Innovation Applications Satellitaires du CEREMA).

Au cours de mon stage, une grande partie des améliorations que j'ai effectuées sur la chaîne de traitement se situe au niveau de la création du MNH, que ce soit par (1) une validation de la méthodologie et des conseils à mettre en place lors de la création du MNH, (2) un choix des outils et des spécifications à utiliser qui a permis au CEREMA de coder une application permettant de créer le MNH, (3) la proposition d'un calcul de biais entre le MNS et MNT rencontré dans cet étude (spécifié par la suite).

Les trois principales étapes de la création du MNH comprenant la vérification du modèle numérique de terrain (MNT), la vérification du modèle numérique de surface (MNS) et la création du MNH, sont détaillées ci-dessous. Lors des précédentes études, certaines précautions méthodologiques n'étaient pas mises en place, comme la vérification de la qualité du MNT qui permet de calculer le MNH ou la prise en compte des zones sans données. Il s'est avéré nécessaire d'effectuer ces améliorations dans le cadre de cette étude, qui sont également détaillées ci-dessous.

- Le Modèle Numérique de Terrain (MNT)

La première étape consiste à vérifier la qualité du MNT. Sur la zone Bordeaux Métropole Ouest, il est de très bonne qualité puisque qu'il est principalement établi à partir de données lidars (figure 5). J'ai noté cependant que certains bâtiments visibles sur le MNT ne devraient pas l'être. Le MNT augmente légèrement en présence de certains bâtiments, ce qui peut amener à des artefacts de quelques centimètres. Ces faibles valeurs ne sont pas suffisantes pour être problématiques.

- Le Modèle Numérique de Surface (MNS)

La seconde étape consiste à vérifier la qualité du MNS. Pour cela, le MNS est comparé au MNT au niveau des sommets des segments des routes de la BDTopo®. Tout d'abord en utilisant l'outil "QualityMnsEstimation" mis en place par le CEREMA, qui va créer une couche points, dans laquelle chaque point correspond à la valeur du MNS et du NDVI. Par la suite l'outil "qgis points sampling tool" permet d'associer les valeurs du MNT aux points créés précédemment puis les points dont les valeurs du NDVI est inférieure à 0,3 (donc recouvert par de la végétation) sont supprimés. En comparant les valeurs du MNH et du MNT, on observe à partir de l'histogramme des valeurs, un biais de 0,8m, malgré que les valeurs soient prises au niveau des routes et qu'elles devraient par conséquent être identiques. Il est donc nécessaire de prendre en compte ce biais pour la création du MNH et de vérifier par la suite dans chaque nouvelle étude si un biais entre le MNS et le MNT existe également. Il reste cependant à démontrer par des études complémentaires la plus-value apportée par cette correction.

Le MNS de la zone de Bordeaux Ouest Métropole se caractérise également par la présence de pixels sans données, présentant des valeurs dites « nodata ». Pour la suite du traitement, il s'avère nécessaire d'interpoler la valeur de l'élévation de ces zones non corrélées. Au cours de ce travail, j'ai testé plusieurs outils, afin de conseiller les futurs utilisateurs et de proposer des pistes pour la création d'un script dans le futur (Figure 8):

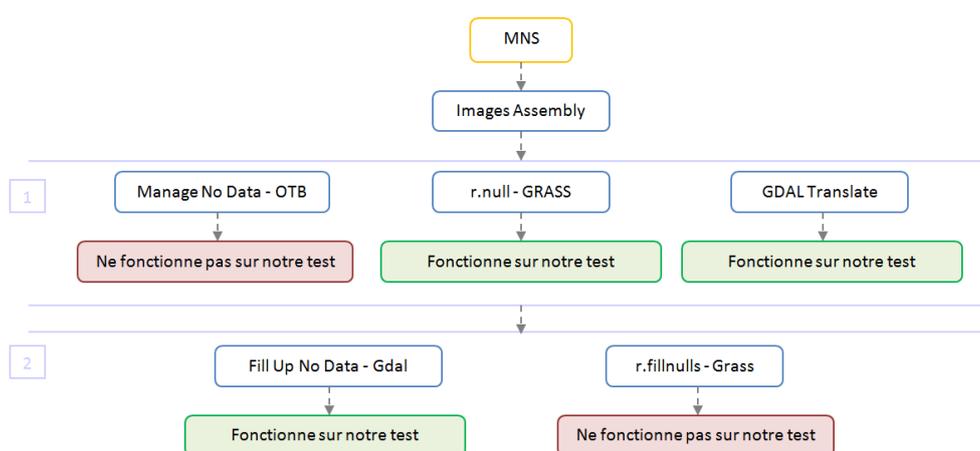


Figure 8: Choix des outils pour l'étape consistant à boucher les zones non corrélées. Pour cela il faut (1) définir/redéfinir les zones en no_data (qui correspondent à 0 sur le MNS), puis (2) combler ces zones par interpolation.

Lorsque deux applications amènent aux mêmes résultats, les applications GDAL sont privilégiées, puisqu'elles sont celles principalement utilisées par l'équipe CEREMA pour la création des scripts python.

- Modèle Numérique de Hauteur (MNH)

La dernière étape consiste à calculer le MNH à partir du MNS et du MNT, en prenant en compte le biais du MNS (Figure 9).

Au cours de cette étape de calcul du MNH, la définition des valeurs « no data » pose plusieurs problèmes :

- L'outil bandmath, entre autres, considère les données hors emprise (qui sont définies comme des valeurs « no data ») comme des valeurs égales à 0.
- Certaines valeurs d'élévations égales à 0 présentes dans la zone d'étude de Bordeaux Ouest Métropole sont considérées comme des valeurs « no data » par certains outils.

Pour pallier à ce problème, j'ai remplacé les valeurs égales 0 qui ne sont pas des « no data » par des valeurs à 0.0001. Puis j'ai découpé le raster par l'emprise de la zone d'étude définie précédemment. Ce traitement permet d'obtenir aucune valeur égale à zéro dans l'emprise, et de définir toutes les valeurs hors de l'emprise, égales à 0, comme des "nodata". Il faut cependant faire attention à l'utilisation de cette méthode et l'utiliser au cas par cas.

Après la création du MNH, il est possible de l'améliorer:

1. En remplaçant les valeurs négatives par des valeurs à 0,0001,
2. En creusant les canyons urbains. En effet, le MNH a tendance à diminuer très faiblement voire ne pas diminuer entre deux bâtiments proches. Pour pallier à cela, la superposition des routes (de la BDtopo®) à l'aide de bandmath permet de faire passer les routes à 0,0001 et donc diminuer la valeur du MNH entre deux bâtiments.
3. En comparant l'élévation donnée par le MNH et la BDtopo®, il est possible de détecter les bâtiments qui ne sont pas pris en compte dans le MNH et de les rajouter.

Toutes ces étapes de construction du MNH sont résumées dans la figure 9. Ce travail a également été une aide pour coder l'outil permettant de créer le MNH à partir du MNT et du MNS.

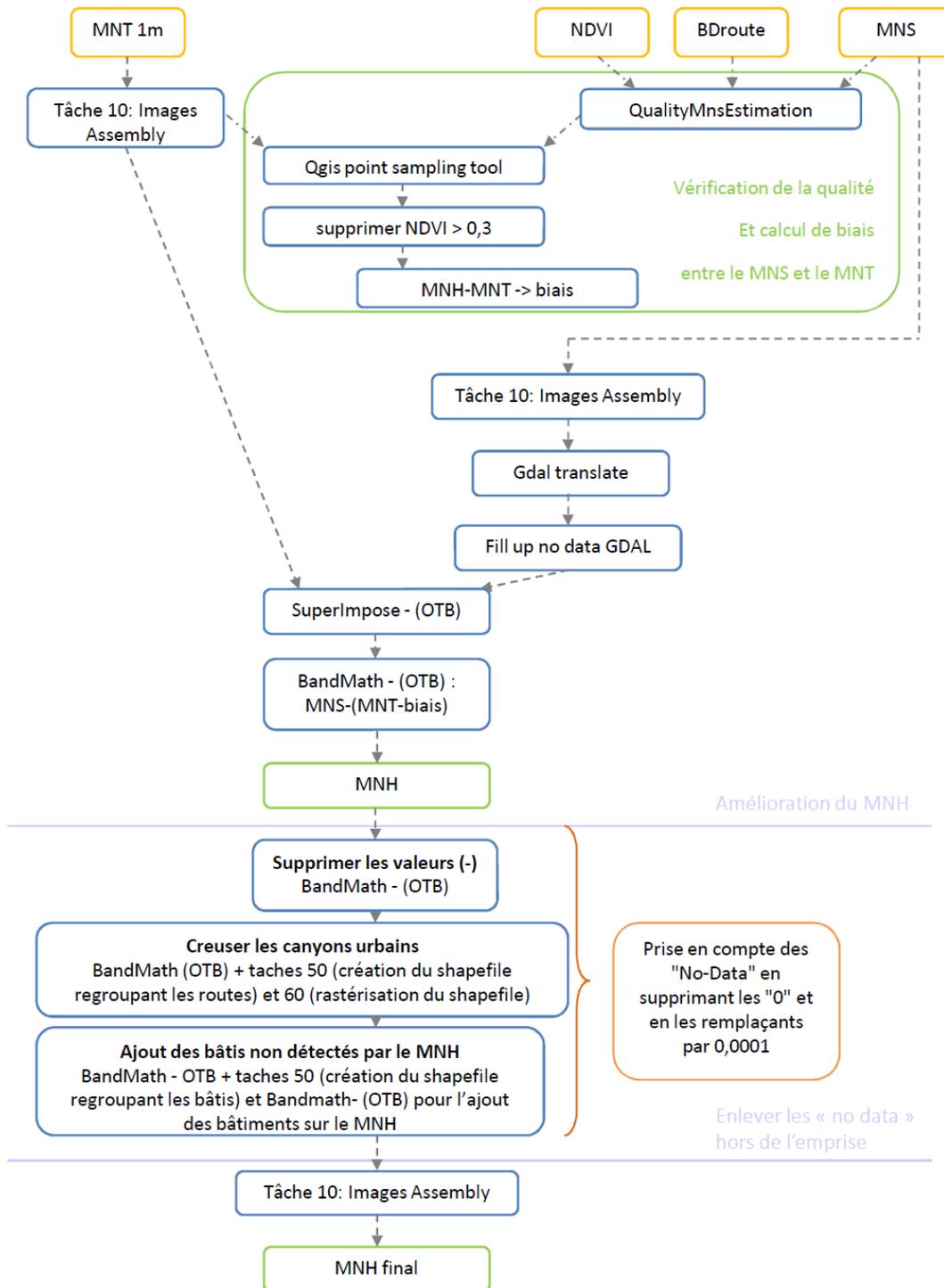


Figure 9: Diagramme de flot pour la création du MNH.

3. Etape 2 et 3: Préparation des échantillons d'apprentissages et classification supervisée

Cette étape a pour objectif d'améliorer les échantillons d'apprentissages en les rendant les plus conformes possibles à la date de l'image ou aux spécifications propres à la classification (comme par exemple ne pas prendre comme échantillons de route, des routes ou des arbres les

recouvrant). Pour cela, de très grandes surfaces d'échantillons d'apprentissages sont sélectionnées, permettant de représenter la diversité et l'hétérogénéité du paysage pour chaque classe.

Cette seconde étape correspond à l'exécution de la tâche 50 jusqu'à la tâche 120 de la méthodologie de classification brute (annexe 1 et figure 6).

Ces différentes tâches se divisent en deux étapes principales:

- La création d'échantillons d'apprentissages, à partir (1) de données exogènes et (2) des shapefiles que j'ai tracé pour cette étude. Ces shapefiles sont ensuite « rasterisés » et nettoyés à partir de données exogènes. Par exemple, toutes zones correspondant à de la végétation sont nettoyées en supprimant du raster les pixels avec un NDVI inférieur à 0,5. Les classes sont ensuite divisées en micro-classes, regroupées dans un fichier tif unique. Enfin un échantillonnage aléatoire de points est effectué sur des zones prédéfinies pour chaque classe (entre 0,2 et 3%). La spécificité de cette chaîne de traitement est la division des classes en micro-classes ayant des caractéristiques identiques. Au sein de chacune de ces classes, une division en plusieurs micro-classes ayant les mêmes particularités radiométriques et des valeurs du NDVI de la texture ou du MNH très proches est faite. Cela permet, lors des post-traitements, de traiter seulement les micro-classes qui posent problèmes (paragraphe V.4. Etape 4: Post-traitements).
- La deuxième étape est la classification supervisée correspondant à la tâche 120 (annexe 1 et figure 6). La classification est effectuée grâce à la méthode « Random Forest ». Une étude a été réalisée au CEREMA (MAO, 2014) pour tester différentes méthodes de classification ; il en résulte que la méthode « SVM » amène une plus-value lorsque l'échantillonnage est réduit. Lorsque le nombre d'échantillons est important, comme pour notre étude, les résultats sont proches mais la méthode SVM prend plus de temps de calcul. La méthode « Random Forest » est donc privilégiée dans notre cas.

Nous obtenons en résultats, une carte présentant la classification en raster (annexe 4), que l'on nomme « classification brute ».

On peut observer sur cette classification brute des anomalies dans les classes. Celles-ci peuvent être présentes sur une ou plusieurs classes ou seulement sur une ou plusieurs micro-classes. Elles sont corrigées en appliquant des séries de post-traitements.

4. Etape 4: Post-traitements

Une série de post-traitements a ensuite été appliquée afin d'améliorer la qualité de la classification (figure 10).

Pour certains post-traitements, il est plus simple de travailler en micro-classes. En effet dans certaines classes, il est possible de ne modifier que la micro-classes correspondant au problème rencontré. Par la suite, une deuxième étape consiste à faire des post-traitements sur les macro-classes.

Cette étude a nécessité de mettre en place de nouveaux post-traitements qui viennent s'ajouter aux post-traitements déjà existants (les post-traitements 2, 3 et 4: PT2, 3 et 4 sur la figure 10 et paragraphe la V.4.a. Micro-classes), ainsi que la modification de certains autres (les post-traitements 7 et 8: PT7 et 8 sur la figure 10 et paragraphe la V.4.b. Macro-classes). Ces post-traitements ont été mis en place afin de répondre aux problèmes spécifiques liées à l'étude Bordeaux Ouest Métropole et également d'améliorer la qualité de la classification en utilisant les outils (ou « tâches ») mises en place par le CEREMA. En effet, les post-traitements déjà définis ne répondaient pas à tous les problèmes que j'ai rencontré.

Les post-traitements que j'ai mis en place vont s'ajouter à ceux existant et vont permettre de créer un catalogue de post-traitements où un utilisateur pourra "piocher" les post-traitements dont il aura besoin en fonction de sa zone d'étude.

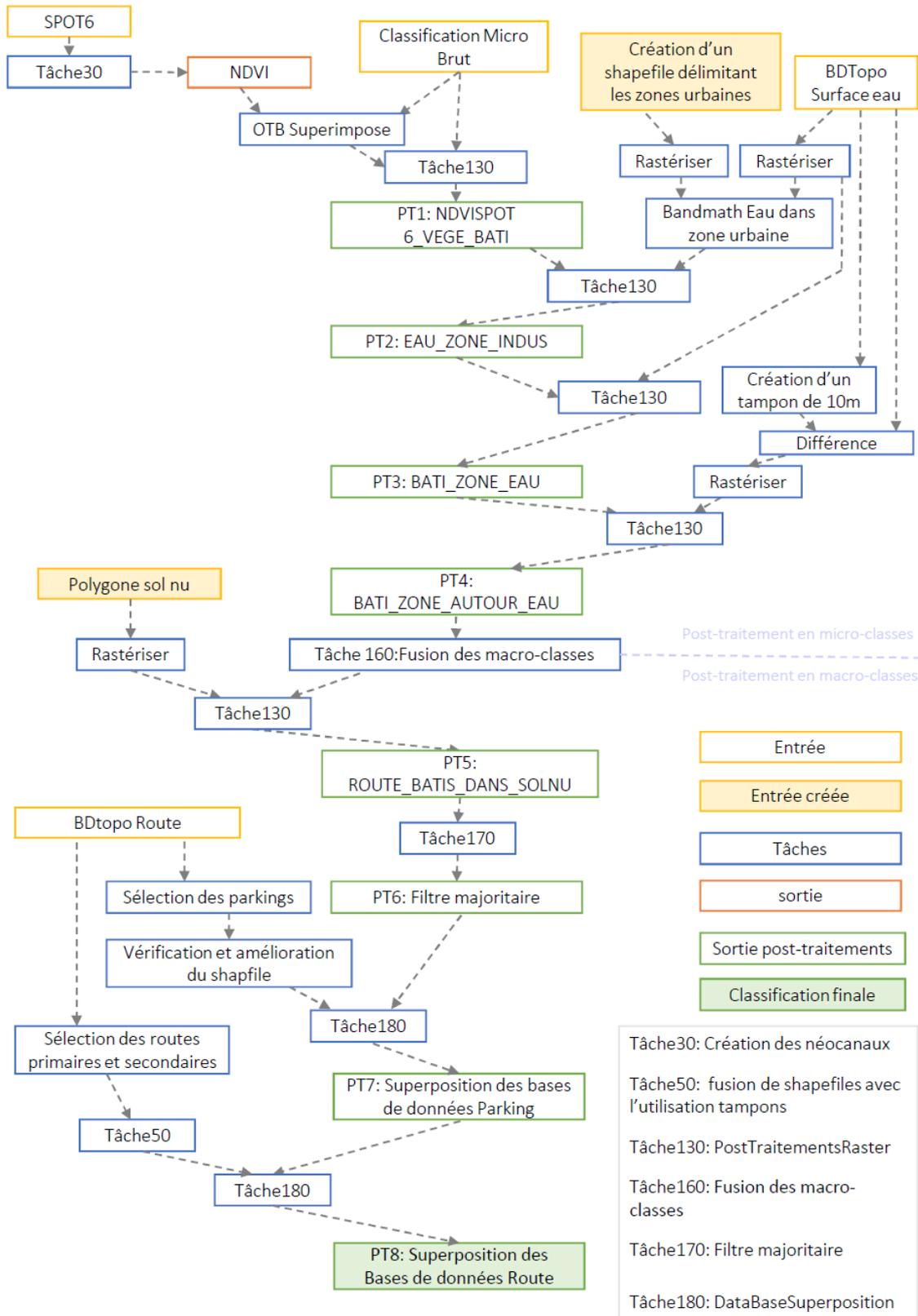


Figure 10: Diagramme de flot des post-traitements.

a. Micro-classes

- Post-traitement 1 : NDVISPOT 6_VEGE_BATI

Ce premier post-traitement est utilisé pour supprimer les pixels correspondant à de la végétation mais faussement interprétés comme des bâtiments dans la classification. Ces problèmes rencontrés dans la classification résultent d'arbre haut ayant un NDVI faible (à cause de la saison). On peut donc effectuer un seuillage du NDVI de spot 6.

Pour cela, il s'agit de définir les zones de bruits au sein de la végétation, définies comme des micro-classes de bâtiments fréquemment situées dans des classes de végétation. Pour l'étude Bordeaux Ouest Métropole, 5 micro-classes de bâtiments montrent cette particularité.

Sur notre zone d'étude, la première étape consiste à calculer le NDVI de l'image spot 6 de 2016 et d'appliquer un masque pour ne sélectionner que les valeurs du NDVI qui sont supérieures à 0,5. A partir de la tâche 130 (annexe 1 et figure 6), les pixels des micro-classes bâti identifiées comme des zones de bruits contenues dans les classes de végétation et localisées au sein du masque créé précédemment sont assimilées à une nouvelle micro-classes de végétation.

La figure 11 montre un exemple de résultat de ce post-traitement sur Bordeaux métropole avec le transfert de la classe bâti (en rouge sur la figure 11.b) en végétation (en vert sur la figure 11.c).

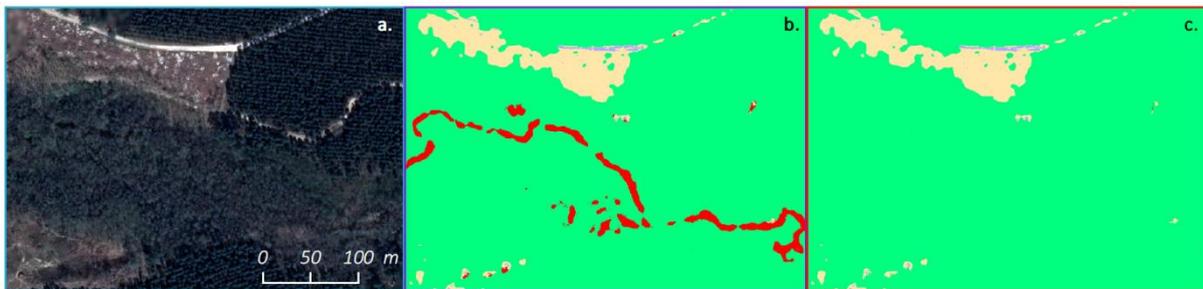


Figure 11: Exemple de résultat pour le post-traitement 1. a) Capture d'une image satellite Pléiade sur la zone Bordeaux Ouest, b) Capture de la classification brute sur la même zone, c) Capture de la classification après avoir appliqué le post-traitement 1. En rouge la classe bâtie, en vert la végétation, en jaune le sol nu et en violet les routes

- Post-traitement 2: EAU_ZONE_INDUS

Ce post traitement est utilisé pour supprimer les pixels faussement définis comme des classes « eau » à l'ombre des bâtiments (zones en bleues dans la figure 12b), le plus souvent sur des zones de routes ou des trottoirs. J'ai modifié ce post-traitement pour cette étude, en ajoutant (1) l'utilisation de la BDtopo® « surface en eau » pour délimiter les zones à traiter et (2) l'utilisation d'un masque pour effectuer ces traitements uniquement dans les zones avec beaucoup de bâtis, ce qui permet de ne pas supprimer les surfaces d'eau réel.

La première étape du traitement consiste à définir les micro-classes concernées par le problème. Sur la zone Bordeaux Ouest Métropole, elles sont au nombre de trois. J'ai ensuite créé un shapefile contenant les zones urbaines que j'ai rastérisé par la suite. Le post-traitement n'est appliqué que sur cette zone présélectionnée. La surface en eau définie par la BDTopo® est parallèlement ajoutée à la zone urbanisée, ce qui permet d'appliquer les modifications du post-traitement seulement sur les zones urbaines, hors des surfaces d'eaux réelles. Ces modifications consistent donc à remplacer, sur la zone d'étude Bordeaux Ouest Métropole, les 3 micro-classes où sont retrouvées des classes « eau » au sein d'ombres de bâtiment. Elles sont ensuite remplacées et insérées dans une nouvelle micro-classe (figure 12c), à l'aide de la tâche 130 (annexe 1 et figure 6).

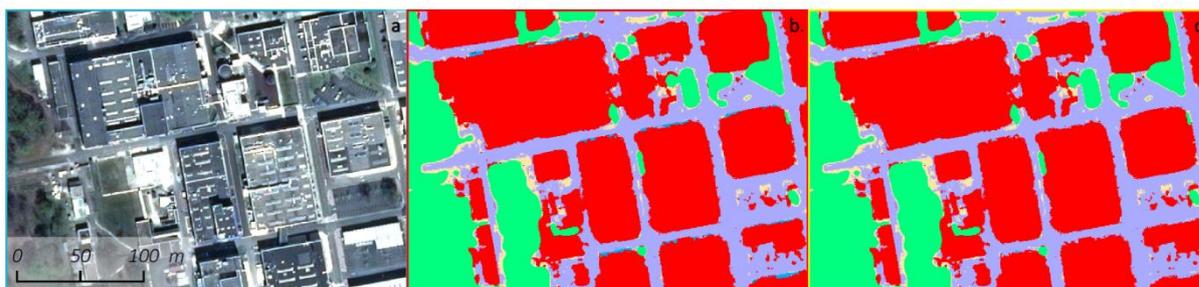


Figure 12: Exemple de résultat pour le post-traitement 2. a) Capture d'une image satellite Pléiade sur la zone Bordeaux Ouest, b) Capture de la classification brute sur la même zone, c) Capture de la classification après avoir appliqué le post-traitement 2. En rouge la classe bâtie, en vert la végétation, en jaune le sol nu, en violet les routes et en bleu les surface en eau

- Post-traitement 3 : BATI_ZONE_EAU

Ce troisième post-traitement que j'ai mis en place est utilisé pour nettoyer certaines zones de bâtis présentes dans des surfaces en eau, ce qui n'est pas le cas sur l'image Pléiade (Figure 13b). Après recherche des micro-classes de bâti posant problème (une seule dans le cas de Bordeaux Ouest Métropole), j'ai créé un raster des surfaces en eau à partir de la BDtopo®, puis les bâtiments de la micro-classe posant problème localisé dans ces zones sont transférées dans une nouvelle micro-classe « eau » (Figure 13c).

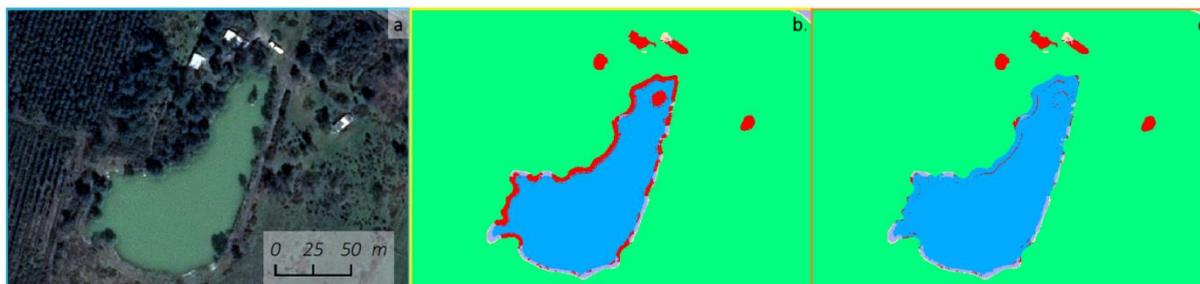


Figure 13: Exemple de résultat pour le post-traitement 3. a) Capture d'une image satellite Pléiade sur la zone Bordeaux Ouest, b) Capture de la classification brute sur la même zone, c) Capture de la classification après avoir appliqué le post-traitement 3. En rouge la classe bâtie, en vert la végétation, en violet les routes et en bleu les surface en eau

- Post-traitement 4 : BATI_ZONE_AUTOUR_EAU

Après avoir effectué le post-traitement 3, on remarque des éléments de bâtis aux alentours des surfaces en eau inexistant sur l'image Pléiade (Figure 14 a, b). J'ai mis en place un autre post-traitement pour cette étude afin de régler ce problème.

Ce post-traitement est identique au précédent, mais il est appliqué sur un buffer de 10m autour des surfaces en eau. La micro-classe à problème est ensuite transférée dans une nouvelle micro-classe « végétation » (figure 14c).

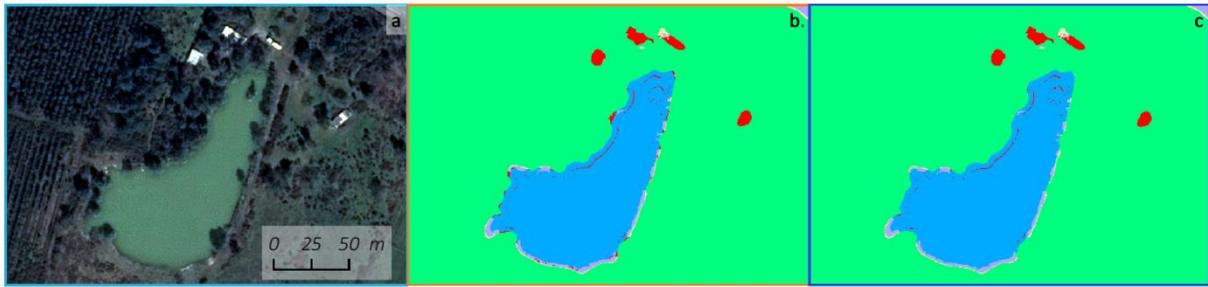


Figure 14: Exemple de résultat pour le post-traitement 4. a) Capture d'une image satellite Pléiade sur la zone Bordeaux Ouest, b) Capture de la classification brute sur la même zone, c) Capture de la classification après avoir appliqué le post-traitement 4. En rouge la classe bâtie, en vert la végétation, en violet les routes et en bleu les surface en eau

b. Macro-classes

Après avoir traité les micro-classes, les macro-classes ont été fusionnées pour permettre des modifications plus générales sur les problèmes rencontrés.

- Post-traitement 5 : ROUTE_BATIS_DANS_SOLNU

Ce post-traitement permet de nettoyer les zones de sols nus définies en zones de bâtis ou en zone de route (figure 15b). Cette opération est rapide lorsque les zones à problème sont étendues. Pour cela, des polygones englobant ces zones de sol nu sont créés manuellement en faisant attention à n'avoir que du sol nu (comme par exemple des champs). Dans ces polygones, tous les pixels définis comme des bâtis ou des routes sont transformés en sol nu (figure 15c).

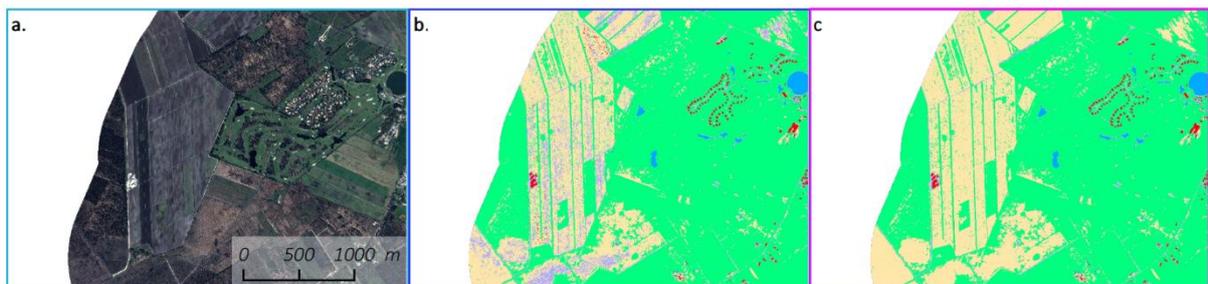


Figure 15: Exemple de résultat pour le post-traitement 5. a) Capture d'une image satellite Pléiade sur la zone Bordeaux Ouest, b) Capture de la classification brute sur la même zone, c) Capture de la classification après avoir appliqué le post-traitement 5. En rouge la classe bâtie, en vert la végétation, en violet les routes, en jaune les sols nus et en bleu les surface en eau.

- Post-traitement 6 : Filtre majoritaire

Le post traitement suivant est un filtre majoritaire, il permet de reclasser les pixels isolés en fonctions des classes qui l'entourent et donc de lisser la classification. Pour cela, j'ai utilisé la tâche 170 des traitements du CEREMA (figure 6 et annexe 1). Un exemple de résultat est montré en figure 16.



Figure 16: Exemple de résultat pour le post-traitement 6. a) Capture d'une image satellite Pléiade sur la zone Bordeaux Ouest, b) Capture de la classification brute sur la même zone, c) Capture de la classification après avoir appliqué le post-traitement 6. En rouge la classe bâtie, en vert la végétation, en violet les routes, en jaune les sols nus et en bleu les surface en eau.

- Post-traitement 7 : Superposition des Bases de données Parking

De nombreux artefacts sont présents sur les zones de parkings, concernant principalement la classe bâtie. En effet, nous retrouvons par exemple, le cas de véhicules qui sont pris pour des bâtiments. L'incrustation des parkings est un post-traitement rapide qui peut apporter des améliorations importantes sur la classification (figure 17). Un développement spécifique a été apporté pour cette étude : le traitement initial consistait à incruster les routes et les parkings en un seul traitement sans vérifications pour les parkings. Certains parkings sont cependant mal découpés sur la BDtopo®, tandis que d'autres sont situés dans des immeubles, faussant la classification puisque ceux-ci sont alors transformé en parking. Pour cette étude, j'ai décidé de diviser le post-traitement initial en deux étapes, l'amélioration sur les parkings (PT7) puis sur les routes (PT8). J'ai par la suite affiné les polygones représentant les parkings dans la BDtopo®. Les polygones correspondant aux pistes d'aéroports ont également été rajoutés car ils possèdent les mêmes caractéristiques que les parkings. En utilisant la tâche 180 des traitements du CEREMA (figure 6 et annexe 1), j'ai rajouté les polygones des parkings sur la classification.

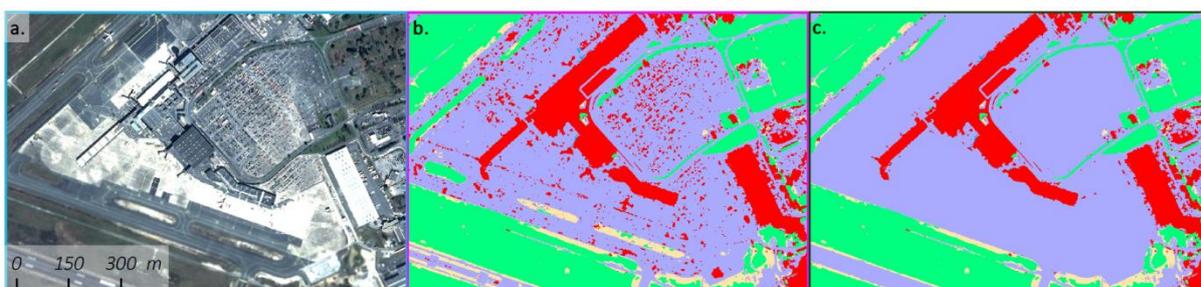


Figure 17: Exemple de résultat pour le post-traitement 7. a) Capture d'une image satellite Pléiade sur la zone Bordeaux Ouest, b) Capture de la classification brute sur la même zone, c) Capture de la classification après avoir appliqué le post-traitement 7. En rouge la classe bâtie, en vert la végétation, en violet les routes et en jaune le sol nu.

- Post traitement 8 : Superposition des Bases de données Route

Le dernier post-traitement consiste à ajouter les routes primaires et secondaires de la BDTopo®, sans prendre en compte la catégorie 'Route empierrée' et en appliquant un buffer négatif afin de ne pas rajouter des routes plus larges que dans la réalité. Elles sont insérées dans la classe « route » grâce à la tâche 180 des traitements du CEREMA (figure 6 et annexe 1). Ce post-traitement permet d'obtenir rapidement une classification plus propre (figure 18).



Figure 18: Exemple de résultat pour le post-traitement 8. a) Capture d'une image satellite Pléiade sur la zone Bordeaux Ouest, b) Capture de la classification brute sur la même zone, c) Capture de la classification après avoir appliqué le post-traitement 8. En rouge la classe bâtie, en vert la végétation, en violet les routes et en jaune le sol nu.

5. Résultats de la classification finale et apport de chaque post-traitements

La classification finale est obtenue après l'exécution de tous les post-traitements ; celle-ci est visible en annexe 5. L'utilisation de l'application « bandmath » d'OTB permet d'obtenir une carte des modifications dans la classification obtenue après l'exécution des post-traitements (annexe 6), afin de connaître l'apport de chaque post-traitements et la localisation des modifications réalisées. J'ai utilisé la formule suivante (tableau 2):

$im5b1!=im4b1?4:(im4b1!=im3b1?3:(im3b1!=im2b1?2:(im2b1!=im1b1?1:0)))$

Tableau 2: Formule utilisée sur Bandmath (OTB) afin de connaître l'apport et la localisation de chaque post-traitement.

Ces résultats m'ont également permis de construire un tableau montrant en nombre de pixels et en pourcentage l'apport des différents post-traitements (Tableau 3).

Post-traitements	1	2	3	4	5	6	7	8	Somme
	PTR1	PTR2	PTR3	PTR4	PTR5	MAJ1	DBS1	DBS2	
Nombre de pixels reclassés	3917726	306332	45780	57821	8408060	11263553	1029815	8989734	34018821
Pixels modifiés par rapport à la somme des pixels modifiés (%)	11.52	0.90	0.13	0.17	24.72	33.11	3.03	26.43	100
Pixels modifiés par rapport à la somme des pixels de la classification (%)	0.33	0.03	0.00	0.00	0.70	0.94	0.09	0.75	2.84

Tableau 3: Résultats de l'apport des différents post-traitements sur la classification.

Sur la zone Bordeaux Ouest Métropole, le nombre de pixels présents dans la classification et compris dans l'emprise est de 1 198 959 049, et le nombre de pixels modifiés est de 34 018 821, ce qui donne un pourcentage de pixels modifiés par les post-traitements dans la classification de 2.84%.

Ces tableaux soulignent l'important écart du nombre de modifications apportées par chaque post-traitements. Certains post-traitements apportent de fortes modifications (c'est le cas des post-traitements 5, 6 et 8, qui cumulent environ 83% des modifications à eux seuls) alors que d'autres beaucoup moins. On remarque cependant que certaines modifications mineures sont très visibles sur la classification (figure 12 et 13): le nettoyage par le post-traitement 2 des zones d'ombres des bâtiments dans les zones industrielles parfois considérées comme des zones d'eaux, par exemple. Ces améliorations apportent un confort visuel, mais elles ne sont pas essentielles pour un résultat quantitatif puisqu'elles ne représentent que 0,03% de pixels sur la classification totale.

6. Bilan

La chaîne de traitement mise en place par le CEREMA permet d'obtenir une classification propre (annexe 5). L'exécution de cette chaîne et des post-traitements qui suivent m'a permis de comprendre la méthodologie et de proposer des améliorations sur certains points : la mise en place d'un fichier XML réutilisable dans l'urbain, la vérification d'outil pour la création du MNH et le calcul du biais, l'apport d'un monitoring de tâche, la mise en place de nouveau post-traitement et la quantification de l'apport lié à ces post-traitement. Cela m'a également permis d'améliorer, de compléter et de créer de la documentation.

Cette étude sur la zone de Bordeaux Ouest Métropole montre que le développement de post-traitements spécifiques peut être nécessaire pour chaque nouvelle étude. On souligne donc l'importance de mettre en place un catalogue de post-traitements pour les regrouper, et ainsi permette aux futurs utilisateurs de la chaîne de traitement d'avoir à disposition de nombreux outils permettant le nettoyage de la classification.

Il ressort de ce travail qu'une question essentielle qui doit se poser en amont de l'exécution de la chaîne de traitement est celle de la précision que l'on cherche à atteindre pour la classification, et donc de jusqu'où les post-traitements doivent être poussés. En d'autres termes, quelle est la plus-value apportée par l'exécution des post-traitements mineurs par rapport au temps passé dessus dans un contexte de production. Pour chaque traitement, il faut ainsi prendre en compte le ratio temps vs apport de chaque post-traitement en fonction de l'utilisation faite de la classification par la suite, c'est-à-dire la qualité externe recherchée.

7. Plan d'assurance qualité

Afin d'obtenir la meilleure qualité dans les classifications obtenues et d'avoir une assurance de celle-ci (plan d'assurance qualité), plusieurs démarches sont mises en place, et correspondent à :

- La formation des opérateurs avant l'utilisation de la méthode de traitement.
- La mise en place d'une note descriptive de la procédure qui est améliorée continuellement, ainsi qu'un tableau décrivant les tâches et les paramètres à utiliser.
- La chaîne de traitement est automatisée au maximum afin d'éviter les erreurs, que ce soit de paramétrage, de nommage de fichier, ou encore de l'oubli d'une tâche sur la chaîne de traitement.
- La traçabilité de la procédure à travers l'archivage des paramètres utilisés lors de chaque étude à travers le fichier XML ou bien la fiche de suivi de l'étude complétée lors de chaque étude qui a été effectuée.

VI. Production qualité

Comme nous avons pu l'observer dans le paragraphe III.2. Sur l'état de l'art à propos de la Qualité, le contrôle de la qualité d'un produit permet de définir s'il est conforme à un besoin, permettant de sécuriser un utilisateur sur l'utilisation qu'il est possible de faire de la donnée.

Il existe plusieurs approches sur la qualité. On définit la qualité interne qui correspond à la qualité du produit obtenue par rapport "au cahier des charges" effectué en amont et la qualité externe correspondant à l'étude de la qualité du produit par rapport à l'usage qu'il est prévu d'en faire (figure 2).

Un des objectifs de ce stage est de mettre en place une méthodologie permettant de définir la qualité du produit avec pour but que cette méthode puisse être réutilisée par la suite.

Pour analyser et définir ces qualités (interne et externe), plusieurs critères sont définis dans la norme ISO 19157 (figure 3) (ORGANISATION INTERNATIONALE DE NORMALISATION, 2013; BONIN & GALLAIS, 2017). Pour cette étude, un des objectifs initiaux a été d'appliquer tous ces critères sur la classification

de la zone Bordeaux métropole. Cependant, j'ai rapidement observé que même si la norme semblait détaillée, il est vite devenu compliqué de l'appliquer. En effet le protocole est très généraliste, de plus la norme s'applique beaucoup plus aux données vecteurs que rasters. Il a donc été décidé de s'éloigner de la norme pour certains critères.

Un autre des problèmes rencontrés concerne le calcul des dynamiques urbaines au sein de Bordeaux métropole. En effet les critères présents ne permettent pas d'obtenir une idée sur la qualité et l'erreur que l'on peut obtenir entre deux classifications concernant l'évolution de la dynamique urbaine. J'ai donc créé pour cette étude un indicateur qualité qui permet de quantifier la qualité de dynamique urbaine entre deux classifications.

Le but de l'étude de ces différents critères de qualité est de savoir si la donnée que nous avons répond à l'usage qu'il est souhaité d'en faire (qualité externe) et d'obtenir une visualisation simple de ces résultats.

1. La cohérence logique

D'après la norme iso 19157 "La cohérence logique se définit comme le degré de cohérence interne des données selon des règles de modélisation et de spécification du produit" (BONIN & GALLAIS, 2017). Elle se décompose en quatre sous-critères:

- **La cohérence conceptuelle** qui se décrit comme "le respect du schéma conceptuel des données " (BONIN & GALLAIS, 2017).
- **La cohérence aux domaines de valeurs:** qui se décrit comme " l'appartenance des valeurs d'attributs aux plages de valeurs spécifiées" (BONIN & GALLAIS, 2017).
- **La cohérence du format** qui se décrit comme " l'adéquation avec la structure physique du jeu de données"(BONIN & GALLAIS, 2017).
- **La cohérence topologique** qui se décrit comme " l'exactitude des caractéristiques topologiques du jeu de données " (BONIN & GALLAIS, 2017).

Plus simplement, la cohérence logique correspond aux règles de structures auquel doit répondre le produit. On peut prendre comme exemple un élément qui doit être conforme à un domaine de valeurs, comme une éolienne qui ne peut pas mesurer plus de 200m, ou bien un fichier qui doit porter un nom particulier ou encore un format particulier. Ce sont également des règles de bon sens, comme par exemple le fait qu'une gare routière ne peut se retrouver dans un lac. Cela peut également correspondre pour des données raster à des amas de pixels devant atteindre une taille minimum (dans notre étude ce n'est pas le cas, nous n'avons pas fixé de taille minimum à un agrégat de pixels).

a. Méthodologie

Ce critère de qualité est normalement appliqué pour des données de type vecteurs, et il est donc difficile d'obtenir des résultats par rapport à la classification finale, après post-traitement, qui est en raster. J'ai choisi de répondre aux différents critères de cohérences par un critère booléen ('VRAI/FAUX'), par rapport à ce qui est voulu en amont de la chaîne de traitement.

Les critères sélectionnés sont (annexe 7 et figure 3):

- **La cohérence conceptuelle:** Pour des vecteurs cela correspond par exemple aux chevauchements entre les polygones. Ce critère n'est pas pris en compte car les données de classifications sont en raster, et il est difficile de trouver une application.
- **La cohérence aux domaines de valeurs:** Le CEREMA a défini 5 classes pour l'occupation du sol avec des codes correspondant :11100 Bâti, 11200 Route, 12200 Eau, 13000 Sol nu, 20000 Végétation. Pour valider ce critère, il faut que notre raster soit bien défini par ces codes.
- **La cohérence du format:** Pour valider ce critère, le nom du fichier doit être conforme, le type et format du raster doivent correspondre à un Geotiff au format Uint16et (nombre entier non signé de 16 bits) et dans le bon système de référence prédéfini.

- **La cohérence topologique.** Correspond pour des vecteurs à vérifier que les données ne se superposent pas. En raster, on peut la définir par le fait que les pixels soient bien définis dans toute la zone d'étude.

b. Résultats pour Bordeaux Métropole zone Est et Ouest

Type de cohérence logique	Règle	Vrai/Faux
La cohérence aux domaines de valeurs	Codes correspondent aux 5 classes prédéfinies	Vrai
La cohérence du format	Le nom du fichier doit être conforme: Il ne contient pas d'espaces (mais des "_") et de caractères spéciaux, il doit comprendre le lieu et la date de l'image Pléiade qui a servi pour la classification	Vrai
La cohérence du format	Le fichier doit être un Geotiff	Vrai
La cohérence du format	Le fichier doit être au format Uint16 (nombre entier non signé de seize bits)	Vrai
La cohérence du format	Respect du système de référence RGF93	Vrai
La cohérence topologique	Pixels bien définis dans toute la zone d'étude	Vrai

Tableau 4: Résultats obtenus pour la cohérence logique sur la zone de Bordeaux métropole.

c. Conclusions

La norme ISO19157 décrit la cohérence logique pour des fichiers vecteur plus que raster, ce qui la rend difficile à définir autrement que par des critères booléens. On note cependant que, par rapport à la qualité qui est recherchée pour cette chaîne de traitement (que toutes les règles soient respectées), les résultats sont bon. J'ai donc créé une notation simple pour pouvoir caractériser ce critère de qualité.

Note	Respect aux règles
5	Toutes les règles sont respectées
4	Une des règles n'est pas respectée
3	deux des règles ne sont respectées
2	Trois des règles ne sont respectées
1	Quatre ou plus des règles ne sont respectées

Tableau 5: Notation du critère de cohérence logique.

Pour la zone Bordeaux Métropole, la note attribuée est de 5/5.

2. L'exhaustivité

D'après la norme iso 19157 "l'exhaustivité se définit comme la présence et l'absence d'entités, de leurs attributs et de leurs relations" (BONIN & GALLAIS, 2017).

Elle se décompose en deux sous-critères :

- **Les excédents**, correspondant aux nombres d'éléments en excès dans un jeu de données.
- **Les omissions**, correspondant aux nombres d'éléments omis dans un jeu de données.

Afin d'avoir la possibilité de connaître les éléments omis ou en excès, il faut avoir une référence pour comparer nos données, que se soit une base de donnée, un contrôle terrain, ou encore des études d'expertises. De plus il est souvent nécessaire d'effectuer un échantillonnage car il peut s'avérer vite difficile voire impossible de vérifier tous les éléments, surtout dans le cas où la référence n'est pas une

base de données déjà construite. Cette échantillonnage doit prendre en compte un nombre d'élément assez important pour être représentatifs de la zone mais doit également prendre en compte tous les cas possibles dans la classification, comme les effets de bords par exemple (annexe 14 partie 2). On peut vérifier par exemple les gares recensées dans deux bases de données, dont une est la référence et définir les gares en excès ou manquantes sur la deuxième base de données.

a. Méthodologie

La méthodologie pour décrire l'exhaustivité de la classification suit les étapes suivantes (annexe 8):

- On effectue un échantillonnage probabiliste orienté surface (figure 19) sur la classification en utilisant la théorie binomiale de probabilité (FOODY, 2009; DE KEERSMAECKER, 1987) pour définir le nombre d'éléments à échantillonner.

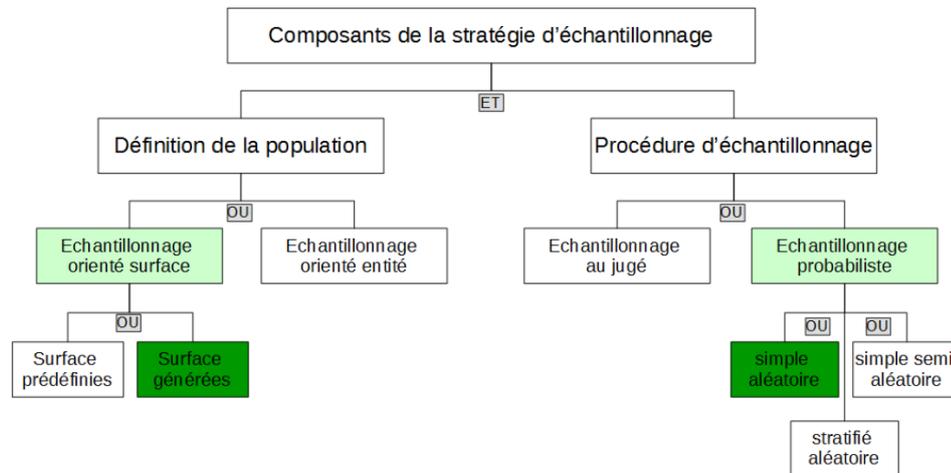


Figure 19: Choix effectués pour faire un échantillonnage (modifié à partir de la norme ISO19157)

Cette échantillonnage doit être réalisé sur l'ensemble de la classification et pas uniquement sur la classe bâtie, car il est nécessaire d'avoir des éléments présents dans la réalité (ou considéré comme tel) mais absent de la classification.

- On compare visuellement la classification finale et l'image Pléiade en utilisant des règles fixées pour être le plus objectif possible, comme par exemple le fait qu'un amas de pixels est considéré comme un bâtiment s'il mesure plus de 10m² (40pixels). On note que la comparaison s'effectue sur des éléments bâtis présents ou absents de la classification et non sur des pixels (annexe 8 partie 2. Comparaison visuelle).
- On calcule le taux de bon classement et le kappa.
- On calcule le nombre d'éléments en excès et omis.
- On calcule le taux d'éléments en excès et omis.
- On calcule le taux d'exhaustivité.

Afin de vérifier la fiabilité de l'échantillonnage et du nombre d'éléments échantillonnés, cette méthodologie est effectuée deux fois.

b. Résultats pour Bordeaux Métropole zone Est et Ouest

Le résultat est identique pour la classification brute et la classification après post-traitement. Il n'y a donc qu'un seul résultat pour les deux classifications.

Résultat 1: effectué sur 190 éléments (tableau6)

		Le réel	
		Bâti	Autre
Classification	Bâti	33	0
	Autre	1	156

Tableau 6: Résultats obtenus lors d'une première comparaison des bâtiments entre l'image pléiade et la classification.

Le tableau 6 présente les comparaisons entre la classification et l'image Pléiade. On peut donc observer 33 bâtiments sur l'image Pléiade sont bien classés comme bâtiments. Un bâtiment présent sur l'image Pléiade n'est pourtant pas défini comme tel dans la classification (figure 20). Les 156 échantillons restant qui ne sont pas des bâtiments sur l'image Pléiade, sont bien classés comme une des quatre classes autre que le bâtiment.

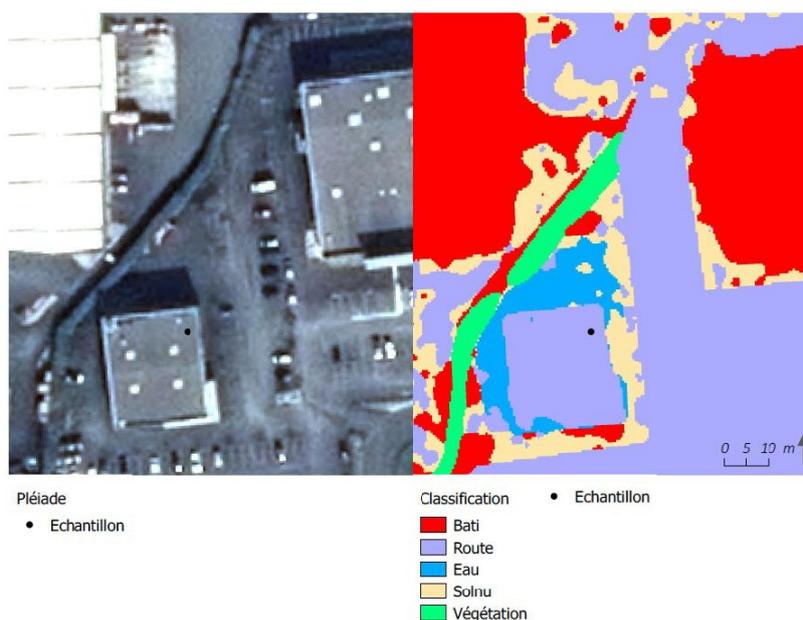


Figure 20: Exemple de comparaison entre l'image Pléiade et la classification (ici bâti présent sur l'image Pléiade et absent de la classification).

A partir du tableau 6, on obtient un coefficient de kappa de 0,98 et un taux de bon classement de 99,5%. Par la suite j'ai calculé le nombre d'éléments en excès et omis ainsi que le taux d'éléments en excès (tableau 7). Pour cela il faut prendre en compte le nombre de bâtiments présents dans l'image Pléiade et présents ou non dans la classification, cela correspond aux vrais positifs (33 éléments) et faux positifs (0 élément). Ce qui donne un nombre d'élément de 33.

Excédent	Nombre d'éléments en excès	0
	Taux d'éléments en excès	0%
Omission	Nombre d'éléments omis	1
	Taux d'éléments omis	1/33 = 3%

Tableau 7 : Éléments en excès et omis obtenu lors d'une première comparaison des bâtiments entre l'image pléiade et la classification.

Cela m'a permis de calculer un taux d'exhaustivité de **97% ± 6%** (annexe 8 partie 3. Résultats).
Résultat 2: effectué sur 186 éléments (tableau 8)

		Le réel	
		Bâti	Autre
Classification	Bâti	25	2
	Autre	0	158

Tableau 8: Résultats obtenus lors d'une deuxième comparaison des bâtiments entre l'image pléiade et la classification.

Le tableau 8 présente les comparaisons entre la classification l'image Pléiade. On peut donc observer que 25 bâtiments observés sur l'image Pléiade sont bien classés comme bâtiments. Deux bâtiments présent sur la classification est pourtant absents sur l'image Pléiade. Les 158 échantillons restant qui ne sont pas des bâtiments sur l'image Pléiade, sont bien classés comme une des quatre classes autre que le bâtiment.

A partir du tableau 8, on obtient un coefficient de kappa de 0,93 et un taux de bon classement de 98,4%. Par la suite j'ai calculé le nombre d'élément en excès et omis ainsi que le taux d'éléments en excès (tableau 9). En prenant en compte les vrais positifs (25 éléments) et faux positifs (2 éléments) présents sur la classification. Ce qui donne un nombre d'élément de 28.

Excédent	Nombre d'éléments en excès	2
	Taux d'éléments en excès	$2/28 = 7,14\%$
Omission	Nombre d'éléments omis	0
	Taux d'éléments omis	0%

Tableau 9: Eléments en excès et omis obtenus lors d'une deuxième comparaison des bâtiments entre l'image pléiade et la classification.

Cela m'a permis de calculer un taux d'exhaustivité de **92% ± 10%** (annexe 8 partie 3. Résultats).

Les deux tests effectués sont donc cohérents. Pour avoir un résultat encore plus précis, il est possible d'ajouter ces résultats.

Résultat final: effectué sur 376 éléments (tableau 10)

		Le réel	
		Bâti	Autre
Classification	Bâti	59	2
	Autre	1	314

Tableau 10: Résultats obtenus lors de l'ajout des deux comparaisons des bâtiments entre l'image pléiade et la classification.

On peut observer à partir du tableau 10, 59 bâtiments présents sur l'image Pléiade qui sont bien classés comme bâtiments. Un bâtiment présent sur l'image Pléiade n'est pourtant pas défini comme tel dans la classification. Deux bâtiments présents sur la classification sont pourtant absents sur l'image Pléiade. Les 314 échantillons restant qui ne sont pas des bâtiments sur l'image Pléiade, sont bien classés en une des quatre classes autre que le bâtiment.

On obtient un coefficient de kappa de 0,97 et un taux de bon classement de 99.2%.

Par la suite j'ai calculé le nombre d'élément en excès et omis ainsi que le taux d'éléments en excès (tableau 11). En prenant en compte les vrais positifs (58 éléments) et faux positifs (2 éléments) présents dans la classification, ce qui donne un nombre d'élément de 61.

Excédent	Nombre d'éléments en excès	2
	Taux d'éléments en excès	2/61 = 3,27%
Omission	Nombre d'éléments omis	1
	Taux d'éléments omis	1/61 = 1,6%

Tableau 11 : Eléments en excès et omis obtenus lors de l'ajout des deux comparaisons des bâtiments entre l'image pléiade et la classification.

Cela m'a permis de calculer un taux d'exhaustivité de **95± 5%** (annexe 8 partie 3. Résultats).

c. Conclusions

Afin de pouvoir noter facilement ce taux d'exhaustivité, j'ai utilisé la notation tirée de la fiche "Critère d'exhaustivité" de la série de fiches « Qualifier les données géographiques » du CEREMA, 2017 (BONIN & GALLAIS, 2017).

Note	Taux d'exhaustivité
5	De 95 % à 100 %
4	De 90 % à 95 %
3	De 75 % à 90 %
2	De 50 % à 75 %
1	Taux < 50 %

Tableau 12: Notation du critère d'exhaustivité.

On obtient donc ici une note de 5 car nous avons un taux d'exhaustivité 95± 5%.

Le critère d'exhaustivité pour la zone Bordeaux Métropole donne donc de bons résultats. Cela signifie que les éléments de bâtis sont très bien pris en compte dans la classification (on note que l'on parle bien d'éléments de bâti et non de pixels de classe « bâti »). On confirme donc que la construction de nouveaux bâtiments a de forte chance d'être bien pris en compte sur des classifications effectuées entre deux dates différentes. La figure 21 illustre cela sur un cas local sur la zone Bordeaux Métropole. On montre la différence entre la classification réalisée sur l'image Pléiade 2014 et l'image Pléiade 2016. On observe que les nouveaux bâtiments sont bien présents sur la classification de 2016.

Résultat des classifications à partir d'images Pléiade (2014-2016) lors de la construction d'un bâtiment

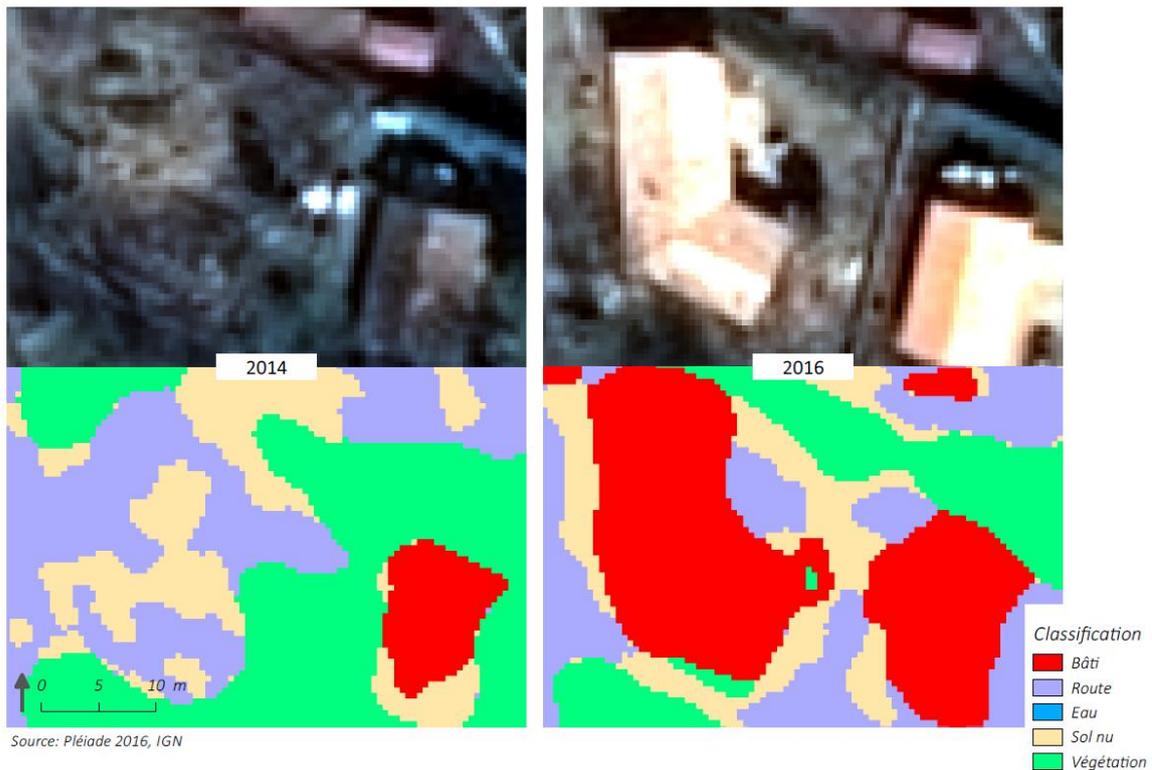


Figure 21: Résultats obtenus sur des classifications entre 2014 et 2016 lors de la construction d'un bâtiment.

3. La précision thématique/sémantique

D'après la norme iso 19157 "la précision thématique se définit comme la précision des attributs quantitatifs, et la justesse des attributs non quantitatifs, du classement des entités et de leurs relations (Bonin & Gallais, 2017).

Elle se décompose en trois sous-critères :

- **Justesse de classement:** correspond à la "comparaison des classes attribuées aux entités ou à leurs attributs par rapport à l'univers du discours (par exemple, monde réel ou données de référence)" (Bonin & Gallais, 2017).
- **Justesse des attributs non quantitatifs:** "mesure permettant d'établir si les valeurs d'un attribut non quantitatif sont correctes ou pas" (Bonin & Gallais, 2017).
- **Précision des attributs quantitatifs:** correspond à la "proximité de la valeur d'un attribut par rapport à la valeur vraie ou reconnue comme vraie" (Bonin & Gallais, 2017).

Plus simplement, la cohérence thématique correspond à la conformité des valeurs des éléments du jeu de données par rapport à un référentiel.

a. Méthodologie

- **Justesse de classement**

La méthodologie pour décrire ce critère de qualité est la suivante (annexe 9):

- On effectue un échantillonnage probabiliste orienté surface (figure 19) sur la classification en utilisant la théorie multinomiale de probabilité (De Keersmaecker, 1987) pour définir le nombre

d'éléments à échantillonner afin d'obtenir les éléments qui seront considérés comme la référence.

- On définit par la suite visuellement la classe auquel appartient chaque point des échantillons sur l'image Pléiade en utilisant des règles fixées pour être le plus objectif possible (annexe 9). On peut citer, par exemple, l'utilisation du NDVI ou l'utilisation d'images satellite additionnelles (Google Satellite, par exemple) pour mieux définir la forme des objets (il faut faire attention, ce n'est pas pour connaître leurs positions, mais seulement avoir une idée plus précise de la forme des éléments). Il est fortement conseillé, voir indispensable, que la caractérisation des points d'échantillonnages soit effectuée par plusieurs personnes pour effectuer une comparaison la plus objective possible. En effet, dans de nombreux cas, il est très difficile de définir visuellement la classe du pixel (figure 22), surtout pour les limites de bâtiments. Pour l'étude Bordeaux Métropole, nous avons choisi un nombre de trois opérateurs pour effectuer les interprétations visuelles, puis sur les pixels où nous étions en désaccord, de débattre ensemble du résultat. Après une étude faite sur la qualité de l'interprétation des échantillons par différents opérateurs, on remarque une différence sur les premières interprétations pouvant atteindre 12% des pixels (annexe 14 partie 1). Cela confirme l'importance d'être plusieurs opérateurs afin d'effectuer cette analyse.



Figure 22: Exemple de difficulté d'interprétation de la classe présente sur l'image Pléiade. Avec ici (a) l'image Pléiade comparée à (b) une image tirée de google satellite permettant d'obtenir en (c) une idée plus précise de la forme des bâtiments et ainsi de les visualiser sur l'image Pléiade.

- On calcule par la suite la matrice de confusion entre l'interprétation faite considérée comme la référence et les résultats de la classification pour connaître les éléments bien classés de la classification.
- On calcule la matrice relative de classement.
- On calcule un taux de bon classement.

- On calcule le coefficient de Kappa.
- **Justesse des attributs non quantitatifs**

La classification ne possède pas d'attributs non quantitatifs, il n'est donc pas possible de décrire ce critère.

- **Précision des attributs quantitatifs**

Pour décrire ce critère de qualité, j'ai décidé de comparer la surface de bâti de la classification avec celle de la BDtopo® (car considérée la plus proche de la réalité) à travers des grilles dont la longueur des côtés des carrés peuvent varier. Cette opération est réalisée sur la classification brute mais également sur la classification finale, pour quantifier l'évolution de la qualité de la classification avant et après post-traitement. Après plusieurs tests, le choix final le plus pertinent pour la grille est défini, en essayant de conserver des grilles ayant des carrés assez petits pour obtenir des résultats visibles tout en minimisant les temps de calculs. Ce choix correspond à une grille de 500m de côté.

La méthodologie pour décrire ce critère de qualité est la suivante (annexe 10):

- Avoir une classification en macro-classe avec l'emprise voulue.
- Fusionner les couches de la BDtopo® E_bâti, comprenant le Bâti_indifférencié, le Bâti_industriel et le Bâti_remarquable.
- Faire un « dissolve » des entités de la couche BDtopo® bâti créée précédemment.
- Faire un découpage par la même emprise que celle utilisée pour la classification sur la couche de BDtopo® Bati
- Utiliser l'outil de comparaison vecteur/raster « QualityClassificationGridQualification.py », mis en place par le CEREMA) qui permet à partir de la couche de la BDtopo® et du raster de la classification de définir pour chaque carré de la grille la surface de bâtiments dans la classification ainsi que la surface de bâtiments dans la BDtopo®. A partir de cela il est possible de comparer ces valeurs.

b. Résultats pour Bordeaux Métropole zone Est et Ouest

- **Justesse de classement**

A partir des résultats obtenus par interprétation visuelle, j'ai réalisé une matrice de confusion avec en entrée les résultats des interprétations et la classification brute et finale, afin notamment de quantifier les améliorations sur la qualité réalisée par l'exécution des post-traitements. Cette étude a donc été effectuée sur la classification brute et la classification finale. Seule la classification finale est présentée dans le rapport (tableau 15).

		Classification						
		Bâti	Route	Eau	Sol nu	Végétation		
Référence	Bâti	62	1	0	0	7	70	10,26 %
	Route	19	55	0	10	11	95	13,93 %
	Eau	0	1	31	1	2	35	5,13 %
	Sol nu	4	8	1	24	7	44	6,45 %
	Végétation	4	11	1	7	415	438	64,22 %
		89	76	33	42	442		
		13,05 %	11,14 %	4,84 %	6,16 %	64,81 %		

Tableau 13: Matrice de confusion entre la classification finale et les résultats de l'interprétation sur l'image Pléiade.

A partir du tableau 13, il est possible de calculer le pourcentage de pixels bien classé qui est ici de 86 % ainsi que le coefficient de kappa qui est de 0,75.

A partir de cette matrice, on peut calculer le pourcentage d'éléments en excès, omis, mais également le nombre d'éléments qui se compensent en prenant en compte le pourcentage de chaque classe (tableau 14).

Par exemple, sur la classe végétation qui représente 64,22% de la classification totale, nous avons 60,85% sur les 64,22% qui sont des données correctes, c'est à dire de pixels qui correspondent entre la classification et la référence. Nous avons ensuite les éléments en excès (de la végétation dans la classification mais une autre classe dans la référence) ou omis (de la végétation dans la référence mais une autre classe dans la classification).

A partir de ces résultats, on calcule les erreurs qui se compensent, c'est à dire le pourcentage de pixels en excès compensé par le pourcentage des pixels omis (3,37%) (figure 6 de l'annexe 9). Si on calcule les excédents et omissions restants en ne prenant pas en compte le pourcentage de pixels compensées, on obtient un pourcentage d'excédents restant de 0,59% (3,96%-3.37%) et un pourcentage de déficit restant de 0% (3,37%-3.37%).

	données correctes (%)	données compensées/ contrebalancés (%)	Total des excédents (%)	total des déficits (%)	Excédents restant (%)	déficits restant (%)
Bâti	9.09	1.17	3.96	1.1	2.79	0
Route	8.06	3.08	3.08	5.87	0	2.79
Eau	4.55	0.29	0.29	0.59	0	0.29
Sol nu	3.52	2.64	2.64	2.93	0	0.29
Végétation	60.85	3.37	3.96	3.37	0.59	0

Tableau 14 : Pourcentage d'éléments en excès et omis pour chaque classe de la classification finale, calculés à partir de la matrice de confusion.

La Figure 23 est un diagramme qui permet de représenter les résultats du tableau 14. Il permet de visualiser le nombre d'éléments corrects, d'éléments en excès ainsi que les éléments omis.

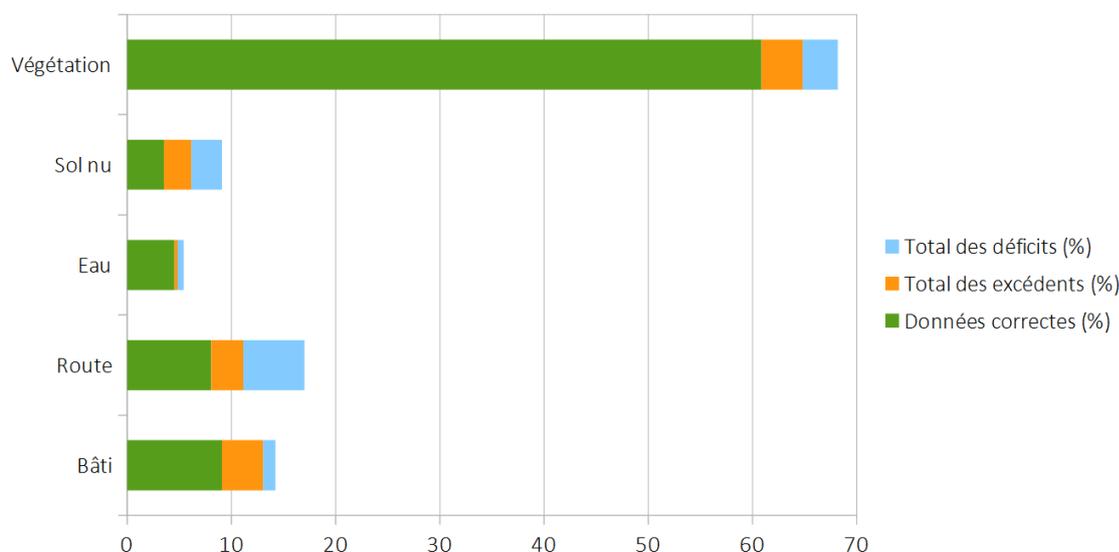


Figure 23: Représentation du nombre d'éléments en déficits, en excédents ou corrects pour chaque classe de la classification finale, calculé à partir de la matrice de confusion.

La Figure 24 est une autre représentation du tableau 14 qui permet de montrer le nombre d'éléments corrects, mais également le nombre d'éléments qui se compensent (figure 6 de l'annexe 9) et le nombre d'éléments en excès ou omis qui n'est pas compensé (figure 24).

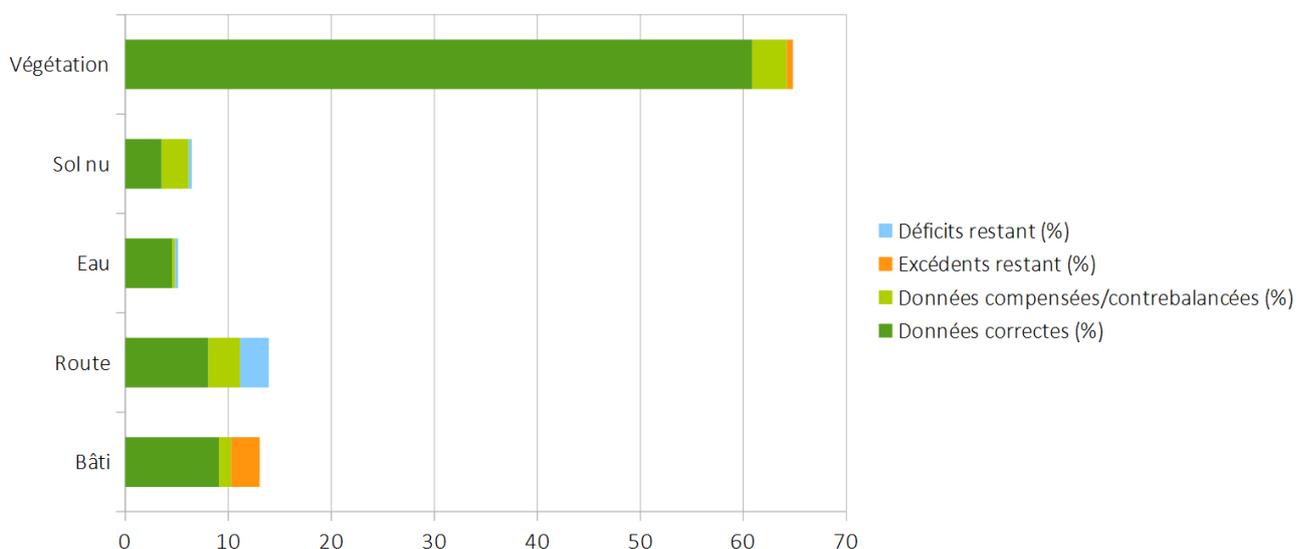


Figure 24: Représentation du nombre d'éléments en déficits, en excédents, corrects et qui se compensent pour chaque classe de la classification finale, calculé à partir de la matrice de confusion.

Cette étude a également été appliquée sur la classification brute. En comparant les résultats, la justesse de classement souligne l'amélioration du nombre d'éléments bien classés et du kappa apporté par l'exécution des post-traitements (Tableau 15).

	Classification brute	Classification finale
Pourcentage d'éléments bien classés	84%	86%
Kappa	0.71	0.75

Tableau 15: Comparaison du nombre d'élément bien classé et du Kappa entre la classification finale et la classification brute.:

- Précision des attributs quantitatifs

Les résultats de la précision des attributs quantitatifs montrent que nous avons une surface de bâti importante au niveau de la classification brute et finale mais celle-ci est très élevée par rapport à la Bdto[®] (figure 25).

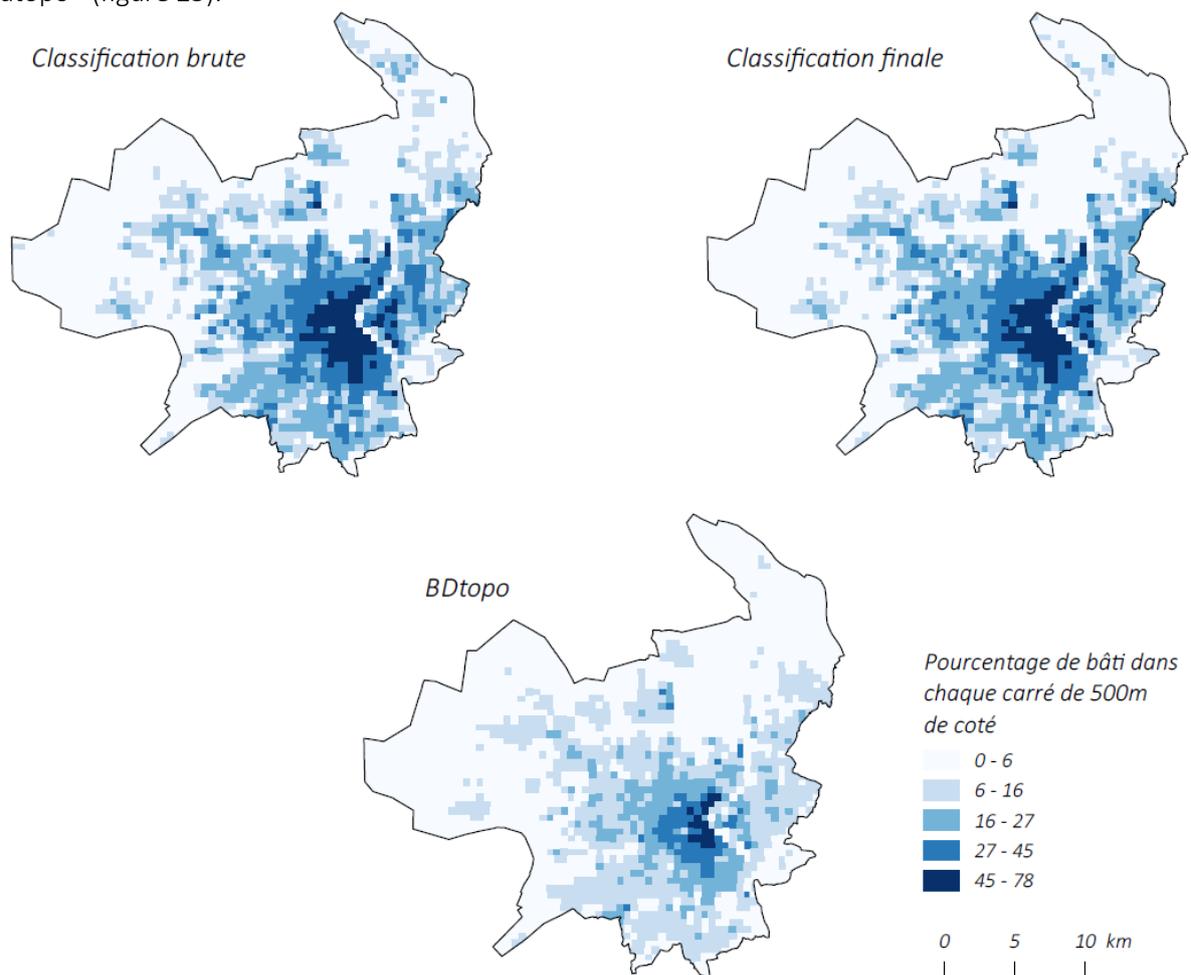


Figure 25 : Pourcentage de bâti sur Bordeaux métropole découpé en des zones de 500m2 réalisé à partir de la classification brute, finale et de la BDtopo.

Ces résultats présentent une surestimation du bâti sur la classification par rapport à la BDTopo[®], légèrement diminuée par les post traitement mais qui reste importante.

On note que cette surestimation peut être légèrement plus importante que la réalité, car bien que la BDtopo[®] soit considérée le plus proche de la réalité, elle est légèrement sous-estimée par rapport à ce qui est observée sur l'image Pléiade (figure 26). Cela provient du fait que la BDtopo[®] n'est pas forcément une référence d'un point de vue millésime.



Figure 26: Comparaison entre les bâtiments de la BDtopo et ceux de l'image pléiade.

c. Conclusions

Afin de noter facilement la précision thématique, j'ai utilisé la notation tirée de la fiche "critère de précision thématique" de la série de fiche "qualifier les données géographiques" du CEREMA 2017 (Bonin & Gallais, 2017) qui utilise le taux de valeurs d'attributs corrects.

Note	Classification
5	De 95 % à 100 %
4	De 90 % à 95 %
3	De 75 % à 90 %
2	De 50 % à 75 %
1	Taux < 50 %

Tableau 16: Notation du critère de précision thématique.

On obtient donc ici une note de 3. On observe en effet un certain nombre de pixels mal classés. Cette note peut être suffisante ; Elle dépend de la qualité externe souhaitée.

4. La précision de position

D'après la norme iso 19157 La précision de position "se définit comme la précision de la position des entités au sein d'un système de référence spatiale" (Bonin & Gallais, 2017). Elle se décompose en deux sous-critères:

- **La précision absolue ou externe** définit comme "la proximité des valeurs de coordonnées reportées par rapport aux valeurs vraies ou reconnues en tant que telles" (Bonin & Gallais, 2017).
- **La précision relative ou interne** se définit comme la "proximité des positions relatives des entités dans un jeu de données par rapport à leurs positions relatives respectives vraies ou reconnues

en tant que telles" (Bonin & Gallais, 2017). Ce deuxième sous-critère ne sera pas abordé dans ce rapport car moins significatif que le premier.

a. Méthodologie

La méthodologie pour décrire ce critère de qualité est la suivante (annexe 11):

- Echantillonnage probabiliste orienté surface (figure 19) sur la classification en utilisant la théorie binomiale de probabilité (Foody, 2009; De Keersmaecker, 1987). Cette échantillonnage sera effectué seulement sur la classe bâtie.
- Comparaison visuelle entre la classification et l'image Pléiade sur la classe bâtie afin de définir la distance les séparant. Pour cela, on utilise des règles pour être le moins subjectif possible. Par exemple, le coin de bâtiment le plus proche de chaque point d'échantillonnage sur l'image Pléiade sera comparé au coin de bâtiment correspondant sur la classification. On aura ainsi la distance entre la classification et l'image Pléiade concernant la classe bâtie.
- Calcul d'incertitudes de position.
- Définir une valeur moyenne des incertitudes.
- Définir une erreur quadratique moyenne.

Nous dérogeons ici à la définition de la norme. En effet, nous faisons la comparaison de la position entre la classification et l'image Pléiade (considérée comme la référence), grâce à cela on peut observer les modifications de position liées à la chaîne de traitement. La précision de l'image Pléiade par rapport à la réalité, correspondant à 0,5m, n'est donc pas prise en compte ici.

b. Résultats pour Bordeaux Métropole zone Est et Ouest

Nous obtenons comme résultats une moyenne des incertitudes de 3,4m et une erreur quadratique de 4,3m.

Pour comprendre d'où proviennent ces erreurs qui correspondent à l'écart entre l'image Pléiade et la classification pour chaque point d'échantillonnage, les erreurs en X et Y sont tracées (avec pour abscisse X et pour ordonnée Y) pour chacun des points échantillonnés (figure 27). L'origine du graphe correspond à l'emplacement de chaque point sur l'image Pléiade. On peut donc visualiser l'écart entre la classification et l'image Pléiade pour chacun des points.

Ce graphique (figure 27) montre une surestimation du bâti vers le nord dans la classification. On observe ainsi que, d'une manière générale, les surfaces des bâtiments sont surestimées. Pour vérifier si cette surestimation ne provient pas des ombres des bâtiments, j'ai sélectionné tous les échantillons de bâti se situant dans l'ombre des bâtiments. Ces points correspondent aux points en oranges sur la Figure 27. On confirme donc que la surestimation des surfaces des bâtiments vers le nord est en relation avec l'ombre des bâtiments.

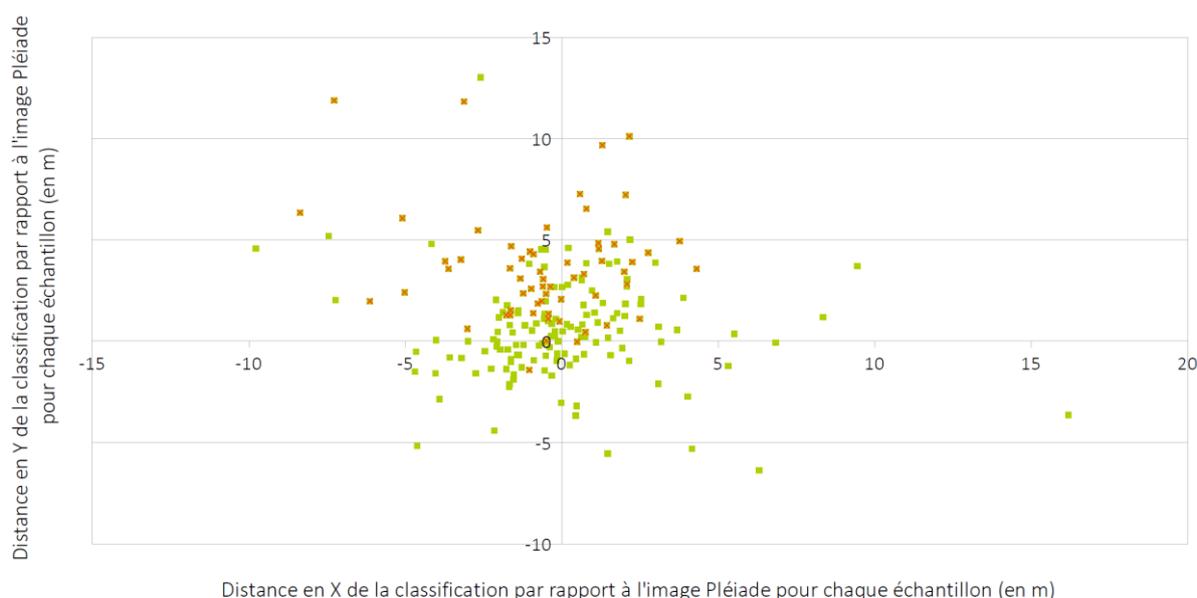


Figure 27: visualisation de l'écart de position entre la classification et l'image Pléiade pour chacun des points d'échantillonnage (les points orange correspondent à des échantillons situés sur des bâtiments présentant des ombres).

c. Conclusions

Pour conclure, on peut observer un écart entre la classification et l'image Pléiade. Cet écart ne correspond pas à un décalage constant en X et Y mais plutôt à une surestimation du bâti, comme démontré par le graphe de la figure 27. On peut également observer que l'ombre des bâtiments peut être une cause de la surestimation vers le nord des bâtiments.

Une possibilité serait de prendre en compte ces ombres dans le calcul des bâtiments:

- Soit en définissant en amont de la classification et en ajoutant aux 5 classes déjà définies une classe "Ombre". Cela pourrait peut-être enlever de toute la classification la surestimation du bâti dans les ombres et donc dans le cas d'étude d'évolution entre deux dates, la surestimation pourrait être amoindrie voire supprimée.
- Soit en sous-estimant la classe bâtis d'un certain pourcentage qui sera défini par rapport à la surface d'ombre des bâtiments.

Afin de pouvoir noter facilement la précision de position, j'ai utilisé la notation tirée de la fiche "Critère de position" de la série de fiches "Qualifier les données géographiques" du CEREMA, 2017 (Bonin & Gallais, 2017).

Note	Valeur moyenne des incertitudes (VMI)
5	$VMI \leq 0,4 \text{ m}$
4	$0,4 \text{ m} < VMI \leq 1,5 \text{ m}$
3	$1,5 \text{ m} < VMI \leq 5 \text{ m}$
2	$5 \text{ m} < VMI \leq 20 \text{ m}$
1	$VMI > 20 \text{ m}$

Tableau 17: Notation du critère de précision de position.

On obtient donc ici une note de 3 avec une moyenne des incertitudes de 3,4m. Cette valeur peut être suffisante en fonction de l'usage que l'on veut faire de ces données.

5. La qualité temporelle

La qualité temporelle se définit comme "la qualité des attributs temporels et des relations temporelles entre les entités" (Bonin & Gallais, 2017).

Initialement, d'après la norme ISO 19157, la qualité temporelle a pour but d'évaluer les attributs qui correspondent à des mesures temporelles et non d'évaluer les informations temporelles contenues dans les métadonnées.

Cependant ce type de données n'est pas présent dans le résultat de la classification. Il est donc nécessaire de redéfinir une qualité temporelle qui sera possible d'appliquer par la suite dans notre étude.

a. Méthodologie

Les classifications obtenues ne contenant pas d'attributs temporelles, les conditions d'applications de la norme ne sont pas entièrement respectées. Un point important de la classification est l'utilisation de données provenant de la même année : contrairement à ce que nous indique la norme, on travaille donc avec les informations des métadonnées (Gerard, 2014) (annexe 12).

b. Résultats

A partir du tableau 18, nous pouvons observer que toutes nos données sont donc situées à moins de 1 an de notre image Pléiade. Le plus grand écart de temps est celui entre le MNT produit 8 mois avant l'image Pléiade. Cependant, 8 mois correspond à la fourchette de temps souhaitée, de plus il est peu probable que le MNT se modifie en quelque mois.

Données d'entrée	Production
Image Pléiade	21/03/2016 et 02/04/2016
MNS	21/03/2016 et 02/04/2016
MNT	02/07/2015
RPG	Version 2016
BDtopo® (bâti, route, aire de triage, voie ferrée)	06/04/2016

Tableau 18: Millésime utilisé pour chaque donnée en entrée de la chaîne de traitement.

c. Conclusions

Le millésime des données en entrée est un critère très important lorsque l'on veut mesurer une évolution de la dynamique urbaine, surtout dans le cas de calcul d'évolution des dynamiques urbaine entre des périodes proches.

Une notation est également mise en place pour ce critère (tableau 19):

Note	Ecart de date entre les données	
5	1 an	Les données d'entrées utilisées datent de la même années.
4	2 ans	Les données d'entrées utilisées proviennent de dates pouvant avoir des écarts allant jusqu'à 2 ans.
3	3 ans	Les données d'entrées utilisées proviennent de dates pouvant avoir des écarts allant jusqu'à 3 ans.
2	4 ans	Les données d'entrées utilisées proviennent de dates pouvant avoir des écart allant jusqu'à 3 ans.
1	>4 ans	Les données d'entrées utilisées proviennent de dates pouvant avoir des écart allant jusqu'à 4 ans.

Tableau 19: Notation du critère de qualité temporelle.

On obtient donc ici une note de 5, qui est la note maximale concernant ce critère.

6. L'écart des dynamiques

Les différents critères de qualité vu précédemment ne permettent pas de répondre à une des problématiques de l'étude, qui est celle de l'évolution des bâtis entre deux dates (la dynamique de bâti). Le critère de dynamique constitue un nouveau critère créé pour cette étude afin de répondre à cette problématique. Cette solution est mise en place pour avoir une idée de la justesse des mesures des dynamiques. Ce critère permet de mesurer l'écart obtenu entre des surfaces de bâtis mesurés sur deux classifications en utilisant des points de contrôle représentatifs de notre zone d'étude.

Ce critère ne permet pas de mesurer des dynamiques mais il permet de vérifier si il est possible de calculer des dynamiques sur la classifications étudiée. A partir des justesses de classements de deux classifications prises à des dates différentes, on peut savoir à partir de ce critère si on mesure du bruit ou bien de la dynamique urbaine.

a. Méthodologie

Ce critère ne peut être appliqué que sur la partie Est de Bordeaux, puisque la classification pour le millésime de 2014 a été faite uniquement sur cette zone (paragraphe IV.1. Zone d'étude). La méthodologie est décrite en annexe 13. La première étape consiste à refaire la justesse de classement (paragraphe VI.3. La précision thématique) pour obtenir les tableaux des résultats associés et ceci pour les deux millésime 2014 (tableau 20) et 2016 (tableau 21) sur la zone Est de Bordeaux métropole.

Ce critère utilise le pourcentage d'éléments en excès ou omis, sans prendre en compte les éléments qui se compensent, présent dans les tableaux de sorties des résultats de la justesse de classement (annexe 9). Dans le cas de calcul des dynamiques de bâti, on s'intéresse aux surfaces et non à la validité de la position des objets, c'est pour cela que seules les pixels en excès ou omis restant sont pris en compte. Le but est ainsi d'évaluer si le bâti est surestimé ou sous-estimé par rapport aux données de contrôles, et ainsi quantifier si cet écart de mesure, correspondant au bruit, est important ou non par rapport aux dynamiques calculés. Ce résultat permet de juger si la dynamique calculée possède une signification réelle, ou si celle-ci ne correspond qu'à du bruit.

b. Résultats pour Bordeaux Métropole zone Est

En utilisant la même méthodologie que la justesse de classement (annexe 9), on obtient le tableau présentant les excès et omissions restant (sans prendre en compte le pourcentage de pixels compensés). Seule la ligne du tableau présentant le bâti est utile:

Résultat de la justesse de classement sur le millésime 2014 (tableau 20)

%	données correctes	Total des excédents	total des déficits	pixels compensés/ contrebalancés	Pixels en excès restant	Pixels omis restant
Bâti	11.49	1.04	2.09	1.04	0	1.04

Tableau 20: Pourcentage de pixels omis sur Bordeaux métropole zone Est pour le millésime 2014.

Pour la classification de 2014 sur Bordeaux métropole Est, on peut donc dire que l'on va obtenir une surface de bâti qui est sous-estimée de 1.04 % (et cela par rapport à la référence).

Résultat de la justesse de classement sur le millésime 2016 (tableau 21):

%	données correctes	Total des excédents	total des déficits	pixels compensés/ contrebalancés	Pixels en excès restant	Pixels omis restant
Bâti	13.13	4.78	1.04	1.04	3.73	0

Tableau 21: Pourcentage de pixels en excès sur Bordeaux métropole zone Est pour le millésime 2016.

Pour la classification de 2016 sur Bordeaux métropole Est, on peut donc dire que l'on va obtenir une surface de bâti qui est surestimée de 3.73 % (et cela par rapport à la référence).

Donc si nous obtenons une sous-estimation de 1.04% en 2014 et une surestimation de 3.73% en 2016, alors on peut donc dire que concernant la dynamique du bâti entre 2014 et 2016 nous avons un écart avec la référence de 4.77%

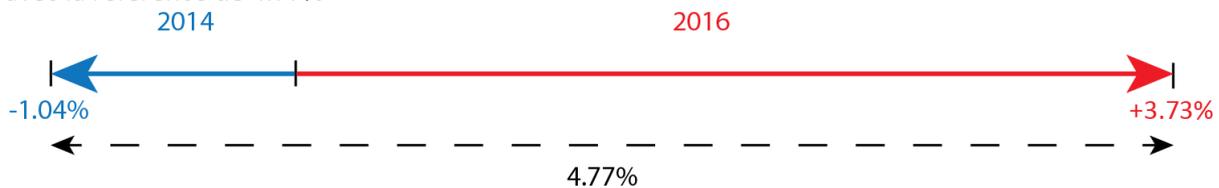


Figure 28: Ecarts d'estimation des dynamiques urbaine entre la classification et la référence.

c. Conclusions

Pour conclure, on observe que l'écart entre la classification et notre référence concernant les dynamiques entre 2014 et 2016 est de 4.77% avec une surestimation en 2016 et une sous-estimation en 2014. Cela veut dire qu'on aura des valeurs très éloignées de dynamique de bâti entre 2014 et 2016 par rapport à la réalité. Cela confirme le fait que la mesure de petites dynamiques comme sur Bordeaux Est qui est très faible, d'environ 0,3% en moyenne par an depuis 10 ans pour Bordeaux d'après les fichiers fonciers. Il est donc impossible d'obtenir des valeurs de dynamiques faible, comme pour l'exemple de Bordeaux métropole Est.

Comme pour les critères précédents, il a été possible de noter ce critère avec la notation suivante (figure 22):

Note	Ecart possible entre deux classifications
5	< 0,5 %
4	De 0,5 % à 1 %
3	De 1 % à 2 %
2	De 2 % à 5 %
1	> 5%

Tableau 22: Notation du critère de dynamique.

Nous obtenons une note de 2. Un écart entre 2 et 5% n'est pas un bon résultat lorsque l'on veut mesurer de faibles dynamiques, il est indispensable d'être à un écart inférieur à 0,5% voire même 0,1%.

VII. Résultats

1. Qualité interne

Concernant le projet mis en place par le CEREMA, aucun cahier des charges précis n'avait été rédigé en amont. La commande initiale était d'expliquer ce qu'il était possible d'obtenir comme informations sur la densité urbaine et ses dynamiques à partir d'image satellitaire de très haute résolution. Le sujet était donc vaste, puis par la suite, d'autres sujets se sont greffés à la commande initiale tel que les problématiques liées aux zones végétalisées en ville ou encore celle liées aux zones imperméables (par exemple pour le Plan de Prévention des Risques d'Inondation PPRI).

2. Qualité externe

Après avoir évalué la classification à partir de critères de qualité, j'ai pu vérifier si la donnée était en adéquation avec les besoins nécessaires pour différents usages. Pour cela j'ai utilisé les résultats obtenus pour chacun des critères étudiés et les notes correspondantes que j'ai attribué.

A partir des notes sur 5 pour chaque critère de la qualité sur la classification finale, j'ai pu mettre au points des diagrammes représentant la qualité externe du produit (figure 29-30). Dans le cas de mesure de qualité liée à la dynamique urbaine, la notation sur le critère des écarts de dynamique est ajoutée (figure 31), nous avons donc un diagramme utilisant 6 critères et non 5 (comme lorsque nous ne travaillons pas sur des phénomènes de dynamique urbaine).

Ces diagrammes correspondent à des exemples que l'on peut obtenir dans le cas de demande d'utilisateur sur la qualité de la donnée.

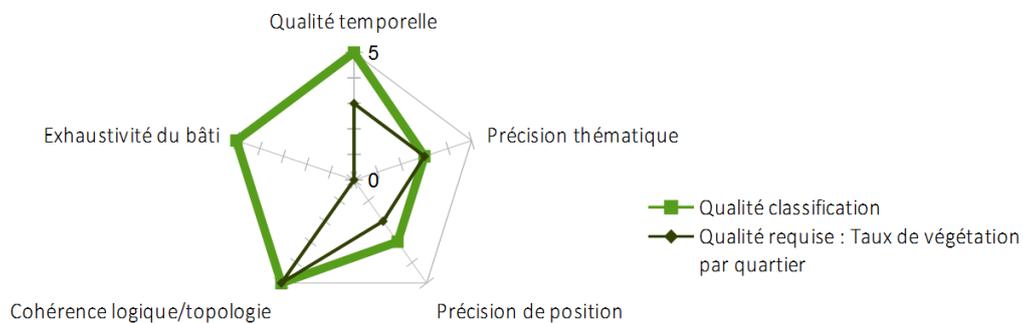


Figure 29: Représentation possible de la qualité externe dans le cas où l'objectif est la mesure de taux de végétation en ville.

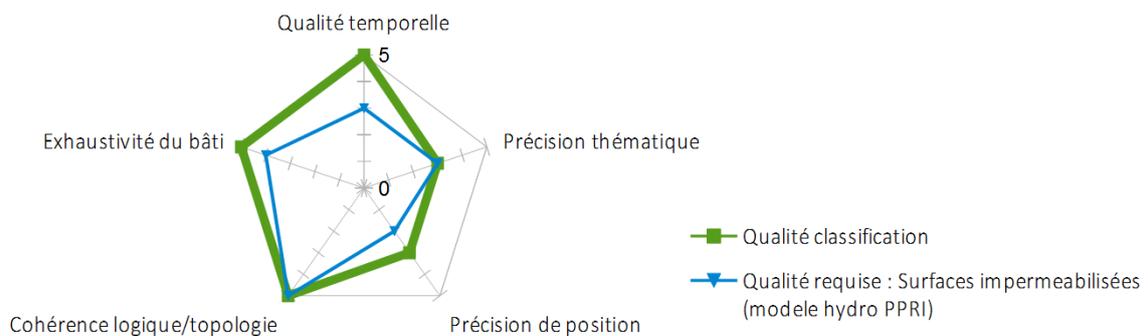


Figure 30: Représentation possible de la qualité externe dans le cas où l'objectif est la mesure des surfaces imperméabilisées.

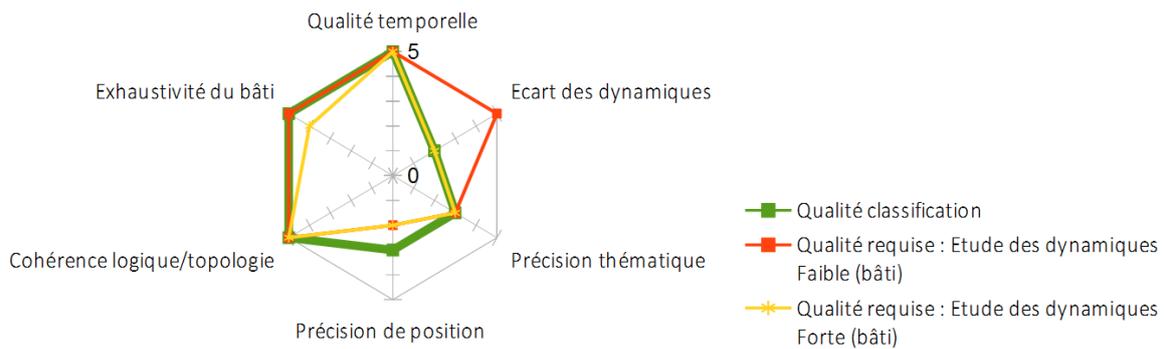


Figure 31 : Représentation possible de la qualité externe dans le cas où l'objectif est la mesure des dynamiques urbaines qu'elle soit considérées comme faibles ou fortes.

Nous avons donc ici des diagrammes à titre d'exemple pour des usages déjà définis. A partir du moment où nous avons obtenu ces critères de qualité et leurs notes associées, il est plus simple de les comparer à une qualité souhaité pour un usage externe défini.

Dans ce cas même si d'après la norme ISO 19157 ou bien INSPIRE, l'étude de la qualité n'est pas obligatoire (CERTU, 2010). Celle-ci peut être exigée par un utilisateur, afin de confirmer que le produit réponde à un besoin spécifique.

Cette étude de qualité de la donnée sur Bordeaux métropole a été résumée dans une fiche type comprenant également l'étude des métadonnées sur cette zone (annexe 15). Cependant il faut tenir compte pour cette étude que le critère d'écart des dynamiques est calculé sur Bordeaux Est à partir du millésime 2014 et 2016, contrairement aux autres critères mesurés sur Bordeaux métropole (Est et Ouest) sur le millésime 2016. Lors de ce stage, le but était de mettre en place une méthodologie pour définir la qualité, et pourvoir par la suite reproduire cette. Pour compléter l'étude sur Bordeaux métropole il faudrait maintenant étudier tous les critères sur Bordeaux métropole Est et Ouest, mais également sur les millésimes 2014 et 2016.

VIII. Etude de dynamique 2014-2016

Les chiffres obtenues pour le critère de qualité de dynamique (paragraphe VI.6.c. Conclusions) pour l'évolution des dynamiques à l'échelle de Bordeaux Métropole n'étant pas concluante sur des dynamiques faibles sur de courtes périodes, j'ai donc essayé d'observer des dynamiques sous un autre angle. C'est à dire de ne pas mesurer les dynamiques de manière générale sur Bordeaux métropole, mais plutôt vérifier si l'on peut observer des évolutions sur des zones beaucoup plus petites.

J'ai donc effectué une comparaison de surface entre 2014 et 2016 sur la zone EST de Bordeaux, puisqu'il n'existe une classification du sol en 2014 que sur la zone EST (figure 4). Le but est d'obtenir un quadrillage comprenant des carrés de 500m de côté recouvrant toute la surface de Bordeaux Est, dans lesquels on calcule le pourcentage de différences entre les deux millésimes par rapport à la surface de chaque carré.

J'ai utilisé un outil interne CEREMA permettant de mesurer la surface de bâtis d'une classification pour chaque carré. Cette opération a été effectuée sur la classification 2014 et 2016 puis la soustraction des résultats entre 2016 et 2014 a été réalisé avec l'aide de QGIS. Cette valeur a été ensuite normée en pourcentage d'évolution de bâti par carré.

L'étape suivante consiste à faire ressortir les éléments avec un fort déficit ou en fort excès par rapport à la classification générale. Afin d'observer si malgré une surestimation et sous-estimation du bâti, il est possible de faire ressortir des différences entre deux millésimes au niveau du bâti (figure 32 et 33).

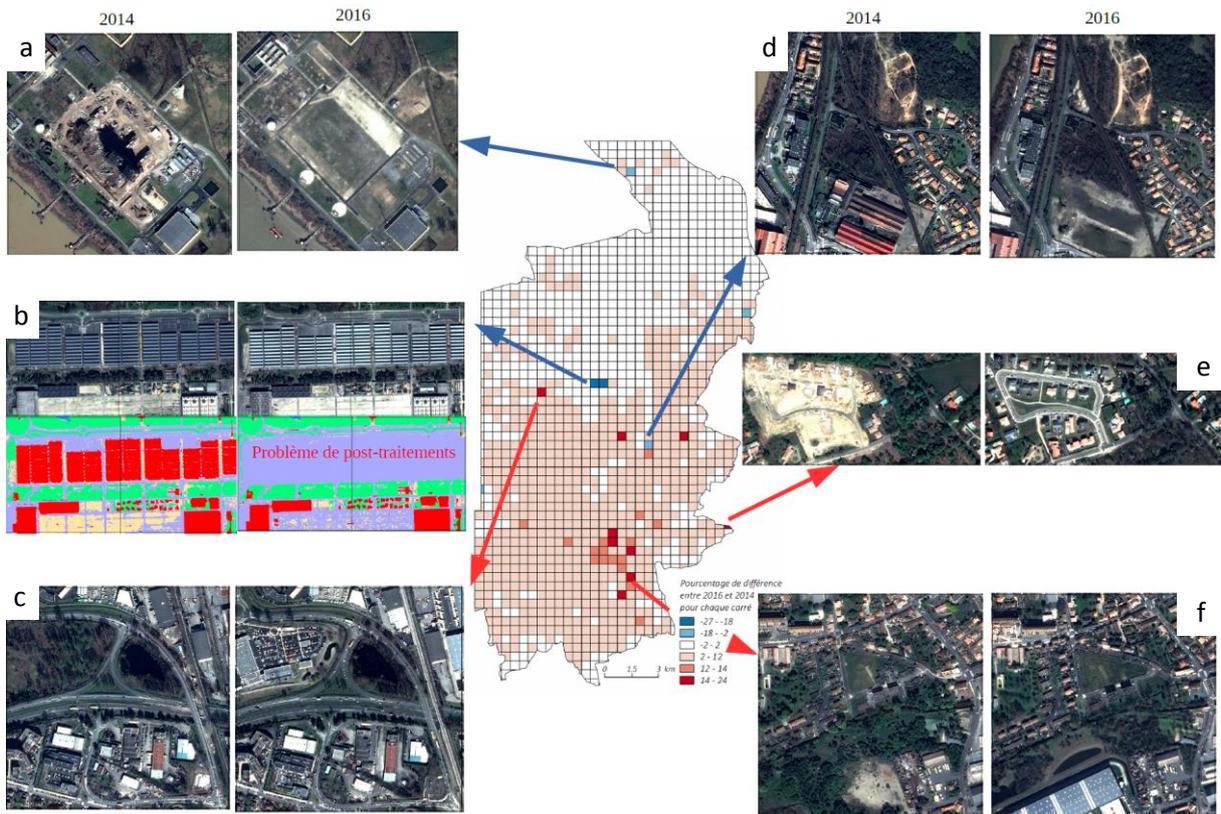


Figure 32: Etude de l'évolution du bâti sur Bordeaux Métropole Est à partir d'un quadrillage de 500m² (partie1)

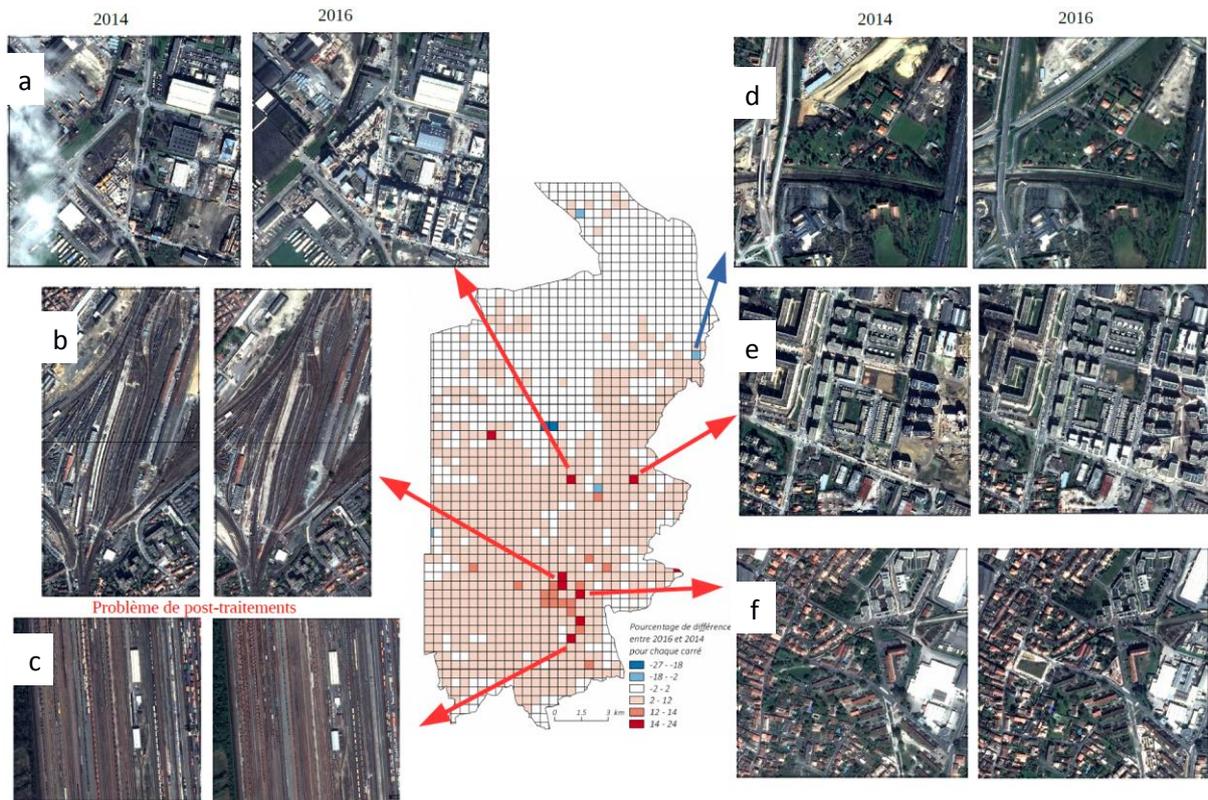


Figure 33: Etude de l'évolution du bâti sur Bordeaux Métropole Est à partir d'un quadrillage de 500m² (partie2)

On peut donc observer dans les figure 32 et 33 que les évolutions de bâti entre deux millésimes ressortent bien, si celle-ci sont importantes (comme par exemple la construction de bâtiments figure 32 c, e et f ou encore la destruction de bâtiments figure 32 a et d).

Il est donc possible d'envisager comme amélioration de l'outil, une carte présentant les zones où il y a potentiellement une forte augmentation ou diminution de bâti. L'utilisateur devra ensuite vérifier s'il y a véritablement une évolution du bâti dans ces zones ou bien si cela provient du bruit (ou d'une erreur de classification).

Il est cependant nécessaire de continuer ces tests sur des quadrillages plus faibles ou auprès d'autres classifications afin d'observer s'il est possible de visualiser des évolutions ou diminutions de bâtis plus fines que celles observées dans notre exemple.

Une autre remarque importante concernant les post-traitements peut être faite à partir des observations où il est possible d'observer des erreurs (figure 32 b, et 33 b, c). En effet, il est nécessaire de faire attention à ce que les post-traitements soient effectués par une même personne ou avec des post-traitements très spécifiés.

IX. Conclusion

Cette étude m'a permis de mettre en place une méthodologie permettant de définir la qualité des données sortant de la chaîne de traitement mise en place par le CEREMA et ainsi de s'assurer que la qualité de la donnée correspond bien à la qualité nécessaire pour son utilisation (qualité externe). Cette méthodologie repose sur 5 critères, définis d'après la norme ISO 19157. Un nouveau critère a également été développé pour prendre en compte la qualité de la dynamique urbaine entre deux millésimes.

L'étude réalisée sur Bordeaux métropole montre tout d'abord l'importance de l'étude la qualité de la donnée. Si celle-ci n'est pas obligatoire, cette étude montre qu'elle est très fortement conseillée. Cela confirme bien le fait qu'il faut obligatoirement connaître la donnée que l'on va utiliser et savoir si elle convient ou pas à l'utilisation que l'on veut en faire. Dans ce stage, il est possible de voir que selon l'usage externe voulu, la donnée en sortie de la chaîne peut ou ne peut pas être utilisée.

Cette étude souligne également la complexité d'une étude de qualité. En effet des normes ou directives existent et donnent des indications. Cependant, il est parfois difficile de les appliquer et les indications pour les mettre en place restent souvent très vagues.

Cette remarque est bien illustrée par l'échantillonnage. En effet, le diagramme des échantillonnages de la norme Iso 19157 propose plusieurs choix, probabiliste, aléatoire, au jugé, mais après une étude un peu plus approfondie sur nos résultats en terme d'échantillonnage, il est possible de voir que selon le choix d'échantillonnage, les résultats obtenus pour chaque type d'échantillonnage sont très différents. Par exemple, dans notre étude, le choix d'un échantillonnage au jugé et non probabiliste prenant en compte 1,05 milliard de pixels est moins représentatif qu'un échantillonnage aléatoire et probabiliste prenant en compte moins de 700 points d'échantillonnages. De nombreuses études peuvent donc surestimer la qualité de la donnée en utilisant un échantillonnage non adapté et surtout non représentatif (ne prenant pas en compte les effets de bords par exemple) et donc poser des questions quant à leurs résultats.

L'objectif à terme du CEREMA est de reproduire cette méthodologie sur de nombreuses études, afin d'étudier la qualité et obtenir un ordre de grandeur sur chaque résultat des différents critères de qualité, et d'intégrer une notion d'incertitude. Cela permettrait de connaître si nous avons une qualité de classification plus ou moins constante sur les cartes d'occupation du sol en sortie de la chaîne de traitement. Ou bien si en fonction de l'échelle ou de la morphologie de la zone étudiée, la qualité de la classification sera fortement différente. En fonction des résultats, plusieurs méthodes peuvent être envisagées.

- La première est dans le cas où les résultats sur la qualité sont constants (les valeurs résultantes des critères sont très proches en fonction des études), sans trop de variabilité sur des données

en sortie après de nombreuses études de qualité à différentes échelles et avec des morphologies variées. Dans ce cas, on aura une valeur générale pour chacun des différents critères de qualité et cela pour toutes les occupations du sol en sortie de la chaîne de traitement.

- La deuxième est dans le cas où les résultats de qualité varient fortement. Dans ce cas, il est peut-être possible d'obtenir des ordres de grandeurs, non systématiques à l'ensemble des études, mais qui dépendront de l'échelle ou bien de la morphologie étudiée.
- La troisième correspond au cas où les résultats (même en fonction de l'échelle ou de la morphologie) sont trop hétéroclites. Il sera donc nécessaire d'effectuer une étude de qualité après chaque classification de l'occupation du sol effectuée.

SOURCES CONSULTEES

Bibliographie:

- ANTONI, Jean-Philippe et YOUSSEFI, Samy. Étalement urbain et consommation d'espace. Étude comparée de Besançon, Belfort et Montbéliard. *Revue Géographique de l'Est*, 2007, vol. 47, no 3.
- BECHET, Béatrice et al. Sols artificialisés et processus d'artificialisation des sols : déterminants, impacts et leviers d'action. Synthèse du rapport d'expertise scientifique collective, Ifsttar-Inra (France), 127 p. 2017.
- BONIN Yves & GALLAIS Arnaud. Qualifier les données géographiques: Un décryptage de la norme ISO 19157. CEREMA, 2017.
- CEREMA. Signature n°60 - Mai 2016 - CEREMA: Territoires et ville, 2016.
- CONSEIL NATIONAL DE L'INFORMATION GEOGRAPHIQUE (CNIG). Guide de saisie des éléments de métadonnées INSPIRE: Appliqué aux données, 2013.
- DE KEERSMAECKER Marie-Laurence. Stratégie d'échantillonnage des données de terrain intégrées dans l'analyse des images satellitaires. In: Espace géographique, tome 16, n°3, 1987. pp. 195-205.
- EUROPEAN COMMISSION. Mapping Guide for a European Urban Atlas, 2011.
- FOODY, Giles M. Sample size determination for image classification accuracy assessment and comparison. *International Journal of Remote Sensing*, 2009, vol. 30, no 20, p. 5273-5291.
- GERARD Marianne. Mesure des espaces agricoles et de leur évolution, Teruti-Lucas au regard de données SIG et qualité de données géographiques d'occupation du sol. Rapport de master, Direction régionale de l'alimentation de l'agriculture et de la forêt de Rhône-Alpes, 2014.
- INSTITUT NATIONAL DE L'INFORMATION GEOGRAPHIQUE ET FORESTIERE (IGN). Guide utilisateur Pléiades. 2015.
- INSTITUT NATIONAL DE L'INFORMATION GEOGRAPHIQUE ET FORESTIERE (IGN). OCS GE Version 1.1: Occupation du Sol à Grande Echelle ; Descriptif de contenu, 2016.
- JOAO Eléanore. Expertise qualité du produit "indicateurs aménagements des territoires urbains" pour le suivi des dynamiques. Rapport de master, Université de Toulouse III et CEREMA, 2017.
- LAUGIER, ROBERT. L'étalement urbain en France. *Centre de Ressources Documentaires Aménagement, Logement et Nature*, 2012.
- MAO Lia. Traitement automatique d'image satellites pour l'extraction d'information d'occupation des sols. Rapport de master, CEREMA, 2014.
- MINISTERE DE L'ALIMENTATION DE L'AGRICULTURE ET DE LA PECHE. Commission de validation des données pour l'information spatialisée, 2010.

MINISTERE DU BUDGET ET DE LA REFORME DE L'ETAT, MINISTERE DU TOURISME DE L'EQUIPEMENT DU TRANSPORT ET DE LA MER. Information Géographique : Recommandation relative à la mise en œuvre de la norme ISO 19115 sur les métadonnées, 2006.

NATIONAL COASTAL DATA DEVELOPMENT CENTER (NCDDC), NATIONAL OCEANOGRAPHIC DATA CENTER (NODC), NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION (NOAA). ISO 19115, Geographic information - Metadata. Guide to Implementing ISO 19115:2003(E), the North American Profile (NAP), and ISO 19110 Feature Catalogue, 2012.

ORGANISATION INTERNATIONALE DE NORMALISATION. ISO 19157:2013 ; Information géographique -- Qualité des données, 2013.

ORGANISATION INTERNATIONALE DE NORMALISATION. ISO 19115-1:2014 ; Information géographique -- Métadonnées -- Partie 1: Principes de base, 2014.

PERALDI, Pascal. Mise en place d'un outil de métadonnées au sein des services de la Communauté d'Agglomération de Montpellier. Rapport de master, Université de RENNES 2, 2007.

Webographie:

BRGM. Géocatalogue. Saisie des métadonnées d'une donnée géographique [en ligne]. Juin 2016. <http://carto.geo-ide.application.developpement-durable.gouv.fr/IHM/metadata/DEPT023A/Publication/metadonnees_L_PERIM_ARCHEO_PREVENTIVE_023.pdf>

CEREMA. <<https://www.cerema.fr/fr>> (a).

CEREMA, Séminaire QuaDoGeo: mise en ratique de la qualification des données. [en ligne]. 20 février 2018. <<https://www.cerema.fr/fr/actualites/seminaire-quadogeo-mise-pratique-qualification-donnees>> (b).

CERTU. La qualité des données géographiques: quels enjeux pour les collectivités?. [en ligne]. Mai 2008. <<https://www.cerema.fr/fr/centre-ressources/boutique/qualite-donnees-geographiques-quels-enjeux-collectivites>>

CERTU. La qualité des données géographiques: Etat des lieux pour un débat. [en ligne]. Décembre 2010. <http://formations-geomatiques.developpement-durable.gouv.fr/NAT009/ADL/Approche_qualite_des_donnees/res/Reflexion_sur_la_qualite_des_donnees_geographique.pdf>

CESBIO. Génération automatique et opérationnelle de cartes d'occupation des sols [en ligne]. 2015. <<http://www.cesbio.ups-tlse.fr/multitemp/?p=6178>>

CONSEIL NATIONAL DE L'INFORMATION GEOGRAPHIQUE (CNIG). INSPIRE [en ligne]. 2018. <http://cnig.gouv.fr/?page_id=8991>

COPERNICUS. Urban Atlas 2012 [en ligne]. 2013. < <https://land.copernicus.eu/local/urban-atlas/urban-atlas-2012>>

INSTITUT NATIONAL DE L'INFORMATION GEOGRAPHIQUE ET FORESTIERE (IGN). RGE ALTI®.
Métadonnées [en ligne]. 2017. < http://professionnels.ign.fr/doc/IGNF_RGEALTIr_2-0.html>

INSTITUT NATIONAL DE L'INFORMATION GEOGRAPHIQUE ET FORESTIERE (IGN). BD TOPO®.
Métadonnées [en ligne]. 2018.
<<http://professionnels.ign.fr/sites/default/files/m%C3%A9tadonn%C3%A9es%20de%20produit%20BD%20TOPO%C2%AE.html>>

MINISTERE DE LA TRANSITION ECOLOGIQUE ET SOLIDAIRE. La commission de validation des données pour l'information spatialisée: COVADIS [en ligne]. 2016. <<https://www.ecologique-solidaire.gouv.fr/commission-validation-des-donnees-linformation-spatialisee>>

MINISTERE DE LA TRANSITION ECOLOGIQUE ET SOLIDAIRE. Commissariat général au Développement durable. Base de données géographique CORINE Land Cover (CLC) [en ligne]. 2018.
<<http://www.statistiques.developpement-durable.gouv.fr/donnees-ligne/li/2539/0/base-donnees-geographique-corine-land-cover-clc.html>>

MINISTERE DU LOGEMENT ET DE L'EGALITE DES TERRITOIRES, loi ALUR : Lutte contre l'étalement urbain. [en ligne]. 24 mars 2014. <http://www.cohesion-territoires.gouv.fr/IMG/pdf/alur_fiche_lutte_contre_l_etalement_urbain.pdf>

THEIA. CES Occupation des sols [en ligne]. 2018. <<http://www.theia-land.fr/fr/ces-occupation-sols>>