

WalousMàJ

Analyse changements

Version	1
Révision	2
Projet	WalousMàJ
Date	28/09/2022

	Nom	Société	Signature
Auteur	Damien Draime	Aerospacelab	
Relecteur / Approbateur	Jonathan Denies	Aerospacelab	
Approbateur	Benoît Deper	Aerospacelab	<i>Benoît Deper</i>

Table des matières

Table des matières	2
Abréviations et acronymes	3
1. Objet du rapport	4
2. Approche	4
2.1. Labélisation de points échantillonnés aléatoirement	4
2.2. Détection des objets d'une classe pour des zones échantillonnées aléatoirement	5
2.3. Segmentation des objets d'une classe pour des zones échantillonnées aléatoirement	6
3. Données de validation	7
3.1. Constitution	7
3.2. Consensus	7
4. Analyse de changements	7

Abréviations et acronymes

SPW	Service Public de Wallonie
WALOUS_OCS__2018	Carte d'occupation du sol de la Wallonie pour 2018 générée par le projet WALOUS
WALOUS_OCS_IA__2019	Carte d'occupation du sol de la Wallonie pour 2019 générée pour ce marché

1. Objet du rapport

Ce rapport a pour but d'analyser la fiabilité des changements observés dans la carte d'occupation du sol de 2019, générée dans le cadre du marché de mise à jour de la carte d'occupation du sol de Wallonie, WALOUSMAJ, lorsqu'elle est comparée à la version de 2018 issue de WALOUS 2018.

Cette analyse diachronique des changements se concentre sur la classe « Constructions artificielles hors sol » et se base sur plusieurs labélisations manuelles et indépendantes d'experts.

2. Approche

Plusieurs approches ont été considérées afin de mener cette analyse diachronique des changements. Chacune présente des avantages et inconvénients qui sont listés, ci-dessous, dans les différentes sous-sections.

2.1. Labélisation de points échantillonnés aléatoirement

Cette méthode demande à l'expert photo-interprétation d'assigner une classe, sur base des orthophotos, à chacun des points échantillonnés aléatoirement sur le territoire wallon, pour chacune des années étudiées.

Les points sont échantillonnés aléatoirement dans les zones de changements considérés. Ces zones de changements sont obtenues en comparant les cartes d'occupation du sol représentant la situation pour 2 années différentes, soient t_1 et t_2 . En limitant l'analyse aux changements liés à une seule classe, k , un point, p , doit, pour être considéré pour l'échantillonnage, avoir changé de classe entre les deux années d'intérêt et avoir appartenu à la classe d'intérêt, k , pour l'une des 2 années :

$$C_p^{t_1} \neq C_p^{t_2} \wedge (C_p^{t_1} = k \vee C_p^{t_2} = k)$$

où C_p^t est la classe du point p au temps t

Pour chacune des années considérées, le point est labélisé sur base de l'orthophoto correspondante. La labélisation se fait sur base de la classe visible sur l'orthophoto, c'est-à-dire « vue du ciel ». Outre l'assignement à une des 11 classes principales, l'expert doit également indiquer si le point se trouve près de la frontière entre plusieurs classes. Ceci permet à cette approche de prendre en compte les effets de bords.

Cette approche a également l'avantage de permettre une grande variété dans l'échantillonnage des points. En effet, cette approche est moins exigeante sur la charge de travail par point et donc permet d'échantillonner plus de points.

Cependant, les erreurs d'omissions ne sont pas prises en compte par cette approche puisqu'elle se base sur la carte des changements. Ainsi, si aucun des deux modèles n'a prédit la bonne classe pour un point donné, alors ce point ne sera jamais échantillonné et donc ne sera pas considéré pour cette analyse.

Cette approche est illustrée avec une étude cas à la Figure 1. Dans cet exemple, la classe d'intérêt est « Constructions artificielles hors sol » (label : 2). Le *Point 1* concerne un point correctement classifié dans WALOUS_OCS__2018 mais pas dans WALOUS_OCS_IA__2019, résultant en un faux positif dans la carte de changement. Ce point, répondant aux critères d'échantillonnage évoqués ci-dessus, a été échantillonné dans cet exemple et il est donc demandé à l'expert de labéliser ce point à la fois pour 2018 et 2019, ainsi qu'indiquer si ce point se trouve à proximité d'une autre classe. Il en va de même pour le *Point 2* qui est cette fois-ci correctement labélisé dans WALOUS_OCS_IA__2019 mais pas dans WALOUS_OCS__2018. Cependant, le *Point 3*, bien qu'il concerne un changement réel, il n'est pas considéré pour l'échantillonnage car aucune des deux cartes d'occupation du sol ne l'a classifié comme appartenant à la classe des « Constructions artificielles hors sol ». Ceci illustre le fait que cette approche ne permet pas de prendre en compte les erreurs d'omission. Le *Point 4* est correctement labélisé par les deux cartes d'occupation du sol. Il représente donc un réel changement et cela est confirmé par la photo-interprétation de l'expert.



	2018		2019		Bords
	WALOUS OCS 2018	Photo-interprétation	WALOUS OCS IA 2019	Photo-interprétation	
Point 1	2	2	1	2	0
Point 2	2	7	7	7	0
Point 3	7	N/A	1	N/A	N/A
Point 4	4	4	2	2	1

Figure 1 - Illustration de l'Approche 1

2.2. Détection des objets d'une classe pour des zones échantillonnées aléatoirement

Dans cette approche, des zones sont échantillonnées aléatoirement et il est demandé à l'expert de détecter, à l'intérieure de celles-ci, tous les objets de la classe d'intérêt, k , et ce pour chacune des années d'intérêt. La détection d'un objet se fait en mettant un point sur ce même objet.

L'avantage de cette approche est qu'elle permet de prendre en compte les erreurs d'omission puisqu'elle ne repose pas sur les cartes d'occupation du sol pour l'échantillonnage des zones.

Cependant, elle présente de nombreux désavantages. La première question que cette approche soulève est de savoir où mettre le point définissant un objet ? Et faut-il mettre plusieurs points pour un même objet ? De plus, faut-il détecter chaque instance de la classe ou plutôt faire une détection sémantique ? Ces questions sont cruciales car dans son application la plus simple, à savoir la détection sémantique de chaque objet par un point, cette approche ne permet, par exemple, pas de détecter les constructions de nouvelles annexes attenantes à une maison déjà construite.

Cette approche ne prend pas non plus en compte les effets de bords.

Finalement cette approche est plus exigeante que la première en termes de charge de travail.



Figure 2 - Illustration de l'Approche 2, application aux "Constructions artificielles hors sol"

2.3. Segmentation des objets d'une classe pour des zones échantillonnées aléatoirement

Cette approche est similaire à la deuxième approche dans son échantillonnage des zones à labéliser. Cependant, elle se distingue de celle-ci par la segmentation, et non la détection, des objets. Ainsi, cette nouvelle approche permet la proportion de la surface de l'objet qui est correctement classifiée par les différents modèles.

Cette approche peut également prendre en compte les erreurs d'omission et les effets de bords.

Cependant, tout comme pour la seconde approche, il convient de décider si la segmentation doit se faire de manière sémantique ou par instance. Sachant que pour certaines classes, il peut être difficile de distinguer chacune des instances.

Néanmoins, cette approche est la plus exigeante des 3 sur la charge de travail qu'elle requiert pour générer le jeu de données nécessaire à cette analyse.



Figure 3 - Illustration de l'Approche 3, application aux "Constructions artificielles hors sol"

La première approche de labélisation de points échantillonnés aléatoirement a été sélectionnée par le SPW. De plus, il a été décidé de concentrer cette analyse uniquement sur les changements liés à la classe « Constructions artificielles hors sol ».

3. Données de validation

3.1. Constitution

Le jeu de données de validation, constitué pour cette analyse, fût initialement composé de 625 points échantillonnés aléatoirement parmi l'ensemble des pixels ayant connu une transition d'état de, ou vers, la classe « Constructions artificielles hors sol » entre 2018 et 2019, sur base de WALOUS_OCS__2018 et WALOUS_OCS_IA__2019. 23 de ces points ont été rejetés soit parce qu'ils tombaient sur des zones floutées (e.g. : sites militaires), soit parce qu'ils n'ont pas pu être classifiés par les experts.

Deux experts au sein du SPW se sont chargés de labéliser les 602 points pour les années 2018 et 2019, en se basant uniquement sur l'orthophoto de l'année correspondante. De plus, pour chacun des points, l'expert a également signalé si ce premier se trouvait à moins de 1 mètre d'une frontière avec une autre classe. Ces deux photo-interprétations ont été conduites indépendamment l'une de l'autre. Dans la suite de ce rapport, ces experts seront référés comme Expert A et Expert B.

3.2. Consensus

Il convient premièrement de noter que même entre experts, il existe des désaccords sur la classification à donner à certains points. En effet, les deux experts n'ont pas donné le même label que celui assigné par la carte d'occupation du sol à 22.26% et 24.42% des points pour l'année 2018 et 2019, respectivement. De même, pour plus de 24% des points, les experts ne sont pas d'accord entre eux sur la proximité du point avec autre objet d'une autre classe. Au final, les experts sont d'accord entre eux, pour l'ensemble des 3 attributs à assigner à un point (i.e. label 2018, label 2019, et bords), dans 49% des cas.

Même en simplifiant le problème, en ne considérant que deux labels : « Constructions artificielles hors sol » et « Autres », les experts ne sont d'accords entre eux, sur les classes à assigner au point en 2018 et 2019, que pour 77.41% des 602 labélisés.

4. Analyse de changements

Pour analyser les changements, il convient d'étudier les matrices de confusion générées en comparant les transitions d'état prédites par les modèles et celles prédites par les experts.

Afin de faciliter la lecture de ces matrices de confusion, les différents changements ont été regroupés en 4 transitions d'état principales. Ces transitions d'état suivent la convention de nomenclature suivante :

$$C^{2018} \rightarrow C^{2019}$$

où C^t est la classe d'un point au temps t , photo-interprétée par un expert ou prédite par un modèle

C^t peut prendre deux valeurs distinctes :

$$C^t = \begin{cases} 2 \\ X \end{cases}$$

où 2 est le label de la classe « Constructions artificielles hors sol » et X désigne toutes les autres classes

En suivant cette convention, les 4 transitions d'état principales sont :

- $2 \rightarrow X$
Celle-ci regroupe tous les changements depuis une « Construction artificielle hors sol » en 2018 vers une autre classe en 2019.
- $X \rightarrow 2$
Elle regroupe les changements d'une classe (excepté « Constructions artificielles hors sol ») vers la classe « Constructions artificielles hors sol » en 2019.
- $X \rightarrow X$
Cette catégorie regroupe les cas n'impliquant pas la classe « Constructions artificielles hors sol » pour les deux années. Il peut donc s'agir de changements d'une classe vers une autre, ou de l'absence de changement.
- $2 \rightarrow 2$
Cette dernière transition d'état regroupe les cas où aucun changement n'aurait été observé entre 2018 et 2019 par l'expert pour un point appartenant à la classe « Constructions artificielles hors sol ».

Alors que les deux premières transitions d'état peuvent décrire les photo-interprétations des experts ainsi les prédictions des deux modèles, les deux dernières ne servent qu'à décrire les photo-interprétations des experts. En effet, pour rappel, l'échantillonnage du jeu de données s'est basé sur les prédictions des modèles en 2018 et 2019 pour ne considérer que les situations où un changement de classes concernant une « Construction artificielle hors sol » était survenu entre ces deux années.

Le regroupement sous ces 4 transitions d'état permet de mitiger les éventuels désaccords entre experts pour les classes autre que « Constructions artificielles hors sol ». En effet, comme mentionné dans la section précédente, de nombreux désaccords subsistent entre les deux experts qui peuvent masquer des tendances utiles à la présente analyse. Cette dernière s'efforce d'établir si l'utilisateur de `WALOUS_OCS_IA_2019` peut avoir confiance lorsque cette carte indique un changement lié à la classe « Constructions artificielles hors sol ». Cette analyse ne cherche pas à déterminer la précision des modèles dans leurs prédictions des autres classes, ceci ayant déjà fait l'objet d'une étude en profondeur lors de l'analyse des performances du modèle. Etant uniquement préoccupé de savoir si un changement lié à la classe « Constructions artificielles hors sol » a en effet eu lieu, le regroupement sous les 4 transitions d'état principales prend tout son sens puisqu'il permet de gommer les éventuels désaccords entre experts sur la définition d'autres classes que celle d'intérêt pour cette analyse. Ceci facilite alors l'observation des tendances propres à la présente analyse.

Les matrices de confusion de la Figure 4 permettent le constat de différents points :

- pour la majorité des points échantillonnés (dans 75.09% et 73.76% des cas pour l'Expert A et B, respectivement), les experts considèrent qu'il n'y a pas de changements liés à la classe « Constructions artificielles hors sol » (i.e. : transitions d'état $2 \rightarrow 2$ et $X \rightarrow X$) ;
- lorsque les experts identifient un changement lié à la classe « Constructions artificielles hors sol » (i.e. : transitions d'état photo-interprétées $2 \rightarrow X$ et $X \rightarrow 2$), ceux-ci confirment les prédictions des modèles dans la toute grande majorité des cas ;
- lorsque les photo-interprétations des experts mènent à une transition d'état $X \rightarrow X$, ces derniers sont dans plus de 66% des cas, en moyenne, d'accord avec la labélisation de la carte d'occupation du sol de 2019 et ils ne partagent pas la labélisation du point en « Constructions artificielles hors sol » proposée dans l'édition de 2018 de la carte d'occupation du sol.

		Expert A			Expert B		
		Transition d'état prédite par les modèles			Transition d'état prédite par les modèles		
Transition d'état photo-interprétée		2 → X	X → 2	Total	2 → X	X → 2	Total
		2 → 2	135 22.43%	126 20.93%	261 43.36%	109 18.11%	116 19.27%
2 → X	64 10.63%	5 0.83%	69 11.46%	59 9.80%	16 2.66%	75 12.46%	
X → 2	18 2.99%	63 10.47%	81 13.46%	28 4.65%	55 9.14%	83 13.79%	
X → X	130 21.59%	61 10.13%	191 31.73%	151 25.08%	68 11.30%	219 36.38%	
Total	347 57.64%	255 42.36%	602 100%	347 57.64%	255 42.36%	602 100%	

Figure 4 - Matrices de confusion pour chacun des deux experts (n=602)

Lorsque les effets de bords sont pris en compte dans la matrice de confusion, la Figure 5 peut être obtenue :

		Expert A			Expert B		
		Transition d'état prédite par les modèles			Transition d'état prédite par les modèles		
Transition d'état photo-interprétée		2 → X	X → 2	Total	2 → X	X → 2	Total
		2 → 2	13 122	16 110	29 232	39 70	50 66
2 → X	7 57	0 5	7 62	7 52	0 16	7 68	
X → 2	0 18	15 48	15 66	1 27	24 31	25 58	
X → X	23 107	5 56	28 163	51 100	16 52	67 152	
Total	43 304	36 219	79 523	98 249	90 165	188 414	

Figure 5 - Matrices de confusion pour chacun des deux experts, avec effets de bords (n=602). Le nombre en haut à gauche d'une cellule représente le nombre de points considérés comme éloignés d'un bord par l'expert. Le nombre en bas à droite d'une cellule est le nombre de points à proximité (i.e.: < 1m) d'un bord et donc considéré comme effet de bords.

Si tous les points échantillonnés sont considérés comme, en effet, refléter un changement lié à la classe « Constructions artificielles hors sol », alors, dans 86.9% et 68.8% des cas, ceux-ci peuvent être le résultat d'un effet de bords, selon les experts A et B, respectivement. Le constat reste néanmoins similaire si l'on ne considère que les points reflétant un changement lié à la classe « Constructions artificielles hors sol » selon la photo-interprétation de l'expert. Dans ce cas, ce sont 85.3% et 79.7% de ces changements qui peuvent être lié à un effet de bords, selon les experts A et B, respectivement.