

WalousMàJ

Constitution des jeux de données VAL_Set et TEST_Set

Version	3
Révision	1
Projet	WalousMàJ
Date	31/01/2022

	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
Disclaimer	3
Abréviations et acronymes	4
1. Objet du rapport	5
2. Méthode	5
2.1. Echantillonnage	5
2.2. Photo-interprétation	7
2.3. Contre-validation	9
3. Résultats	10
4. Références	11

Disclaimer

Les informations contenues dans ce document sont confidentielles, privilégiées et uniquement destinées à l'information du destinataire et ne peuvent être utilisées, publiées ou redistribuées sans le consentement écrit préalable d'AEROSPACELAB. Les opinions exprimées sont de bonne foi et bien que tout le soin ait été apporté à la préparation de ce document, les auteurs ne font aucune déclaration et ne donnent aucune garantie de quelque nature que ce soit concernant ce document, y compris, mais sans s'y limiter, l'exactitude ou l'exhaustivité des informations, faits et/ou opinions qu'il contient.

Abréviations et acronymes

CRA-W	Centre Wallon De Recherches Agronomiques de Wallonie
SOL_2018	Solution développée dans le cadre du projet WALOUS afin de générer la carte d'occupation du sol de la Wallonie pour 2018
SOL_2019/20	Solution proposée par AEROSPACELAB à l'issue de ce marché
SPW	Service Public de Wallonie
TEST_Set	Jeu de données servant à évaluer la solution finale durant la phase III
TRAIN_Set	Jeu de données servant à entraîner les modèles d'une solution intermédiaire
VAL_Set	Jeu de données servant à valider une solution intermédiaire durant la phase de prototypage. Inclut VAL_Set_2018 et, potentiellement, VAL_Set_2019/20
VAL_Set_2018	Jeu de données de validation se basant sur les données de 2018
VAL_Set_2019/20	Jeu de données de validation se basant sur les données de 2019 (ou 2020)
WAL_OCS_2018	Carte d'occupation du sol de la Wallonie pour 2018 générée par le projet WALOUS
WAL_OCS_2019/20	Carte d'occupation du sol de la Wallonie pour 2019 (ou 2020) générée pour ce marché
WAL_CHGT_OCS__2018_2019/20	Produit vectoriel regroupant les changements survenus entre 2018 et 2019 (ou 2020) et labélisés

1. Objet du rapport

Ce rapport décrit les jeux de données VAL_Set et TEST_Set servant, respectivement, à la validation des solutions intermédiaires et à l'évaluation de la solution finale. La méthode d'échantillonnage des points est détaillée ainsi que le protocole de labélisation basée sur la photo-interprétation.

VAL_Set est constitué, d'une part, du jeu de données VAL_Set_2018 ayant servi à valider SOL_2018 et, d'autre part, d'un nouveau jeu de données se basant sur les orthophotos de 2019, VAL_Set_2019 (ou de 2020, VAL_Set_2020).

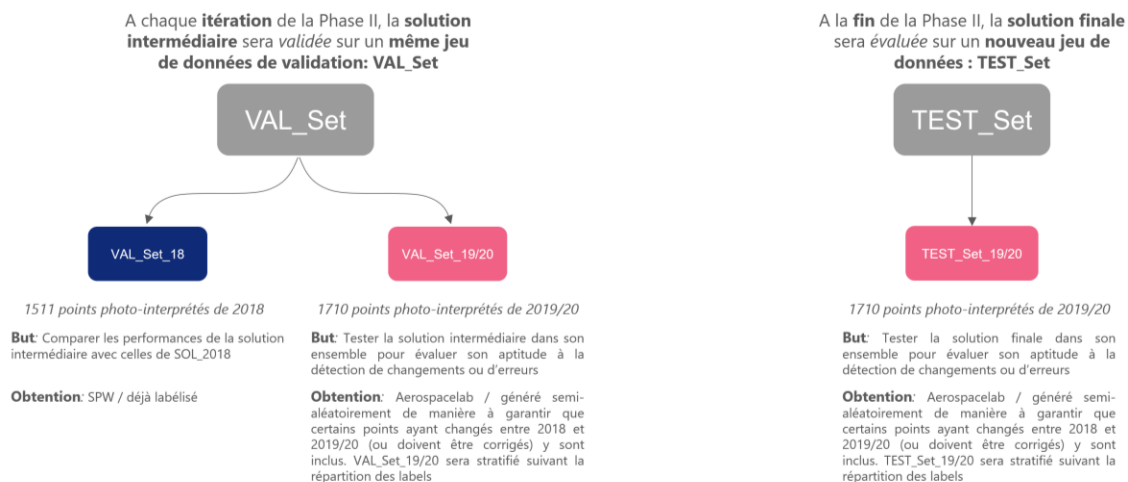


Figure 1 - Constitution des VAL_Sets et TEST_Sets

Le jeu de données d'évaluation TEST_Set est, quant à lui, constitué uniquement de nouvelles données annotées pour 2019 ou pour 2020, dépendamment de l'année pour laquelle la solution fait son inférence.

2. Méthode

2.1. Echantillonnage

Afin de rester fidèle à la méthode d'échantillonnage utilisée pour générer VAL_Set_2018, les jeux de données VAL_Set_2019/20 et TEST_Set_2019/20 ont également été constitués par un échantillonnage stratifié. Cependant, et contrairement à VAL_Set_2018, cet échantillonnage fut semi-aléatoire. Cela découle du fait que les couvertures de certaines strates ne sont pas connues.

Outre les 9 strates¹ de VAL_Set_2018, une nouvelle strate a été ajoutée. Cette dixième strate regroupe 124 points pour lesquels des changements ont été observés entre les orthophotos de 2018 et celles de 2019 ou 2020. Pour les 9 strates communes aux différents jeux de données, la même répartition des points a été respectée, à l'exception de la strate « Ombres » pour laquelle 75 points ont été rajoutés à la demande du SPW, qui jugeait le nombre de points pour cette strate, dans VAL_Set_2018, trop faible pour en tirer de conclusions. Ainsi la distribution des points pour chaque jeu de données est reprise dans la Table 1. Ce tableau reprend également les classes principales et étiquettes doubles qui constituent chacune de ces strates.

¹ Une strate peut se référer directement à une des 11 classes principales ou à un regroupement de plusieurs de ces classes principales.

Table 1 - Distribution des points photo-interprétés par strate et jeu de données

STRATE (ID / TITRE)		CLASSES	VAL_SET_2018	VAL_SET_2019	VAL_SET_2020	TEST_SET_2019	TEST_SET_2020	POIDS 2018	POIDS 2019/20
100	Revêtement artificiel au sol	1, 3, 11, 31, 51, 71, 73, 81, 83, 91, 93	181	181	181	181	181	1,2	2,85
200	Constructions artificielles hors sol	2, 62	96	96	96	96	96	2,6	2,01
300	Eau	5, 55, 75, 85, 95	59	59	59	59	59	0,6	1,21
400	Couvert herbacé toute l'année	7	526	526	526	526	526	4,8	5,44
500	Couvert herbacé en rotation dans l'année	6	218	218	218	218	218	13,3	10,46
600	Sols nus	4	89	89	89	89	89	1,4	0,67
700	Feuillus	9, 90, 29, 39, 59	103	103	103	103	103	21	22,21
800	Résineux	8, 80, 28, 38, 58	189	189	189	189	189	5,2	3,73
900	Ombres	Toutes	50	125	125	125	125	15	8,12
1000	Changements	Toutes	0	124	124	124	124	N/A	0,05
			1511	1710	1710	1710	1710		

Pour les 8 premières strates (i.e. : strates 100 à 800, incluse), obtenues sur base de WAL_OCS_2018, l'échantillonnage a été fait aléatoirement au sein de chacun d'entre elles. Cependant les points se qualifiant pour les strates « Ombres » ou « Changements » n'ont pas été retenus, sauf pour les points situés dans une partie ombragée d'une forêt. De plus, afin de tenir compte de la précision souhaitée de 1 mètre, les points se trouvant à la bordure de deux classes ont été disqualifiés si la distance qui les séparait de la frontière était de moins de 1 mètre.

Concernant les strates « Ombres » et « Changements », leurs points ont été sélectionnés semi-aléatoirement puisqu'il n'était pas possible de dériver leurs recouvrements respectifs sur base de WAL_OCS_2018. Les rôles de ces strates sont de tester la performance de la solution sur les zones ombragées (strate 900) ainsi que d'évaluer la capacité du modèle à détecter et corriger les changements survenus entre 2018 et 2019/20 (strate 1000). L'évaluation de la capacité du modèle à détecter et corriger les erreurs commises dans WAL_OCS_2018 pourra se faire au moyen de VAL_Set_2018.

Afin d'accorder à chaque point une importance relative à la couverture de la strate sur le territoire wallon, un poids leur a été assigné. Ces poids ont été calculés avec la même méthodologie que celle employée pour le projet WALOUS (Card 1982), c'est-à-dire en calculant le ratio entre la surface de la strate dans laquelle le point a été sélectionné et la surface totale du territoire wallon ainsi que le nombre de points sélectionnés dans cette strate, voir Equation 1.

$$p_{ij} = \frac{1}{n_j} \times \frac{S_j}{\sum_k S_k}$$

Equation 1 - Détermination du poids du point i de la strate j

Où p_{ij} est le poids du point i de la strate j , n_j le nombre de points dans la strate j , S_j est la surface de la strate j et son dénominateur représente la somme des surfaces de toutes les strates (i.e.: surface de la Wallonie)

La couverture de chacune des strates se base sur les conclusions du rapport « Validation de la carte d'occupation du sol de Wallonie 2018 » du projet WALOUS (Service Public de Wallonie (SPW) - Walous 2020), et plus précisément les surfaces cartographiées par classe de WAL_OCS_2018. La couverture de la strate 1000 se base sur les résultats d'une labélisation manuelle sur des zones sélectionnées aléatoirement.

2.2. Photo-interprétation

Chacun des nouveaux jeux de données de validation (VAL_Set_2019 et VAL_Set_2020) et d'évaluation (TEST_Set_2019 et TEST_Set_2020) contient 1710 points qui ont été photo-interprétés, dans un premier temps, par AEROSPACELAB dans les règles de l'art. La photo-interprétation a été réalisée pour la classe visible uniquement : les doubles labels n'ont donc pas été pris en compte. En tout, cela représente donc 6840 photo-interprétations.

Cette photo-interprétation humaine vise à assigner à chaque point du jeu de données une classe sur base d'une image. Le point est donc représenté par le pixel le contenant et une classe unique définit ce pixel. Dans le cadre de ce marché, cette tâche s'est faite principalement sur base des orthophotos du SPW, mais également grâce aux données SENTINEL-2 et GOOGLE STREET VIEW. Elle fut confiée à membre de l'équipe d'AEROSPACELAB, l'annotateur, qui fut la seule personne, au sein d'AEROSPACELAB, à réaliser cette tâche dans son entièreté.

La méthodologie suivie par l'annotateur pour cette photo-interprétation fut la suivante :

- a) Dans un premier temps, la photo-interprétation était entreprise sur base de la paire WAL_OCS_2018 et orthophoto². Alors que l'orthophoto était choisie en fonction de l'année à labéliser, la carte d'occupation des sols de la Wallonie de 2018 fut toujours utilisée en support. Le pixel à labéliser est centré sur les deux images. Celui-ci est mis en évidence par un carré d'un mètre de côté et doté d'une bordure de couleur blanche (WAL_OCS_2018) ou rouge (orthophoto). Le contexte fourni par ce carré est donc de 1 m² ce qui bien souvent ne permettait

² En format RGB.

pas une labélisation fiable. C'est pourquoi le contexte offert par le reste de l'orthophoto et WAL_OCS_2018 fut crucial dans ce processus de labélisation. Ainsi, la surface au sol couverte par chacune des images est de 1.6384 ha (128 m de côté), et cela dans leur résolution spatiale native.

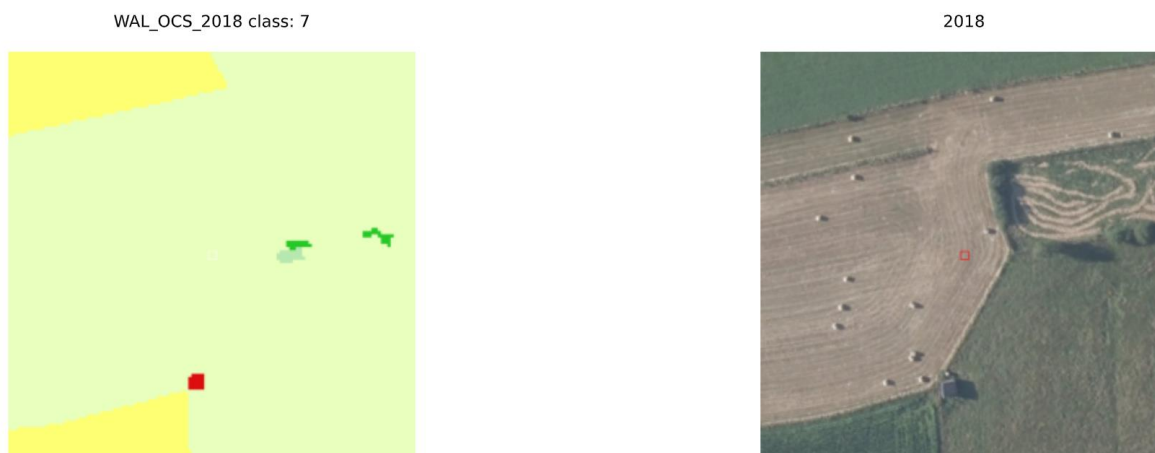


Figure 2 - Exemple de pair WAL_OCS_2018 et orthophoto utilisée en premier recours pour la photo-interprétation

- b) Dans certains cas, et plus particulièrement pour les classes « Couvert herbacé permanent », « Couvert herbacé en rotation dans l'année », « Sols nus », « Résineux », et « Feuillus », l'orthophoto n'offrait pas assez d'information pour pouvoir précisément discerner la classe à laquelle le pixel devait être associé. Dès lors d'autres sources de données furent utilisées afin d'apporter plus de contexte. La première alternative fut souvent de considérer les images disponibles sur GOOGLE STREET VIEW. Cependant, celles-ci présentent de nombreuses limitations. En effet, puisque ces images sont capturées depuis le toit d'une voiture, il faut que le point à labéliser soit suffisamment proche d'une route pour que celui-ci puisse être aperçu sur les images. De plus, la vue ne doit également pas être obstruée par de la végétation, une construction artificielle, ou un masque de floutage. Finalement, la date de capture de l'image peut ne pas correspondre avec l'année à laquelle notre jeu de données se réfère. Et puisque l'historique de ces images n'est pas accessible gratuitement, il se peut qu'aucune image exploitable ne soit accessible.



Figure 3 - Exemple d'une image GOOGLE STREET VIEW prise en aout 2019

- c) Si les images GOOGLE STREET VIEW n'étaient pas exploitables ou pas suffisantes pour labéliser avec certitude le point, l'alternative suivante fut de considérer les images aériennes de GOOGLE EARTH. L'avantage de cette source c'est qu'elle offre un historique et nous permet donc de remonter dans le temps afin d'éventuellement trouver des images pour l'année souhaitée.
- d) Une dernière alternative fut l'utilisation des données SENTINEL-2 via le portail SENTINEL-HUB³. Cependant, les résolutions spatiales de 10 mètres par pixel ne permettent de labéliser que des points faisant parties de larges objets tels que les parcelles agricoles, prairies, et forêts. Les bandes RGB et la dimension temporelle des données SENTINEL-2 furent principalement utilisées pour faire l'interprétation.

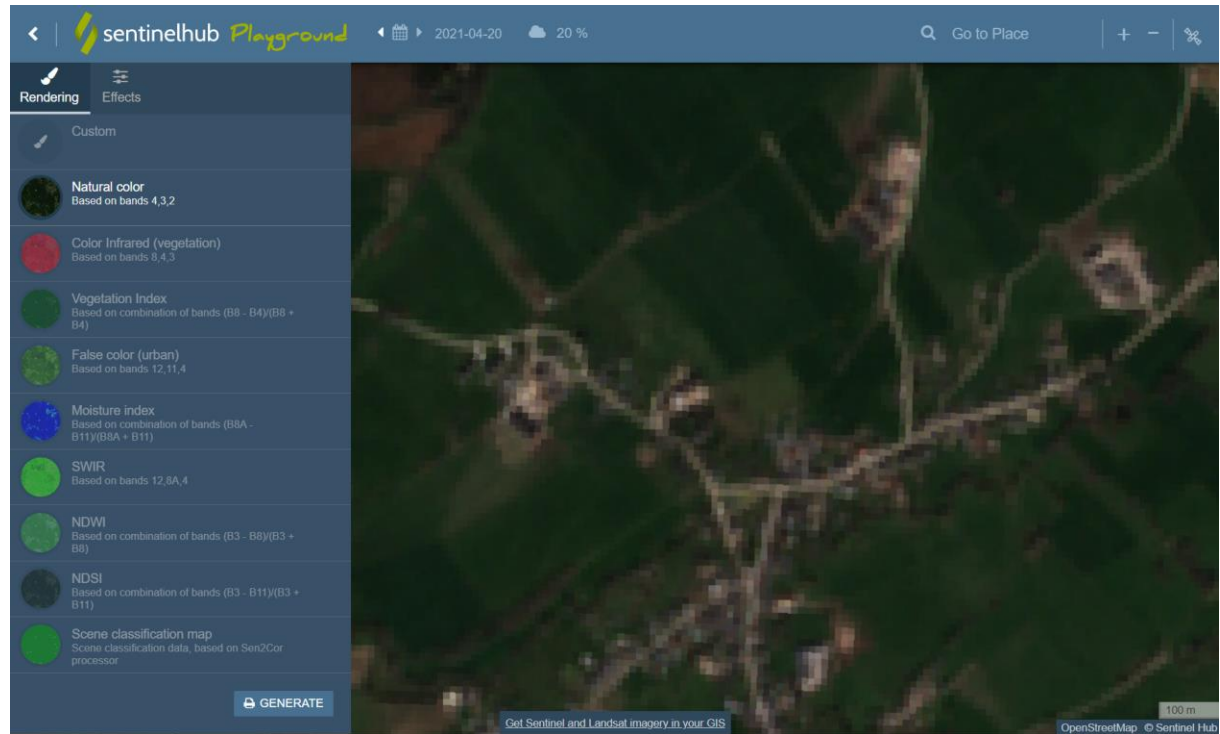


Figure 4 - Portail SENTINEL-HUB PLAYGROUND

Comme mentionné plus haut, il est important de préciser que certaines strates sont difficiles à distinguer sur base des orthophotos. C'est le cas par exemple pour les « Sols nus » et les « Revêtements artificiels au sol », les couverts herbacés en rotation ou non, ainsi que les « Feuillus » et les « Résineux ». Cette tâche de photo-interprétation s'est énormément complexifiée pour les données 2020 où la sécheresse du sol sur les orthophotos rend ardu la distinction des strates susmentionnées.

2.3. Contre-validation

Afin de contrôler la photo-interprétation d'AEROSPACELAB, les nouveaux points ont été, une seconde fois et de manière indépendante, photo-interprétés par un expert dans le domaine du SPW.

Initialement, par suite d'une mauvaise compréhension de la définition des deux strates de couverts herbacés, la photo-interprétation faite par le SPW a dû être reconduite pour les points de ces strates.

Les cas de désaccords entre les deux photo-interprétations ont été tranchés par une autre personne du SPW, avec l'aide pour les classes liées à l'agriculture des deux personnes du CRA-W et d'une personne du consortium WALOUS 2018. Cette contre-validation a permis de quantifier la fiabilité des métriques de performance du modèle lorsqu'il est évalué sur le VAL_Set ou le TEST_Set. En effet, la marge d'erreur pour ces métriques est de 1.6%, en pondérant les points suivant leur strate et en faisant l'hypothèse que, pour les points en désaccord, la photo-interprétation d'AEROSPACELAB est correcte dans la moitié des cas.

³ <https://apps.sentinel-hub.com/sentinel-playground/> et <https://apps.sentinel-hub.com/eo-browser/>

L'étape de contre-validation a également permis de fixer la classe à associer à certains lieux spécifiques, tels que :

- Terrain de tennis (peu importe le revêtement) : « Revêtement artificiel au sol »
- Terrain de football (gazon synthétique) : « Revêtement artificiel au sol »
- Terrain de football (gazon naturel) : « Couvert herbacé permanent »
- Vignoble : « Couvert herbacé en rotation »

3. Résultats

Les figures suivantes montrent la distribution des points pour VAL_Set_2019 et le TEST_Set_2019:

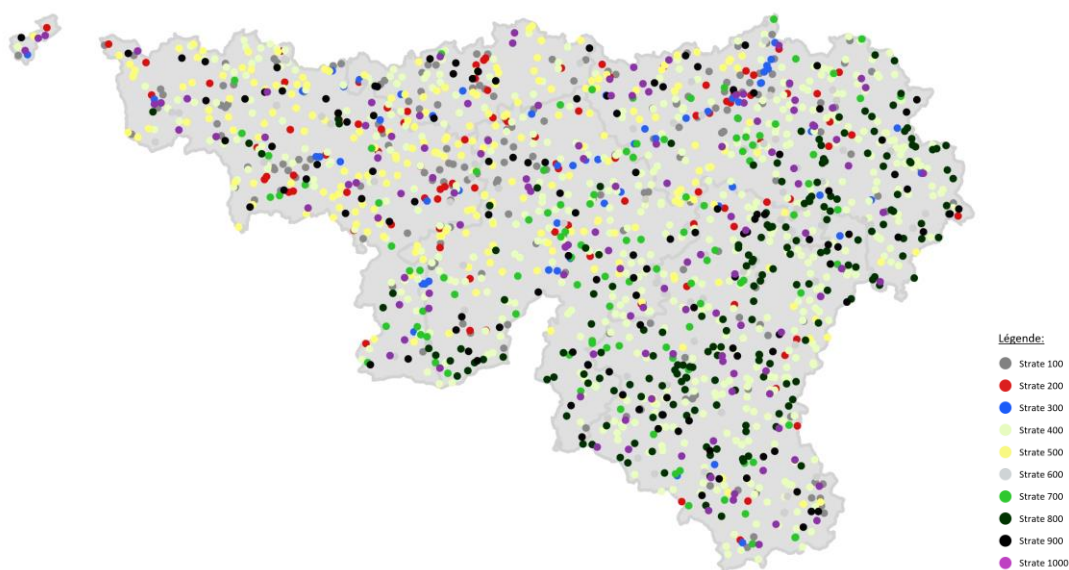


Figure 5 - Distribution des points du VAL_Set_2019



Figure 6 - Distribution du TEST_Set_2019

4. Références

- Card, Don H. 1982. «Using known map category marginal frequencies to improve estimates of thematic map accuracy.» *Photogrammetric Engineering and Remote Sensing* (NASA Ames Research Center) 48: 431-439. https://www.asprs.org/wp-content/uploads/pers/1982journal/mar/1982_mar_431-439.pdf.
- Serice Public de Wallonie (SPW) - Walous. 2019. «Guide méthodologique finalisé et recommandations d'usages pour la fusion des données de la carte d'occupation du sol de Wallonie 2018.» 29. https://geoportail.wallonie.be/files/PDF/annexes_walous/WALOUS_WP5_D5.1_Fusion.pdf.
- Service Public de Wallonie (SPW) - Walous. 2020. «Validation de la carte d'occupation du sol de Wallonie 2018.» 7. https://geoportail.wallonie.be/files/PDF/annexes_walous/WALOUS_WP6_validation.pdf.