



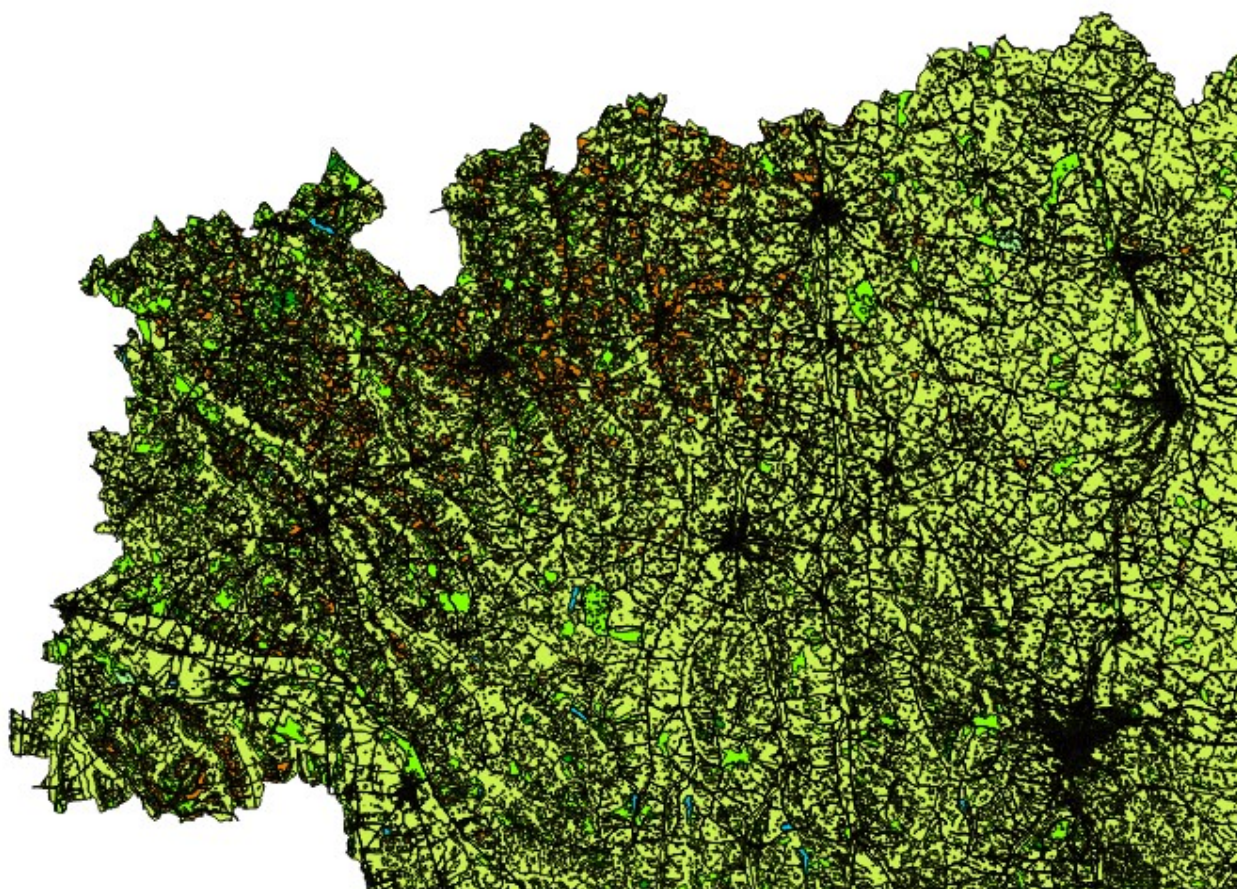
INSTITUT NATIONAL
DE L'INFORMATION
GÉOGRAPHIQUE
ET FORESTIÈRE

***OCS GE DANS LE CADRE
DU DISPOSITIF NATIONAL DE MESURE
DE L'ARTIFICIALISATION DES SOLS***

Rapport de qualification

Département du Gers (32)

Livraison du 15 mars 2022



SOMMAIRE

I- PRESENTATION	3
I -1 LE PRODUIT OCS GE	3
I -2 LA NOUVELLE METHODE DE PRODUCTION	5
I -3 POURQUOI ET COMMENT QUALIFIER LES DONNEES	5
II- INFORMATIONS DE PRODUCTION	6
II -1 CONTEXTE DE LA PRODUCTION	6
II -2 GENEALOGIE DES SOURCES	6
II -3 COMPTAGES	6
II-3.1 DONNEES INITIALES 2019	6
II-3.2 MISE A JOUR	9
II -4 MODELE IA UTILISE POUR LE MILLESIME INITIAL	9
III- QUALIFICATION DU PRODUIT LIVRE	11
III -1 COHERENCE LOGIQUE DES DEUX MILLESIMES	11
III-1.1 COHERENCE DE FORMAT	11
III-1.2 COHERENCE CONCEPTUELLE	11
III-1.3 COHERENCE DES DOMAINES DE VALEUR	12
III-1.4 COHERENCE TOPOLOGIQUE	12
III-1.5 CONCLUSION	13
III -2 PRECISION THEMATIQUE DU MILLESIME INITIAL	14
III-2.1 PRECISION THEMATIQUE DE LA COUVERTURE	15
III-2.2 PRECISION THEMATIQUE DE L'USAGE	18
III -3 EXHAUSTIVITE DU DIFFERENTIEL	20
III -4 PRECISION THEMATIQUE DES DONNEES MISES A JOUR	21
IV- BILAN	22
V- ANNEXE : QUALIFICATION DU MODELE IA	23
V -1 LIMITES DE LA QUALIFICATION	23
V -2 METHODOLOGIE : VALIDATION CROISEE PAR LA METHODE « K-FOLD »	23
V -3 INFORMATIONS GENERALES SUR LE MODELE	25
V -4 METRIQUES SUR LA CLASSIFICATION DETAILLEE	26
V -5 METRIQUES SUR LES CLASSES REGROUPEES	29
V -6 CONCLUSION : FORCES ET LIMITES D'UTILISATION DU MODELE IA	31

I- Présentation

I -1 Le produit OCS GE

Le référentiel OCS GE (occupation du sol à grande échelle) est une base de données vectorielle de description de la couverture et de l'usage du sol qui vise à couvrir l'ensemble du territoire métropolitain et des départements et régions d'outre-mer (DROM).

Les spécifications de l'OCS GE ont été élaborées par un groupe de travail piloté par le CNIG. Les principes de l'OCS GE sont les suivants :

- Base de données géographiques décrivant la couverture et l'usage des sols sur le territoire national.
 - couverture : que voit-on ?
 - usage (ou fonction principale) : à quoi ça sert ?
- Partition du territoire constituée de polygones (aucun vide, aucun recouvrement).
- Nomenclature hiérarchique, emboîtée : possibilité de détailler plus ou moins le territoire.
- Production millésimée sur une référence image.
- Compatibilité Inspire¹ et prescriptions nationales pour la production de données Occupation du sol Grande Echelle du CNIG².
- Ossature socle basée sur le réseau routier et ferré principal de la BD TOPO.
- Seuils surfaciques (500m² en zone urbaine, 2500m² en zone naturelle, agricole et 5000 m² en zone forestière; 200m² pour les zones bâties), et certains seuils de largeur.
- La mise à jour ne doit montrer que les vrais changements sur le terrain, pas du bruit de mesure ni des différences d'interprétations.



*Figure 1 : OCS GE est une partition du territoire.
Chaque polygone est renseigné en deux dimensions : couverture (à gauche) et usage (à droite)*

¹ Transposée par ordonnance n° 2010-1232 du 21 octobre 2010

² Conseil National de l'Information Géographique

OCS GE SOCLE COUVERTURE DU SOL (CS)

CS 1 Sans végétation	CS 1.1 Surfaces anthropisées	CS 1.1.1 Zones imperméables	CS 1.1.1.1 Zones bâties
			CS 1.1.1.2 Zones non bâties
		CS 1.1.2 Zones perméables	CS 1.1.2.1 Matériaux minéraux
			CS 1.1.2.2 Matériaux composites
	CS 1.2 Surfaces naturelles		CS 1.2.1 Sols nus
			CS 1.2.2 Surfaces d'eau
CS 1.2.3 Névés et glaciers			
CS 2 Avec végétation	CS 2.1 Végétation ligneuse	CS 2.1.1 Formations arborées	CS 2.1.1.1 Feuillus
			CS 2.1.1.2 Conifères
			CS 2.1.1.3 Mixte
		CS 2.1.2 Formations arbustives, sous-arbrisseaux	CS 2.1.3 Autres formations ligneuses
			CS 2.2 Végétation non ligneuse
	CS 2.2.2 Autres formations non ligneuses		

UMI | • Zones bâties : 200 m² • Intérieur Zone construite : 500 m²
• Hors Zone construite : 2 500 m²

OCS GE SOCLE USAGE DU SOL (US)

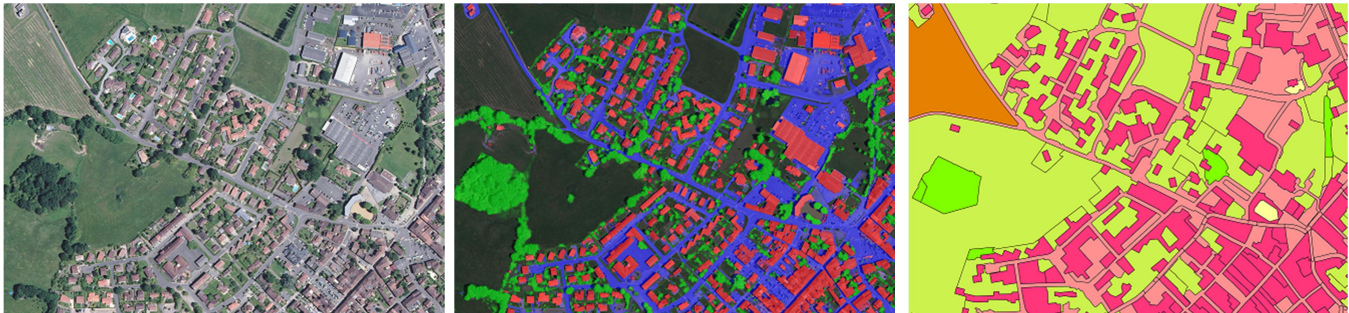
US 1 Production primaire		US 1.1 Agriculture
		US 1.2 Sylviculture
		US 1.3 Activité d'extraction
		US 1.4 Pêche et aquaculture
		US 1.5 Autres prod. primaires
US 2 Production secondaire		US 2 Production secondaire
US 235 Usage mixte		US 235 (US 2, US 3, US 5)
US 3 Production tertiaire		US 3 Production tertiaire
US 4 Réseaux de transport et d'utilité publique, logistique	US 4.1 Réseaux de transport	US 4.1.1 Routier
		US 4.1.2 Ferré
		US 4.1.3 Aérien
		US 4.1.4 Navigable
		US 4.1.5 Autres
		US 4.2 Services logistiques et de stockage
		US 4.3 Réseaux d'utilité publique
US 5 Usage résidentiel		US 5 Usage résidentiel
US 6 Autre usage		US 6.1 Zones en transition
		US 6.2 Zones abandonnées
		US 6.3 Sans usage
		US 6.6 Inconnu

Produit conforme aux préconisations nationales du CNIG

Figure 2 : Nomenclature hiérarchique, couverture et usage

I -2 La nouvelle méthode de production

Les photos aériennes sont initialement traitées et classifiées par intelligence artificielle (Deep Learning ou apprentissage profond). Le résultat de la classification, appelé la « carte de chaleur », est ensuite vectorisé pour répondre aux spécifications de l'OCS GE. Les données des bases existantes (BD TOPO, RPG, Fichiers fonciers, BD Forêt) sont ensuite intégrées pour compléter la couverture et l'usage. Les données ainsi traitées sont corrigées automatiquement et finalement photo-interprétées, avec un guidage par des alertes, pour atteindre le niveau de qualité attendu pour l'OCS GE.



PVA

Carte de chaleur vectorisée

OCS GE

Figure 3 : La chaîne de traitement résumée en trois images

A cette production d'un premier millésime, s'ajoute la production d'un second millésime par recherche des zones de changement en exploitant aussi des bases existantes, et une comparaison des cartes de chaleur produites sur chacun des millésimes. Le second millésime est mis à jour de manière incrémentale à partir du premier pour conserver une stabilité globale de l'information.

I -3 Pourquoi et comment qualifier les données

Le présent rapport permet d'« objectiver » la qualité des données. Il permettra de :

- Fournir aux utilisateurs une information transparente pour fiabiliser l'utilisation qu'ils feront des données ; ceci est d'autant plus nécessaire que la production est innovante et largement automatisée.
- Dégager les forces et les limites d'utilisation des données.
- Dégager des pistes d'amélioration.

Une évaluation est toujours relative : ses conclusions ne sont jamais valides dans l'absolu mais en rapport avec des choix et des hypothèses initiales (un cadre de référence).

La qualification s'effectue selon deux méthodes, selon « deux visions de la réalité » :

- **Méthode d'évaluation indirecte** : qualification du modèle IA ayant permis la production des cartes de chaleur, avec des informations au pixel. On trouvera en annexe de nombreux indicateurs relatifs au modèle IA.
- **Méthode d'évaluation directe** : qualification du produit final, vecteur, de sa conformité aux spécifications.
 - *Cohérence logique* : contrôles automatisés sur l'ensemble des données (conformité au schéma de données, topologie, etc.).
 - *Exhaustivité et précision thématique* (contrôle de la justesse de classement de la couverture et de l'usage, de l'exhaustivité des zones de changement entre millésimes) : par échantillonnage aléatoire, vérification par des experts et inférence statistique.

II- Informations de production

II -1 Contexte de la production

Le département du Gers est une production « pilote » qui permet de consolider toutes les phases de production, depuis la saisie des annotations en amont jusqu'aux contrôles des photo-interprètes. Ce département est produit sur une chaîne de production déjà fiabilisée et qui continuera à être améliorée avec des itérations successives.

Le présent rapport est aussi pilote, comme les données : sa forme et son contenu pourraient être amenés à évoluer selon les échanges avec le commanditaire et les utilisateurs des données, selon l'évolution des procédures de contrôle, et selon la capacité à le générer de manière automatisée dans un contexte de production massive.

II -2 Généalogie des sources

Département du Gers (32)

	Millésime initial	Millésime mis à jour
<i>Année</i>	2019	2016
<i>Date de prise de vue aérienne (PVA)</i>	Du 01/06/2019 au 17/06/2019	Du 14/08/2016 au 23/08/2016
<i>Résolution de la PVA de référence</i>	20 cm	20 cm
<i>Actualité des données complémentaires</i>		
<ul style="list-style-type: none"> • <i>BD UNI (source de la BD TOPO®)</i> • <i>RGFor (source de la BD FORÊT®)</i> • <i>RPG</i> • <i>Fichiers fonciers DGFIP/CEREMA</i> 	17/06/2019 (routier/ferré) et 04/03/2021 (autres classes) 2011 2019 2020 (plus proche de la réalité de la PVA 2019)	16/12/2016 (routier/ferré) et 27/06/2018 (autres classes) 2011 2016 2017

II -3 Comptages

II-3.1 Données initiales 2019

Surface couverte : ~ 6 300 km².

Le Gers est un département largement dominé par les terres agricoles.

a) Couverture

	nb objets	% objets	surface km ²	% surface	surface km ²	% surface
Zones bâties	102538	40.6%	55.23	0.9%	171.64	2.7%
Zones non bâties	23929	9.5%	103.51	1.6%		
Matériaux minéraux	5083	2.0%	12.78	0.2%		
Matériaux composites	23	0.0%	0.12	0.0%		
Sols nus	44	0.0%	0.28	0.0%	52.10	0.8%
Surfaces d'eau	3569	1.4%	51.81	0.8%		
Névés et glaciers	0	0.0%	0.00	0.0%		
Feuillus	47266	18.7%	952.31	15.1%	6 079.97	96.5%
Conifères	1269	0.5%	26.30	0.4%		
Mixte	1730	0.7%	23.05	0.4%		
Formations arbustives, sous-arbrisseaux	4045	1.6%	52.58	0.8%		
Autres formations ligneuses	3744	1.5%	222.69	3.5%		
Formations herbacées	59138	23.4%	4803.04	76.2%		
Autres formations non ligneuses	0	0.0%	0.00	0.0%		
	252378	100.0%	6303.70	100.0%		

Tableau 1 : Répartition des objets du Gers pour la couverture 2019, en nombre d'objets et en surface

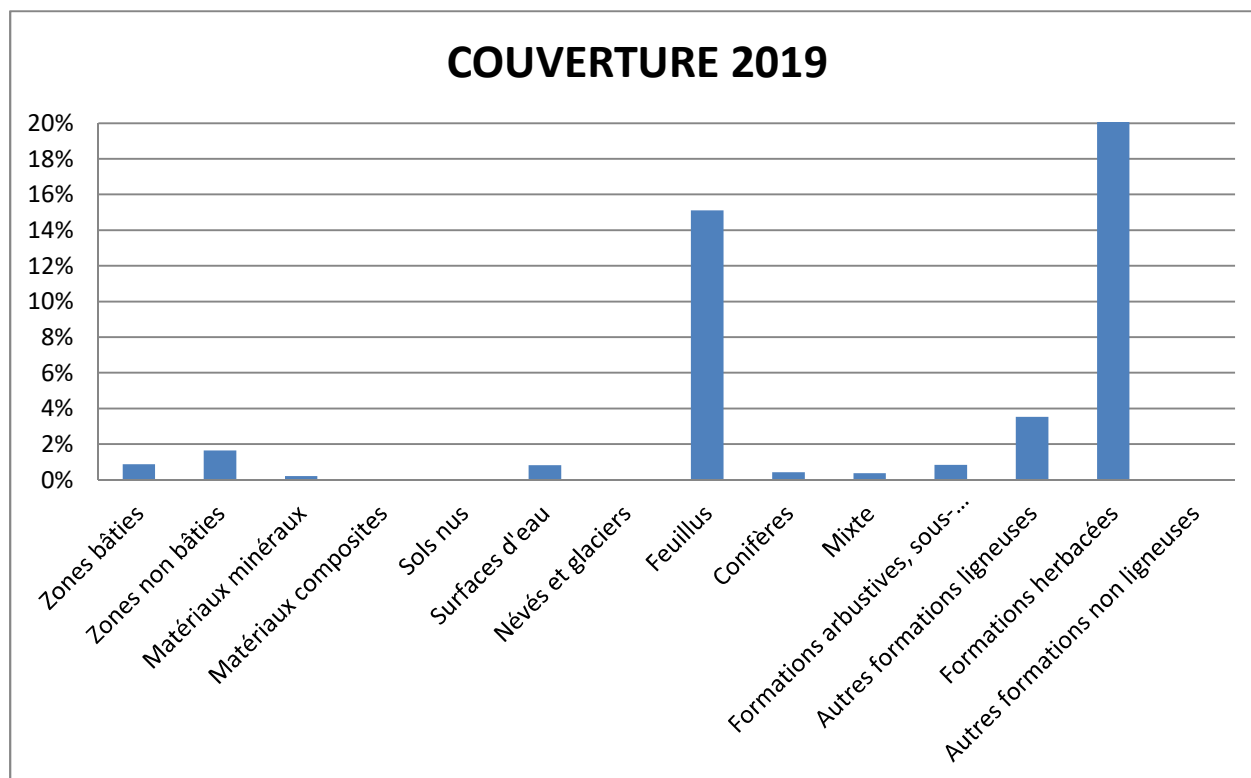


Figure 4 : Histogramme de la répartition des couvertures 2019, en surface, tronqué à 20%. Les Formations herbacées représentent 76 % du jeu en surface.

b) Usage

USAGE		nb objets	% objets	surface km ²	% surface
US1.1	Agriculture	42101	16.7%	4833.91	76.7%
US1.2	Sylviculture	16940	6.7%	865.37	13.7%
US1.3	Activité d'extraction	72	0.0%	1.47	0.0%
US1.4	Pêche et aquaculture	19	0.0%	0.08	0.0%
US1.5	Autres prod. primaire	0	0.0%	0.00	0.0%
US2	Production secondaire	764	0.3%	3.76	0.1%
US235	Usage mixte	21698	8.6%	47.85	0.8%
US3	Production tertiaire	12801	5.1%	33.23	0.5%
US4.1.1	Routier	17848	7.1%	94.79	1.5%
US4.1.2	Ferré	323	0.1%	0.76	0.0%
US4.1.3	Aérien	38	0.0%	1.22	0.0%
US4.1.4	Navigable	4	0.0%	0.02	0.0%
US4.1.5	Autres	0	0.0%	0.00	0.0%
US4.2	Services logistiques et de stockage	19	0.0%	0.14	0.0%
US4.3	Réseau d'utilité publique	848	0.3%	3.25	0.1%
US5	Usage résidentiel	117579	46.6%	240.93	3.8%
US6.1	Zones en transition	420	0.2%	1.63	0.0%
US6.2	Zones abandonnées	355	0.1%	0.83	0.0%
US6.3	Sans usage	20520	8.1%	174.40	2.8%
US6.6	Inconnu	29	0.0%	0.04	0.0%
TOTAL		252378	100%	6303.70	100.0%

Tableau 2 : Répartition des objets du Gers pour l'usage 2019, en nombre d'objets et en surface

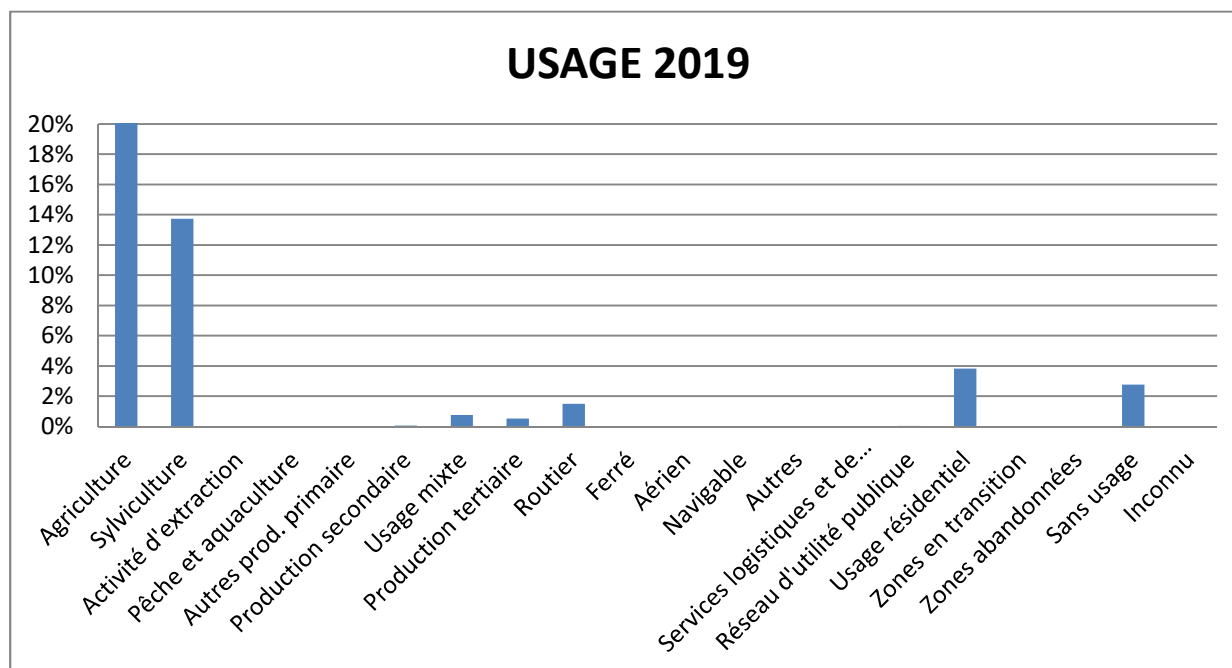


Figure 5 : Histogramme de la répartition des usages 2019, en surface, tronqué à 20 %.
L'usage Agricole représente 77 % du jeu en surface.

II-3.2 Mise à jour

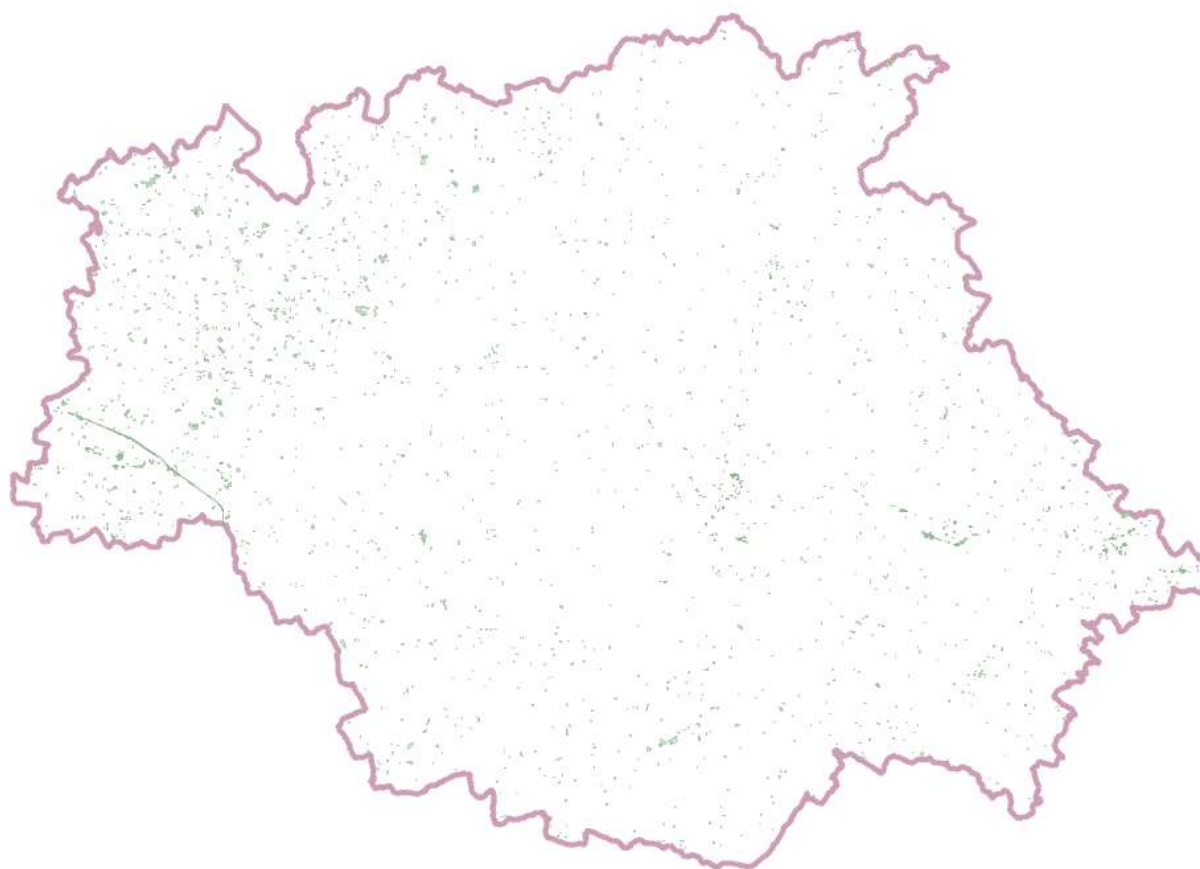


Figure 6 : Visualisation des mises à jour 2019 vers 2016

9274 objets du millésime initial 2019 ont été mis à jour, totalisant une surface d'environ 25 km². Ceci représente 3.7 % de changement en nombre d'objets, et 0.4 % en surface.

II -4 Modèle IA utilisé pour le millésime initial

Le modèle IA a appris sur des annotations du département du Gers uniquement. Les prochains modèles apprendront sur des annotations France entière : cette différence sera à prendre en compte lors des futures comparaisons.

Le modèle IA sert à renseigner la couverture. Les résultats de l'IA sont relatifs à des cartes de chaleur ; les indicateurs renseignés dans ce paragraphe sont calculés sur des pixels, avant la vectorisation, les post-traitements et la photo-interprétation.

Indicateurs clés :

Surface des annotations	150 km ² (exclusivement sur le département du Gers).
Millésime de la PVA des annotations	2019
Architecture du modèle	DeepLab v3 Plus (~ 7 millions de paramètres)
Métrique OA (overall accuracy) : mesure le taux de pixels bien classés <ul style="list-style-type: none"> • Pour une classification fine (15 classes de couverture) • Pour une classification agrégée (3 classes : surfaces anthropisées, surfaces naturelles sans végétation, surfaces avec végétation) 	OA = 83 % OA = 97 %

L'indicateur « OA » sera fortement influencé par les classes fortement représentées (cultures, forêt).

L'annexe du présent rapport offre de nombreux indicateurs complémentaires. Ils ont permis d'établir les conclusions suivantes :

Forces et limites d'utilisation du modèle

Le modèle obtient des résultats limités pour les classes sur lesquelles les données de référence étaient trop faiblement représentées (par exemple les sols nus ou les coupes). Le modèle est également moins performant pour les zones perméables, dont la délimitation est floue ou la variabilité élevée. Enfin, on trouve des confusions entre feuillus, conifères ou broussailles, liées probablement à des confusions dans les annotations elles-mêmes, qui gagneraient à être améliorées par des experts forestiers.

Malgré ces rares écueils, le modèle **obtient de très bons résultats** : plus de 83 % des pixels sont correctement prédits avec la classification la plus fine, et plus de 97 % avec la classification agrégée, ce dernier chiffre étant certes favorisé par l'excellent score de la végétation, très majoritaire sur ce département. Plusieurs classes parviennent à obtenir des indicateurs qui frôlent ou dépassent les 80 %, comme les bâtiments, les surfaces d'eau ou les cultures, bien que cette classe ait souvent une délimitation floue et une forte variabilité. Ces chiffres valident la fiabilisation de la chaîne IA, permettent de confirmer les orientations IA prises pour la production des cartes chaleur, et sont également rassurants pour la suite des traitements de la chaîne de production OCS GE (vectorisation, croisement avec d'autres sources, photo-interprétation guidée par des alertes) en permettant de limiter la photo-interprétation pour la couverture.

III- Qualification du produit livré

Les critères qualité utilisés sont inspirés de la norme ISO 19157 « Information géographique – Qualité des données ».

III -1 Cohérence logique des deux millésimes

La cohérence logique se définit comme le degré de cohérence interne des données selon les règles de modélisation et de spécification du produit. Elle se décompose en quatre sous-critères :

- Cohérence de format : adéquation avec la structure physique du jeu de données. Les données peuvent-elles être lues dans un SIG ?
- Cohérence conceptuelle : respect du schéma conceptuel de données.
- Cohérence des domaines de valeurs : appartenance des valeurs d'attributs aux plages de valeurs spécifiées (respect de la nomenclature couverture / usage).
- Cohérence topologique : exactitude des caractéristiques topologiques du jeu de données.

Les contrôles sont effectués sur l'intégralité du jeu de données.

La spécification utilisée est celle de l'OCS GE v1.1 produite de manière classique, disponible sous ce lien : https://geoservices.ign.fr/sites/default/files/2021-07/DC_OCS_GE_1-1.pdf . La spécification sera modifiée pour permettre le détail de l'usage US235 avec les valeurs US2, US3 ou US5.

Les taux d'erreurs indiqués sont en nombre d'objets (et non en surface), relativement au nombre total d'objets dans le jeu, pour chaque millésime.

III-1.1 Cohérence de format

Contrôle	Millésime 2019	Millésime 2016
Format des fichiers : shapefile	OK	OK
Présence du fichier de projection	OK	OK
Ouverture dans un SIG (QGIS 3.16)	OK	OK

III-1.2 Cohérence conceptuelle

Pour le respect des UMI, les erreurs sont comptées une et une seule fois (par exemple si un objet a été comptabilisé comme non conforme pour le seuil de 200 m², il ne sera pas recompté pour les autres seuils).

Contrôle	Millésime 2019	Millésime 2016
Conformité au schéma de données	Conforme	Conforme
Pas de faces adjacentes de mêmes couverture et usage hors ossature	238 objets non conformes Taux d'erreur : 0.09 %	137 objets non conformes Taux d'erreur : 0.06 %

Respect des UMI : seuil 200 m ² (pas d'objet de surface < 200 m ²) sauf ossature	917 objets non conformes Taux d'erreur : 0.36 %	785 objets non conformes Taux d'erreur : 0.31 %
Respect des UMI : seuil 500 m ² (pas d'objet de surface inférieure à 500 m ² sauf zones bâties) sauf ossature	333 objets non conformes Taux d'erreur : 0.13 %	361 objets non conformes Taux d'erreur : 0.14 %
Respect des UMI : seuil 2500 m ² (pas d'objet de surface inférieure à 2500 m ² hors zones construites) sauf ossature (les surfaces < 2500 m ² sont acceptées si elles touchent la zone construite, avec un partage de géométrie)	523 objets non conformes Taux d'erreur : 0.21 %	531 objets non conformes Taux d'erreur : 0.21 %
Respect des UMI : seuil 5000 m ² (pas d'objet d'usage US1.2 de surface inférieure à 5000 m ²)	81 objets non conformes Taux d'erreur : 0.03 %	78 objets non conformes Taux d'erreur : 0.03 %

III-1.3 Cohérence des domaines de valeur

Contrôle	Millésime 2019	Millésime 2016
Nomenclature couverture et usage respectée	Conforme	Conforme
Croisements couverture / usage autorisés	11 cas non conformes Taux d'erreur < 0.01 %	15 cas non conformes Taux d'erreur < 0.01 %
Valeurs autorisées des différents champs autres que couverture / usage	Cohérence OSSATURE/CODE_OR/ID_ORIGINE/SOURCE : 134 objets non conformes Taux d'erreur : 0.05 %	Cohérence OSSATURE/CODE_OR/ID_ORIGINE/SOURCE : 160 objets non conformes Taux d'erreur : 0,06 %

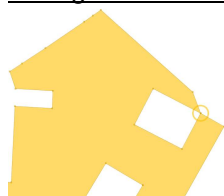
III-1.4 Cohérence topologique

Contrôle	Millésime 2019	Millésime 2016
Intégrité des géométries (validité au sens de l'OGC)	596 polygones non conformes au sens de l'OGC (voir le commentaire ci-dessous) Taux d'erreur : 0.24 %	581 polygones non conformes au sens de l'OGC (voir le commentaire ci-dessous) Taux d'erreur : 0.23 %
Pas de superposition	Aucune erreur	264 micro-recouvrements (défaut de topologie)

		Taux d'erreur : 0.11 %
Pas de trou	4 trous en bord de département (effet de l'ossature) Taux d'erreur < 0.01 %	670 trous : 4 trous en bord de département (effet de l'ossature) – les autres sont liés au défaut de topologie Taux d'erreur : 0,27 %
Pas d'objet avec multi-géométrie hors ossature	1 objet hors ossature avec 2 primitives (bordure de département) Taux d'erreur < 0.01 %	2 objets hors ossature avec 2 primitives (dont un en bord de département) Taux d'erreur < 0.01 %

Commentaires

- Les géométries invalides trouvées au sens de l'OGC sont des polygones « en banane » :



La frontière extérieure est d'un seul tenant, avec un point coïncidant, faisant apparaître un trou qui n'en est pas un. Ces polygones sont valides pour les outils ESRI utilisés en production. Ils ne le sont pas au sens de l'OGC qui modéliserait ici avec un anneau extérieur et un trou. Aucune définition n'est meilleure, c'est une question de convention (<https://postgis.net/workshops/postgis-intro/validity.html> § 23.3). La validité OGC a ici été évaluée.



- Exemple de trous en bordure de département, effet d'ossature :

Un petit morceau d'ossature débordant légèrement de l'emprise génère un trou.

III-1.5 Conclusion

La cohérence logique des données est satisfaisante, avec des taux d'erreurs tous constatés sous le seuil de 0.5 %.

Des contrôles et des opérations complémentaires en fin de chaîne de production permettront de viser une conformité parfaite.

III -2 Précision thématique du millésime initial

Le critère évalué est la **justesse de classement** pour la couverture et l'usage. La **méthodologie** est la suivante :

- Au préalable, les objets de grandes surfaces (seuil de largeur de 1 km) sont découpés par dichotomie. Les surfaces trop importantes peuvent en effet biaiser les résultats.
- Un tirage aléatoire stratifié par poste US et poste CS est réalisé. 30 objets sont tirés par classe de couverture et 30 objets par classe d'usage, quand l'effectif de la classe le permet. Pour les classes comportant moins de 30 objets, l'intégralité de la population est contrôlée.
- Les polygones sont contrôlés par une nouvelle photo-interprétation. Le photo-interprète indique s'il est en désaccord avec les valeurs de couverture et usage mentionnées, propose celles qu'il juge correctes et évalue le pourcentage de désaccord.
- Un polygone est considéré comme correct dès lors que la proportion de surface bien codée a été évaluée à plus de 85%, ou dès lors que la proportion de surface mal codée a été évaluée à moins de 15%.
- Les taux d'accord sont calculés et des statistiques établies.
- **Pour la couverture**, deux taux d'accord par classe sont proposés :
 - o Pourcentage d'objets de l'échantillon dont la classe est correctement renseignée.
 - o Pourcentage d'objets de l'échantillon dont la classe appartient à un niveau supérieur dans la nomenclature emboîtée.

Le regroupement de classes proposé est le suivant :

CS1.1 : surfaces anthropisées

CS1.2 : surfaces naturelles sans végétation

CS2 : surfaces avec végétation

Pour la couverture, les attendus sont les suivants :

- o 85% des entités de l'échantillon doivent être correctement renseignées
- o 95% des entités doivent appartenir au niveau supérieur selon le regroupement ci-dessus.

Exemple : 85% des objets de l'échantillon de la classe CS1.2.1 - Sols nus doivent être validés par la recette. Et 95% doivent être dans la classe CS1.2 – Surfaces naturelles sans végétation. Un objet classé en CS1.2.1 par la production et classé lors du contrôle en surface d'eau CS1.2.2 remplit la deuxième condition sans remplir la première. Un objet classé en CS1.2.1 par la production et classé lors du contrôle en matériaux minéraux CS1.1.2.1 ne remplit aucune des deux conditions (CS1.2 vs. CS1.1).

- **Pour l'usage**, un seul taux d'accord est proposé par classe : pourcentage d'objets de l'échantillon dont la classe est correctement renseignée. Un taux d'accord de 85 % est aussi attendu pour ce niveau le plus fin.
- On applique la technique des **tests statistiques**, expliquée ici pour l'appartenance à la classe fine : on suppose qu'au moins 85 % des objets sont corrects (hypothèse nulle H0). **Apparaissent en rouge dans les tableaux**, les classes pour lesquelles l'attendu sur l'échantillon n'est pas respecté avec un niveau de confiance significatif de 95 % (valeur classique en statistique). **Autrement dit, pour les classes en rouge, on peut affirmer, avec moins de 5 % de chance de se tromper, que le jeu n'est pas conforme pour cette classe.** Pour les autres classes, on ne peut pas affirmer avec une confiance statistique suffisante la non-conformité. Enfin, quand il y a moins de 30 objets dans l'échantillon, aucune notion de confiance statistique n'intervient puisque tous les objets sont contrôlés.

Le même raisonnement est fait pour l'appartenance à la classe regroupée, et ensuite pour l'usage.

III-2.1 Précision thématique de la couverture

Les tableaux ci-dessous présentent les statistiques issues des contrôles sémantiques.

Pour chaque poste, deux résultats sont présentés :

- Le résultat des tests d'appartenance à la classe regroupée (CS1.1, CS1.2 ou CS2)
- Le résultat des tests d'appartenance à la classe fine

Code CS	Taille échantillon	Nombre objets OK	Taux d'accord	Nombre objets OK	Taux d'accord
		Classe fine		Classe regroupée	
CS1.1.1.1 Zones bâties	30	30	100%	30	100%
CS1.1.1.2 Zones non bâties	30	23	77%	27	90%
CS1.1.2.1 Matériaux minéraux	30	26	87%	28	93%
CS1.1.2.2 Matériaux composites	23	6	26%	16	70%
CS1.2.1 Sols nus	30	21	70%	21	70%
CS1.2.2 Surfaces d'eau	30	29	97%	29	97%
CS2.1.1.1 Feuillus	30	29	97%	30	100%
CS2.1.1.2 Conifères	30	28	93%	30	100%
CS2.1.1.3 Mixte	30	20	67%	30	100%
CS2.1.2 Formations arbustives et sous-arbrisseaux	30	28	93%	30	100%
CS2.1.3 Autres formations ligneuses	30	29	97%	29	97%
CS2.2.1 Formations herbacées	30	30	100%	30	100%

Tableau 3 : Contrôles sémantiques pour la couverture 2019

Conclusion

Les résultats sont conformes aux attentes ; ils sont même excellents pour les classes *Zones bâties* (qui est la classe la plus représentée en nombre d'objets) et *Formations herbacées* (qui est la classe la plus représentée en surface).

Pour le niveau fin, trois classes n'atteignent pas les objectifs de manière statistiquement significative : *Matériaux composites*, *Sols nus* et *Végétation mixte*. Pour le niveau regroupé, la *Végétation mixte* devient conforme : il y a des confusions avec les valeurs *Feuillus* et *Conifères* (voir la matrice de confusion dans la suite).

Matériaux composites et *Sols nus* sont des classes rares sur le Gers (respectivement 23 et 44 objets). Pour les *Matériaux composites*, il s'agit d'un problème de compréhension des spécifications : des dépôts de matériaux (bois, voitures.....) ont été codés en CS1.1.2.2 alors que dans ces cas, on aurait dû coder le sol sous les dépôts. Pour les *Sols nus*, on trouve des objets en méandre de rivière avec de la présence de végétation, qui aurait dû conduire à un classement végétatif.

On note aussi le score moyen des *Zones non bâties*, même si on ne peut pas affirmer avec une confiance statistique suffisante que cette valeur n'est pas conforme. Le contrôle qualité a montré des routes trop étroites, en dessous d'une largeur minimale de 5 m. Un correctif a été apporté dans les outils de production pour les prochaines versions.

Histogramme

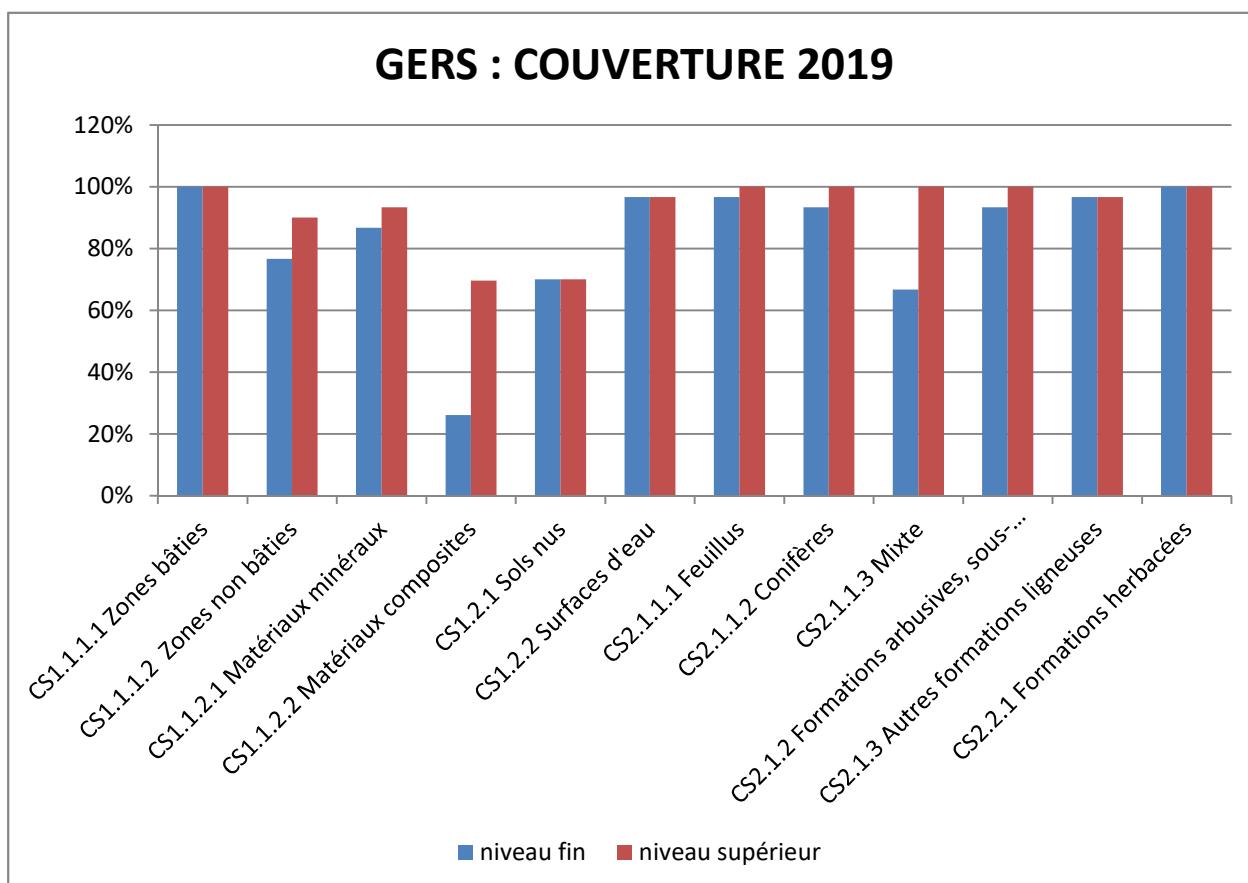


Figure 7 : Histogramme des taux d'accord pour la couverture 2019

Matrice de confusion :

Ctrl	CS1.1.1.1	CS1.1.1.2	CS1.1.2.1	CS1.1.2.2	CS1.2.1	CS1.2.2	CS2.1.1.1	CS2.1.1.2	CS2.1.1.3	CS2.1.2	CS2.1.3	CS2.2.1	All
CS1.1.1.1	30	0	0	0	0	0	0	0	0	0	0	0	30
CS1.1.1.2	0	23	2	0	0	0	0	0	0	0	0	0	25
CS1.1.2.1	0	4	26	10	2	0	0	0	0	0	0	0	42
CS1.1.2.2	0	0	0	6	0	0	0	0	0	0	0	0	6
CS1.2.1	0	0	0	0	21	0	0	0	0	0	0	0	21
CS1.2.2	0	0	0	0	0	29	0	0	0	0	0	0	29
CS2.1.1.1	0	0	0	0	2	1	29	1	8	2	0	0	43
CS2.1.1.2	0	0	0	0	0	0	1	28	2	0	0	0	31
CS2.1.1.3	0	0	0	0	0	0	0	1	20	0	0	0	21
CS2.1.2	0	0	0	0	3	0	0	0	0	28	0	0	31
CS2.1.3	0	0	0	0	0	0	0	0	0	0	29	0	29
CS2.2.1	0	3	2	7	2	0	0	0	0	0	1	30	45
All	30	30	30	23	30	30	30	30	30	30	30	30	353

Tableau 4 : Matrice de confusion pour la couverture 2019

Taux d'accord global sur l'échantillon : 299 (somme de la diagonale de la matrice) / $353 = 85 \%$.

Il s'agit d'un taux non pondéré par la représentativité réelle des classes dans le jeu.

La matrice se lit ainsi :

- En colonne, examen des données du jeu livré et indication du classement qu'elles auraient dû prendre. Cela permet de mesurer les sur-détections ; c'est l'indicateur retenu dans les tableaux précédents. C'est comparable à la mesure de précision du moteur IA.
- En ligne, les données théoriques avec les valeurs qu'elles ont reçues dans le jeu. Cela permet de mesurer les sous-détections. C'est comparable à la mesure de rappel du moteur IA.

Par exemple pour CS1.1.1.2 (zones non bâties) :

- Lecture en colonne : sur les 30 objets de l'échantillon, 23 ont été correctement classés. 4 objets ont été classés à tort en CS1.1.2.1, et 3 objets classés à tort en CS2.2.1.
- Lecture en ligne : sur les 25 objets de la vérité terrain présents dans l'échantillon, 23 ont été bien trouvés correctement classés dans le jeu. 2 se trouvent à tort classifiés en CS1.1.2.1.

On observe ici une sous-détection pour 3 classes (en rouge dans le tableau) : les matériaux minéraux (CS1.1.2.1) qui se confondent parfois avec les matériaux composites (CS1.1.2.2), les feuillus (CS2.1.1.1) qui se confondent parfois avec les mixtes (CS2.1.1.3) et les formations herbacées (CS2.1.1) qui se confondent parfois avec les matériaux composites. Des actions correctives sont en cours afin d'améliorer la détections de ces classes dans les prochaines versions de la chaîne de production. Pour le reste des classes, il y a très peu de confusions.

III-2.2 Précision thématique de l'usage

Pour l'usage, on estime seulement l'accord au niveau le plus fin de la classification.

Apparaissent en rouge les classes non conformes, avec un seuil de confiance de 95 %, au sens du test statistique (hypothèse nulle H0 : au moins 85 % des objets sont conformes).

Code US	Taille échantillon	Nombre objets OK	Taux d'accord sur l'échantillon
US1.1 Agriculture	30	29	97%
US1.2 Sylviculture	30	30	100%
US1.3 Activité d'extraction	30	27	90%
US1.4 Pêche et aquaculture	19	18	95%
US2 Production secondaire	30	25	83%
US235 Production secondaire, tertiaire et usage résidentiel	30	29	97%
US3 Production tertiaire	30	20	67%
US4.1.1 Routier	30	28	93%
US4.1.2 Ferré	30	9	30%
US4.1.3 Aérien	30	25	83%
US4.1.4 Navigable	4	3	75%
US4.2 Services logistique et de stockage	19	19	100%
US4.3 Réseau d'utilité publique	30	30	100%
US5 Usage résidentiel	30	29	97%
US6.1 Zones en transition	30	29	97%
US6.2 Zones abandonnées	30	28	93%
US6.3 Sans usage	30	25	83%
US6.6 Inconnu	29	0	0%

Tableau 5 : Contrôles sémantiques pour l'usage 2019

Conclusion :

Les chiffres sont en accord avec ce qu'on attendrait d'un produit OCS GE CNIG. Quatre valeurs obtiennent cependant des résultats perfectibles. On rappelle que l'usage est majoritairement issu des fichiers fonciers DGFIP/CEREMA.

- **US3 Production tertiaire** : Cet usage est issu des fichiers fonciers DGFIP. Il est difficile à établir par photo-interprétation ; le contrôle qualité a préféré ramener en US235 ou US1.1 quand il y avait un doute sérieux. Les consignes seront à affiner : sur cette classe, un curseur sera à trouver entre qualité et efficacité.
- **US4.1.2 Ferré** : Les cas relevés concernent une voie déclassée pour laquelle les fichiers fonciers indiquent encore à tort un usage ferré. Un meilleur croisement avec la BD TOPO devrait améliorer.
- **US4.1.4 Navigable** : Une erreur relevée, mais le jeu ne contient que 4 objets, il n'est pas significatif.
- **US6.6 Inconnu** : Il s'agit d'alertes incorrectement traitées en photo-interprétation (manque de temps ou d'expertise), les consignes seront à renforcer pour cette valeur. Classe rare et difficile à renseigner, 29 objets sont concernés, la plupart auraient dû être codés en usage mixte US235.

Histogramme :

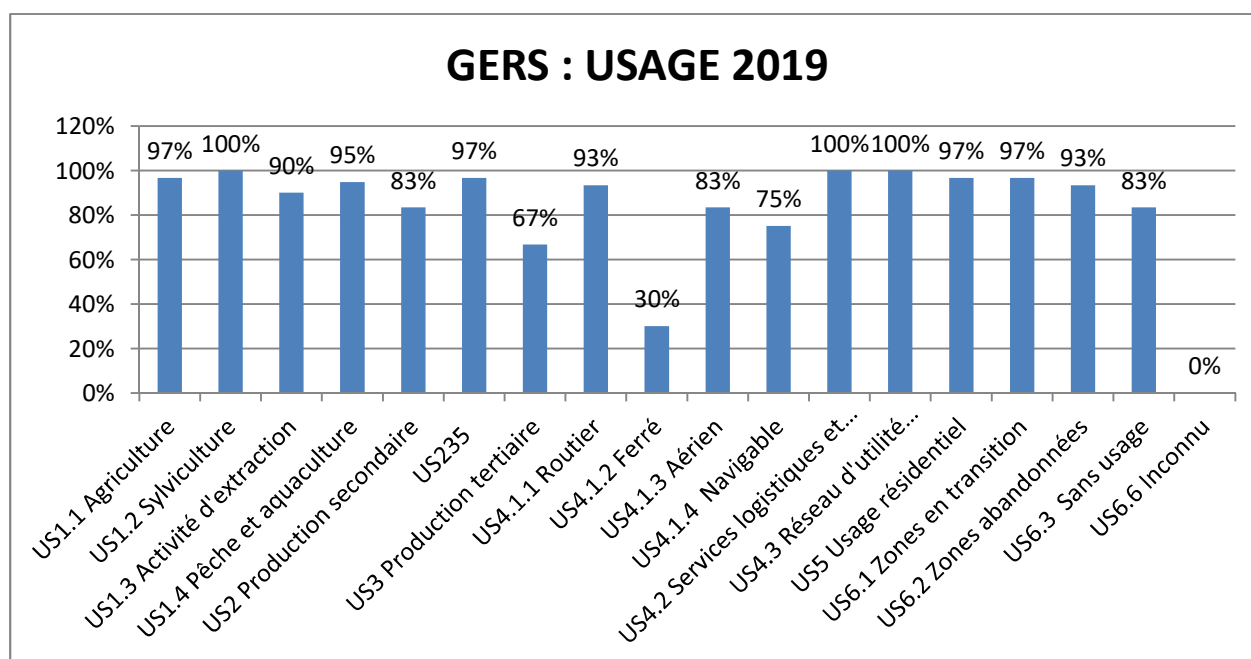


Figure 8 : Histogramme des taux d'accord pour l'usage 2019

Matrice de confusion :

Ctrl	US1.1	US1.2	US1.3	US1.4	US2	US235	US3	US4.1.1	US4.1.2	US4.1.3	US4.1.4	US4.2	US4.3	US5	US6.1	US6.2	US6.3	US6.6	All
US1.1	27	0	0	0	2	0	3	2	7	0	0	0	0	0	1	0	0	3	45
US1.2	1	30	0	0	0	0	0	0	1	0	0	0	0	1	0	1	1	0	35
US1.3	0	0	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	26
US1.4	0	0	0	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17
US2	0	0	0	0	25	0	0	0	0	0	0	0	0	0	0	0	0	0	25
US235	2	0	3	1	1	29	7	0	11	5	1	0	0	0	0	0	4	25	89
US3	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	20
US4.1.1	0	0	0	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0	28
US4.1.2	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	8
US4.1.3	0	0	0	0	0	0	0	0	0	25	0	0	0	0	0	0	0	0	25
US4.1.4	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	3
US4.2	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	0	19
US4.3	0	0	0	0	1	1	0	0	0	0	0	0	30	0	0	0	0	0	32
US5	0	0	0	0	0	0	0	0	0	0	0	0	0	29	0	0	0	0	29
US6.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	29	0	0	0	29
US6.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28	0	0	28
US6.3	0	0	1	1	1	0	0	0	3	0	0	0	0	0	0	1	25	1	33
US6.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
All	30	30	30	19	30	30	30	30	30	30	4	19	30	30	30	30	30	29	491

Tableau 6 : Matrice de confusion pour l'usage 2019

Taux d'accord global sur l'échantillon : 398 (somme de la diagonale de la matrice) / 491 = 81 %.

Il s'agit d'un taux non pondéré par la représentativité réelle des objets dans le jeu.

Les explications de lecture sont données pour la matrice de confusion de la couverture.

On observe une sous-détection de l'usage agricole US 1.1, alors que celui-ci représente déjà la grande majorité de la surface couverte par les données. L'usage US235 est également sous-détecté, puisqu'auraient dû y être classifiés des objets d'usage ferré et des objets sans usage.

III -3 Exhaustivité du différentiel

Le différentiel 2019 → 2016 est analysé. L'exhaustivité se définit comme la présence ou l'absence d'entité. On vérifie ici

- Les omissions : les changements qui ont été oubliés.
- Les excédents : les changements qui ne sont pas légitimes (fausses évolutions).

Un contrôle exhaustif a été réalisé sur une zone représentative du Gers de 360 km² (soit 5,7 % de la surface du département). Il s'agit donc d'un contrôle massif. Un opérateur a repéré manuellement les changements avec un outil dédié, sur fond d'orthophotos, sans appui de l'OCS calculée pour éviter tout biais. L'opérateur n'avait pas non plus déjà travaillé sur la zone. Ensuite ces changements ont été comparés au différentiel calculé pour déduire les métriques ci-dessous.

Contrôle	Statistique mesurée
Omissions	Taux d'éléments manquants dans le différentiel : 12,1 %
Excédents	Taux d'éléments en excès dans le différentiel : 14,2 %

Commentaire :

- **Le résultat est acceptable pour les omissions.** On rappelle que seuls 3.7 % des objets de l'OCS initiale ont bougé. Le processus « oublié » donc environ 12 % de ces 3,7 %. Une analyse complémentaire est à mener pour détecter des systématismes et prendre des mesures correctives.
- **Le résultat est aussi acceptable pour les excédents**, d'autant que 92 % de l'excédent est constitué d'objets de surface inférieures à 50 m². Il s'agit plutôt de bruit de traitements, qui pourrait être éliminé.

III -4 Précision thématique des données mises à jour

Sur le différentiel, sont tirés au hasard au moins 30 objets par classe de couverture et 30 objets par classe d'usage. Plusieurs contrôleurs vérifient les nouvelles valeurs de couverture et d'usage. La statistique indique le pourcentage d'objets dans l'échantillon pour lesquelles les nouvelles valeurs sont correctes. Le chiffre est global (toutes valeurs confondues) et n'est pas pondéré par le nombre d'objets ni la surface.

Contrôle	Taille échantillon	Taux d'accord sur l'échantillon
Justesse de classement selon la couverture	822	86 %
Justesse de classement selon l'usage	1137	85 %

Commentaire :

Ces valeurs sont comparables aux résultats trouvés dans le millésime initial. On s'attend donc à un millésime mis à jour de qualité comparable à celle du millésime initial.

IV- Bilan

L'évaluation effectuée fait ressortir **les forces et les limites des données** :

- **Pour le millésime initial 2019 : une bonne qualité de la couverture et de l'usage, majoritairement conforme aux attentes.**
 - **Pour la couverture**, 85 % des objets tirés dans l'échantillon ont été classés correctement. Le résultat est notamment excellent pour les *Zones bâties* (100 % de conformité constaté sur l'échantillon, sans sous-détection ni sur-détection). Il est également très bon pour les *Formations herbacées* qui couvrent plus des trois-quarts de la surface du département. On note cependant quelques classes fragiles, en-dessous des objectifs, mais représentant une surface infime de l'emprise (*Matériaux composites*, *Sols nus*). Les *Végétations mixtes* sont trop confondues avec les peuplements de feuillus ou de conifères. Enfin, *les Zones non bâties* doivent être améliorées. L'ensemble de ces points identifiés fera l'objet d'ajustements dans la chaîne de production et dans les consignes de photo-interprétation. Les améliorations attendues du modèle IA impacteront aussi positivement la couverture.
 - **Pour l'usage**, le taux de conformité toutes classes confondues, non pondéré, est de 81 %. Il est bon sur quasiment toutes les valeurs, excellent sur les classes fortement représentées. Le résultat est moyen pour *Production tertiaire* (usage pour lequel la dépendance aux fichiers fonciers est forte) ; les résultats sont limites pour les usages *Ferré* et *Inconnu* : pour ces deux classes, l'amélioration viendra aussi d'un ajustement des consignes de photo-interprétation.
- **Pour le millésime mis à jour 2016 : une bonne qualité comparable à celle du millésime 2019.** La mesure d'exhaustivité ne fait pas apparaître de risque particulier sur la bonne détection des mises à jour. La qualité de la mise à jour effectuée est aussi satisfaisante, mais perfectible grâce à un ajustement de certaines consignes.

La délimitation des zones construites a aussi été contrôlée. L'évaluation fait ressortir une conformité de **91,5 % sur le millésime initial 2019** (échantillon de 200 objets), et de **96 % pour la partie mise à jour du millésime 2016** (échantillon de 200 objets touchant les objets du différentiel). Ce résultat est tout à fait rassurant.

Cette analyse statistique sera utilement complétée **par la prise en compte des retours qualitatifs des utilisateurs, qui seront essentiels pour l'amélioration des données sur ce département, et les autres départements qui restent à produire.**

V- Annexe : qualification du modèle IA

V -1 Limites de la qualification

On cherche à estimer la qualité du modèle IA relativement aux annotations saisies par des photo-interprètes. Le modèle IA permet de prédire une information au pixel sous la forme de cartes de chaleur. Les cartes de chaleur ne sont pas livrées par défaut. Il s'agit d'images sur 15 canaux (un canal par classe), qui pour chaque pixel indique la probabilité d'appartenance à la classe. Ces images peuvent souffrir d'effets de bords aux frontières entre classes. Le calcul des cartes de chaleur est suivi d'une vectorisation, des post-traitements, de croisements avec d'autres sources et des reprises par des photo-interprètes.



Figure A1 : Carte de chaleur (gauche) et produit vecteur (droite)

L'évaluation qui suit est donc indépendante des spécifications du produit final. Il s'agit d'une évaluation indirecte de la qualité du produit final.

On rappelle que le modèle IA prédit uniquement la couverture, mais pas l'usage.

Le modèle IA a appris sur des annotations du département du Gers uniquement. Les prochains modèles apprendront sur des annotations France entière : cette différence sera à prendre en compte lors des futures comparaisons.

V -2 Méthodologie : validation croisée par la méthode « k-fold »

Un jeu d'annotation spécifique à la prise de vue aérienne du Gers (150 km²) a été saisi, sous forme de 62 zones de tailles variables avec une nomenclature « NAF » à 15 classes d'occupation du sol. L'ensemble des annotations du Gers a été réparti en 5 ensembles (fold). La répartition n'a pas été faite de façon aléatoire mais par un expert en essayant d'optimiser les critères suivants :

- la répartition du nombre de pixels annotés par classe dans chaque *fold* doit être la plus similaire possible entre les classes,
- chacune des 62 zones saisies va entièrement dans un *fold* (ces zones sont ensuite redécoupées en petites vignettes pour l'algorithme de deep learning et on ne veut pas que des vignettes voisines servent à des rôles différents (apprentissage/validation/test)),
- les superficies de chaque *fold* doivent être homogènes,
- les zones d'annotations de chaque *fold* doivent être bien réparties sur le département du Gers,
- les zones d'annotations peuvent être de type *urbain* / *naturel* / *végétation*. Chaque *fold* doit être équilibré entre ces trois types de zone.

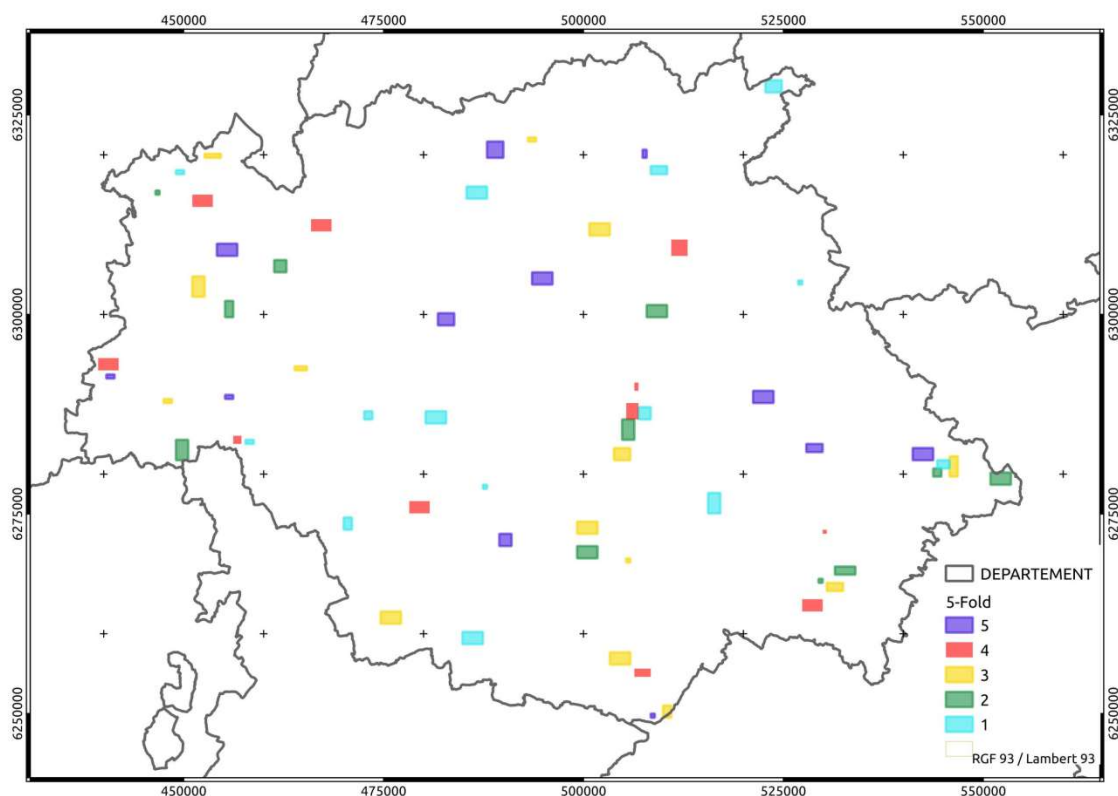


Figure A2 : Choix de la répartition en 5 fold des zones d'annotations sur le Gers (32)

Les *fold*s peuvent jouer plusieurs rôles dans l'apprentissage et évaluation du modèle :

- **train** (apprentissage) : les pixels annotés de cet ensemble servent à déterminer les paramètres du modèle IA, cette détermination se fait de façon itérative. A chaque "époque" on met à jour tous les paramètres du modèle.
- **valid** (validation) : les pixels annotés de cet ensemble servent à choisir la meilleure "époque" (meilleure version du modèle) lors du processus d'apprentissage.
- **test** (test) : les pixels annotés de cet ensemble servent à évaluer la qualité du modèle. Ils ne servent pas au calcul du modèle. Une matrice de confusion et des indicateurs issus de celle-ci sont calculés sur ces données test.

La méthode de validation croisée (k-fold) consiste en réalité à entraîner et évaluer **5 modèles** avec la répartition train/valid/test indiquée dans le tableau suivant. L'évaluation de la qualité du modèle proposée par la suite correspondant à la somme des 5 matrices de confusion des 5 modèles. Les indicateurs de qualité sont calculés depuis cette matrice sommée. Cette méthode permet donc de produire des **indicateurs de qualité sur tout l'ensemble du jeu de données annotées**. Cette méthode est bien éprouvée par la littérature.

fold	1	2	3	4	5
Modèle 1	test	<i>valid</i>	train	train	train
Modèle 2	train	test	<i>valid</i>	train	train
Modèle 3	train	train	test	<i>valid</i>	train
Modèle 4	train	train	train	test	<i>valid</i>
Modèle 5	<i>valid</i>	train	train	train	test

Tableau A1 : Répartition des folds en train/valid/test pour les 5 modèles

V -3 Informations générales sur le modèle

Surface des annotations : 150 km² (exclusivement sur le département du Gers).

Millésime de la PVA : 2019

Nomenclature : nomenclature « naf » en classes qui ne correspond pas exactement à la nomenclature OCS GE, car elle correspond mieux à la réalité de ce qu'on voit sur les images.

Classe du modèle IA	Classe OCS GE	Classes regroupées
batiment	Zones bâties	Surface anthropisée
zone_impermeable	Zones non bâties	
piscine		
zone_permeable	Zones à matériaux minéraux ----- Zones à matériaux composites	
sol_nu	Sols nus	Surface naturelle
surface_eau	Surfaces d'eau	
neige	Névés et glaciers	
Coupe	Peuplement de feuillus ou Peuplement de conifères	Végétation
Feuillus	Peuplement de feuillus ou Peuplement mixte	
coniferes	Peuplement de conifères ou Peuplement mixte	
broussaille	Formations arbustives et sous- arbrisseaux	
Vigne	Autres formations ligneuses	
culture	Formations herbacées	
terre_labouree		

Tableau A2 : Correspondance entre les classes du modèle IA et celles de l'OCS GE

Dans toute la suite, la classe « neige » n'apparaît pas : elle n'est présente ni dans les données de référence, ni dans les données prédites. Il était aussi prévu une classe « autre » qui n'a pas été utilisée et qui n'apparaîtra pas.

Répartition des annotations entre classes :

classe	% pixels	% pixels	regroupement
batiment	4.1%	12.2%	surface anthropisée
zone_permeable	2.1%		
zone_impermeable	5.9%		
piscine	0.0%		
sol_nu	0.1%	2.3%	surface naturelle
surface_eau	2.2%		
coupe	1.0%	85.6%	végétation
feuillus	23.2%		
conifere	2.0%		
brousaille	3.2%		
vigne	3.5%		
culture	44.6%		
terre_labouree	8.0%		

Tableau A3 : Répartition des pixels annotés

Architecture : DeepLab v3 Plus (~ 7 millions de paramètres)

Imagerie utilisée : 5 canaux

- Canaux R.V.B. des orthophotos – résolution 20 cm
- Canal I (infrarouge) des images IRC – résolution 20 cm
- Canal E (élévation) : différence MNS-MNT

V -4 Métriques sur la classification détaillée

Tous les chiffres suivants sont calculés sur la somme des cinq prédictions effectuées au cours de la validation croisée dont la méthodologie a été expliquée ci-dessus. Chaque pixel est évalué une fois.

a) Métriques par classe

Les indicateurs retenus sont ceux classiquement utilisés dans la littérature. Il existe de nombreuses ressources facilement accessibles expliquant ces indicateurs et en donnant les formules de calcul.

- **Précision** : parmi tous les pixels prédits d'une classe, pourcentage de pixels correctement classifiés. **Plus le taux est bas, plus le modèle « sur-détecte » la classe.** Montre la capacité du modèle à effectuer une prédiction correcte d'une classe. Plus on se rapproche de 100% moins on a de sur-détection.
- **Rappel** : parmi tous les pixels de référence d'une classe, pourcentage de pixels correctement classifiés. **Plus le taux est bas, plus le modèle « sous-détecte » la classe.** Montre la capacité du modèle à bien reconnaître une classe. Plus on se rapproche de 100% moins on a de sous-détection.

- **F-score** : moyenne harmonique de la précision et du rappel, indicateur populaire dans la communauté télédétection. Plus on se rapproche de 100%, moins on a, en moyenne, de sous-détection et de sur-détection.
- **IoU** (intersection over union) : parmi la somme de tous les pixels de référence d'une classe et de tous les pixels prédits d'une classe, pourcentage de pixel correctement classifiés. Indicateur intégrant les aspects sur-détection et sous-détection, populaire dans la communauté deep learning.

classe	Precision	rappel	f-score	iou
batiment	86.7%	92.3%	89.4%	80.9%
zone_permeable	51.8%	23.5%	32.3%	19.3%
zone_impermeable	78.0%	82.5%	80.2%	66.9%
piscine	76.8%	49.5%	60.2%	43.1%
sol_nu	30.8%	18.5%	23.1%	13.1%
surface_eau	89.3%	95.0%	92.0%	85.3%
coupe	41.1%	30.5%	35.0%	21.2%
feuillus	85.6%	86.8%	86.2%	75.8%
conifere	54.1%	63.1%	58.3%	41.1%
broussaille	44.2%	31.1%	36.5%	22.3%
vigne	88.8%	92.7%	90.7%	82.9%
culture	88.8%	88.4%	88.6%	79.5%
terre_labouree	72.3%	84.7%	78.0%	64.0%

Tableau A4 : Métriques par classe

Le tableau est à lire ainsi, par exemple pour la première ligne, celle des bâtiments :

- Précision = 86,7 % : « parmi tous les pixels classés **comme bâtiment par le modèle IA**, 86,7 % étaient bien labélisés **comme bâtiment dans les données référence (vérité terrain)**. C'est-à-dire 13,3 % des pixels détectés/classés **comme bâtiment** par le modèle ne sont pas des pixels de bâtiments dans les données référence (sur-détection) ».
- Rappel = 92,3 % : « parmi tous les pixels **labélisés comme bâtiment dans la référence**, 92,3 % été classés **comme bâtiment par le modèle IA**. C'est-à-dire que 7,7 % des pixels **bâtiment** dans les données de référence n'ont pas été détectés/classés **comme des pixels bâtiment par le modèle**. (sous-détection) ».

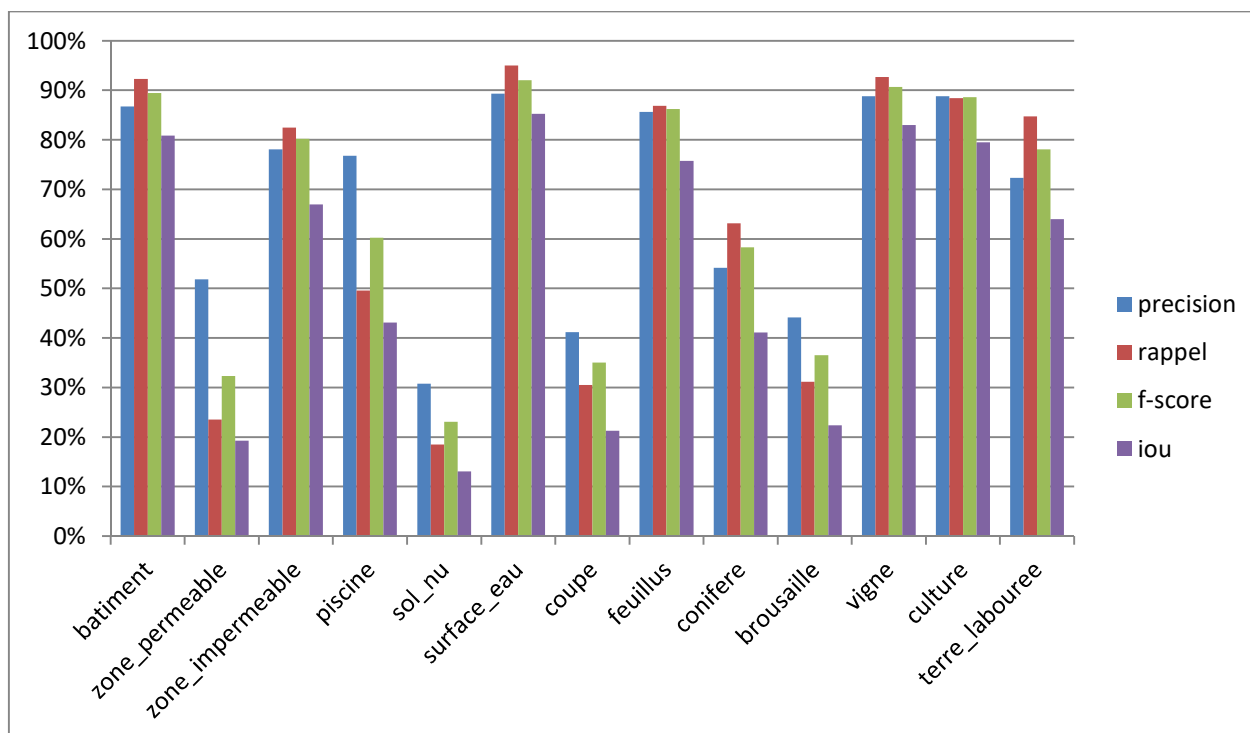


Figure A3 : Histogramme des métriques par classe

Les classes les moins bien « apprises » sont : *zone perméable*, *sol nu*, *coupe*, *broussaille*.

b) Métrique globale

- **Overall Accuracy (OA)** : cet indicateur global calcule simplement le rapport entre nombre de pixels bien classés par rapport à la référence (les saisies des photo-interprètes), et le nombre total de pixels. Il est dépendant du nombre de pixels par classe. Il sera influencé par les classes fortement représentées.

OA	83.4%
----	-------

c) Matrice de confusion normalisée par classe

Une matrice de confusion a été calculée au pixel, mais elle est difficilement lisible. Elle a été normalisée dans le sens des lignes, c'est-à-dire par rapport aux pixels de référence. Les lignes indiquent dans quelle classe de référence (établie par un photo-interprète) ont été classifiés les pixels par le moteur IA. Les cellules diagonales de la matrice correspondent à la métrique « Rappel » du paragraphe précédent, c'est-à-dire la capacité du modèle à détecter correctement une classe.

		PREDICTION													
		batiment	zone_permeable	zone_impermeable	piscine	sol_nu	surface_eau	coupe	feuillus	conifere	brousaille	vigne	culture	terre_labouree	TOTAL
REFERENCE	batiment	92.3%	0.4%	5.0%	0.0%	0.0%	0.1%	0.0%	0.5%	0.0%	0.0%	0.0%	1.7%	0.0%	100%
	zone_permeable	4.0%	23.5%	26.3%	0.0%	0.3%	0.5%	1.3%	4.4%	0.5%	0.8%	1.4%	35.5%	1.6%	100%
	zone_impermeable	5.8%	3.2%	82.5%	0.1%	0.1%	0.6%	0.0%	1.9%	0.1%	0.0%	0.3%	5.1%	0.2%	100%
	piscine	3.9%	3.1%	35.8%	49.5%	0.0%	2.6%	0.0%	0.9%	0.0%	0.0%	0.0%	4.2%	0.0%	100%
	sol_nu	0.3%	4.8%	4.1%	0.0%	18.5%	12.4%	0.9%	6.4%	0.5%	3.7%	0.1%	46.7%	1.6%	100%
	surface_eau	0.1%	0.0%	0.7%	0.0%	0.1%	95.0%	0.0%	1.9%	0.0%	0.2%	0.0%	1.9%	0.1%	100%
	coupe	0.1%	0.3%	0.1%	0.0%	0.1%	0.1%	30.5%	9.2%	5.9%	30.6%	0.5%	22.7%	0.0%	100%
	feuillus	0.2%	0.0%	0.5%	0.0%	0.0%	0.3%	0.6%	86.8%	2.7%	2.6%	0.1%	6.2%	0.1%	100%
	conifere	0.2%	0.1%	0.5%	0.0%	0.0%	0.1%	0.7%	30.4%	63.1%	1.5%	0.0%	3.4%	0.0%	100%
	brousaille	0.0%	0.1%	0.2%	0.0%	0.1%	0.6%	4.8%	35.2%	4.4%	31.1%	0.4%	22.9%	0.1%	100%
	vigne	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.1%	0.2%	0.0%	0.1%	92.7%	6.1%	0.7%	100%
	culture	0.2%	0.5%	0.9%	0.0%	0.1%	0.2%	0.3%	2.6%	0.1%	0.6%	0.6%	88.4%	5.6%	100%
	terre_labouree	0.0%	0.3%	0.1%	0.0%	0.0%	0.1%	0.0%	0.3%	0.0%	0.0%	0.9%	13.6%	84.7%	100%

Tableau A5 : Matrice de confusion normalisée par classe. En rouge, les confusions supérieures à 10 %

La matrice est à lire ainsi, par exemple pour la première ligne : « parmi tous les pixels étiquetés comme *bâtiment* dans les données de référence, 92,3 % ont été classés comme *bâtiment* par le modèle, 0,4% comme *zone perméable* par le modèle, etc. »

Une analyse de la matrice dans le sens des lignes permet de montrer les erreurs type du modèle :

- "*bâtiments*" classés en "*zones imperméables*" (5% des bâtiments en référence),
- "*zones perméables*" classés en "*culture*" (36%) et en "*zones imperméables*" (26%),
- "*zones imperméables*" classées en "*culture*" (5%) et en "*bâtiments*" (6%),
- "*piscines*" classées en "*zones imperméables*" (36%),
- "*sol nus*" classés en "*culture*" (47%), en surface d'eau (12%), en "*feuillus*" (6%), en "*zones perméables*" (5%) et en "*zones imperméables*" (4%),
- "*coupes*" classés en "*broussailles*" (31 %), en "*cultures*" (23%) et en "*feuillus*" (9%)
- "*conifères*" classés en "*feuillus*" (30%),
- "*feuillus*" classés en "*culture*" (6%)
- "*broussailles*" classées en "*feuillus*" (35%) et en "*cultures*" (23%)
- "*vignes*" classées en "*culture*" (6%)
- "*culture*" classées en "*terre labourée*" (6%)
- "*terre labourée*" classé en "*culture*" (14%)

V -5 Métriques sur les classes regroupées

Les trois classes agrégées sont :

- **Surfaces anthropisées** : regroupe les classes *batiment*, *zone perméable*, *zone imperméable*, *piscine*.
- **Surfaces naturelles** : regroupe les classes *sol nu*, *surface d'eau*, et *neige* (absente dans ce jeu).
- **Végétation** : tout le reste, *conifères*, *coupes*, *feuillus*, *broussailles*, *vignes*, *cultures*, *terre labourée*

Les métriques qui suivent sont déduites de la classification fine, en agrégeant les pixels de chaque regroupement de classe.

a) Métriques par classe

classe	precision	rappel	f-score	iou
surface_anthropisee	91.4%	87.1%	89.2%	80.5%
surface_naturelle	88.1%	91.4%	89.7%	81.4%
vegetation	98.1%	98.6%	98.3%	96.7%

Tableau A6 : Métriques par classe agrégée

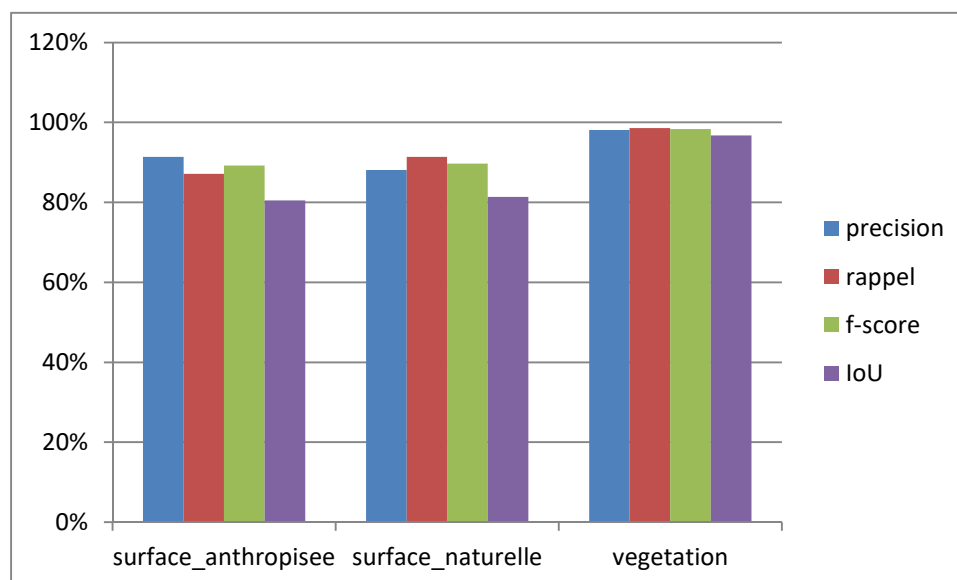


Figure A4 : Histogramme des métriques par classe agrégée

b) Métriques globales overall accuracy

OA	97.0%
----	-------

c) Matrices de confusion

La première matrice est non normalisée, elle montre les confusions effectives entre classes en nombre de pixels. Les chiffres sont en millions de pixels, arrondi avec une seule décimale significative.

La seconde matrice est normalisée par lignes, comme pour les classes détaillées.

Les métriques par classe et globales sont aussi rappelées dans le tableau suivant, pour disposer de tous les indicateurs en un seul tableau lisible.

en millions de pixels, arrondi

		PREDICTION				rappel	f-score	IoU
		surface_anthropisee	surface_naturelle	vegetation	TOTAL			
REFERENCE	surface_anthropisee	253.2	1.5	36.0	290.6	87.1%	89.2%	80.5%
	surface_naturelle	0.7	50.1	4.0	54.8	91.4%	89.7%	81.4%
	vegetation	23.3	5.3	2 017.7	2 046.2	98.6%	98.3%	96.7%
	TOTAL	277.1	56.9	2 057.7	2 391.7			
precision		91.4%	88.1%	98.1%				

OA	97.0%
-----------	-------

normalisée par ligne

		PREDICTION			
		surface_anthropisee	surface_naturelle	vegetation	TOTAL
REFERENCE	surface_anthropisee	87.1%	0.5%	12.4%	100.0%
	surface_naturelle	1.3%	91.4%	7.3%	100.0%
	vegetation	1.1%	0.3%	98.6%	100.0%

Tableau A7 : Matrices de confusion et métriques par classe agrégée

Commentaire :

- Les surfaces anthropisées sont un peu plus sous-détectées que sur-détectées. Environ 12% des pixels de référence de surfaces anthropisées sont classifiés à tort en végétation.
- Au contraire, les surfaces naturelles sont un peu plus sur-détectées que sous-détectées. Environ 7% des pixels de référence de surfaces naturelles sont classifiés à tort en végétation.
- Enfin, la classification de la végétation est excellente.
- La végétation étant la classe la plus représentée (~85 % des pixels), elle tire vers le haut la métrique globale « overall accuracy » qui est excellente sur l'ensemble du jeu (97 % des pixels sont correctement classifiés). Ceci masque un chiffre moins bon pour les surfaces anthropisées, pour lesquelles 80% des pixels sont correctement classifiés.

Ces chiffres sont cohérents avec la fonction de coût et donc l'objectif du modèle IA : maximiser l'indicateur OA « overall accuracy » sur l'ensemble du jeu.

V -6 Conclusion : forces et limites d'utilisation du modèle IA

Le modèle obtient des résultats limités pour les classes sur lesquelles les données de référence étaient trop faiblement représentées (par exemple les sols nus ou les coupes). Le modèle est également moins performant pour les zones perméables, dont la délimitation est floue ou la variabilité élevée. Enfin, on trouve des confusions entre feuillus, conifères ou broussailles, liées probablement à des confusions dans les annotations elles-mêmes, qui gagneraient à être améliorées par des experts forestiers.

Malgré ces rares écueils, le modèle **obtient de très bons résultats** : plus de 83 % des pixels sont correctement prédits avec la classification la plus fine, et plus de 97 % avec la classification agrégée, ce dernier chiffre étant certes favorisé par l'excellent score de la végétation, très majoritaire sur ce département. Plusieurs classes parviennent à obtenir des indicateurs qui frôlent ou dépassent les 80 %, comme les bâtiments, les surfaces d'eau ou les cultures, bien que cette classe ait souvent une délimitation floue et une forte variabilité. Ces chiffres valident la fiabilisation de la chaîne IA, permettent de confirmer les orientations IA prises pour la production des cartes chaleur, et sont également rassurants pour la suite des traitements de la chaîne de production OCS GE (vectorisation, croisement avec d'autres sources, photo-interprétation guidée par des alertes) en permettant de limiter la photo-interprétation pour la couverture.