

# LA QUALITE DANS LA PRODUCTION DE L'OCS GE

Intervention au GT CNIG Qualité des Données Géographiques  
13 octobre 2022



# Question

*Comment aborde-t-on la question de la qualité des données pour la production de l'OCS GE ? La qualité dans l'OCS GE (et non la qualité de l'OCS GE)*

## Plan

- 1.Contexte : l'OCS GE et sa nouvelle chaîne de production
- 2.Evaluation du modèle IA
- 3.Contrôles qualité en fin de production
- 4.Qualification à froid

# 1. Contexte et présentation de la nouvelle chaîne de production

# Présentation de l'OCS GE

IGN  
OCS GE SOCLE  
COUVERTURE DU SOL (CS)

IGN  
OCS GE SOCLE  
USAGE DU SOL (US)

|                      |                               |                                                |                                         |
|----------------------|-------------------------------|------------------------------------------------|-----------------------------------------|
| CS 1 Sans végétation | CS 11 Surfaces anthropisées   | CS 11.1 Zones imperméables                     | CS 11.1.1 Zones bâties                  |
|                      |                               | CS 11.2 Zones perméables                       | CS 11.2.1 Zones non bâties              |
|                      |                               | CS 11.2 Zones perméables                       | CS 11.2.2 Matériaux minéraux            |
|                      |                               |                                                | CS 11.2.3 Matériaux composites          |
|                      | CS 12 Surfaces naturelles     | CS 12.1 Sols nus                               | CS 12.1.1 Sols nus                      |
|                      |                               |                                                | CS 12.2 Surfaces d'eau                  |
| CS 2 Avec végétation | CS 21 Végétation ligneuse     | CS 21.1 Formations arborées                    | CS 21.1.1 Feuillus                      |
|                      |                               |                                                | CS 21.1.2 Conifères                     |
|                      |                               |                                                | CS 21.1.3 Nuda                          |
|                      |                               | CS 21.2 Formations arbrutives, sous-arbrutives | CS 21.2.1 Formations ligneuses          |
|                      |                               |                                                | CS 21.2.2 Autres formations ligneuses   |
|                      |                               | CS 21.3 Formations herbacées                   | CS 21.3.1 Formations herbacées          |
|                      |                               |                                                | CS 21.3.2 Autres formations herbacées   |
|                      | CS 22 Végétation non ligneuse | CS 22.1 Formations herbacées                   | CS 22.1.1 Formations herbacées          |
|                      |                               |                                                | CS 22.2 Autres formations non ligneuses |
|                      |                               | CS 22.2 Autres formations non ligneuses        | CS 22.2.1 Formations herbacées          |

|                                                             |                                            |                                              |
|-------------------------------------------------------------|--------------------------------------------|----------------------------------------------|
| US 1 Production primaire                                    | US 1.1 Agriculture                         | US 1.1.1 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 2.1 Production secondaire               | US 2.1.1 Production secondaire               |
| US 2.2 Usage mixte                                          | US 2.2.1 (US 2.2.1, US 2.2.2)              | US 2.2.1 (US 2.2.1, US 2.2.2)                |
| US 3 Production tertiaire                                   | US 3.1 Production tertiaire                | US 3.1.1 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 Ferre                               |
|                                                             |                                            | US 4.1.3 Aérien                              |
|                                                             |                                            | US 4.1.4 Navigable                           |
|                                                             |                                            | US 4.1.5 Autres                              |
| US 5 Usage résidentiel                                      | US 5.1 Services logistiques et de stockage | US 5.1.1 Services logistiques et de stockage |
|                                                             |                                            | US 5.1.2 Services logistiques et de stockage |
|                                                             |                                            | US 5.1.3 Services logistiques et de stockage |
| US 6 Autre usage                                            | US 6.1 Usage résidentiel                   | US 6.1.1 Usage résidentiel                   |
|                                                             |                                            | US 6.1.2 Zones en transition                 |
|                                                             |                                            | US 6.1.3 Zones abandonnées                   |
|                                                             |                                            | US 6.1.4 Sans usage                          |
|                                                             |                                            | US 6.1.5 Inconnu                             |

UMI | Zones bâties : 200 m<sup>2</sup> | Intérieur Zone construite : 500 m<sup>2</sup> | Hors Zone construite : 2 500 m<sup>2</sup>

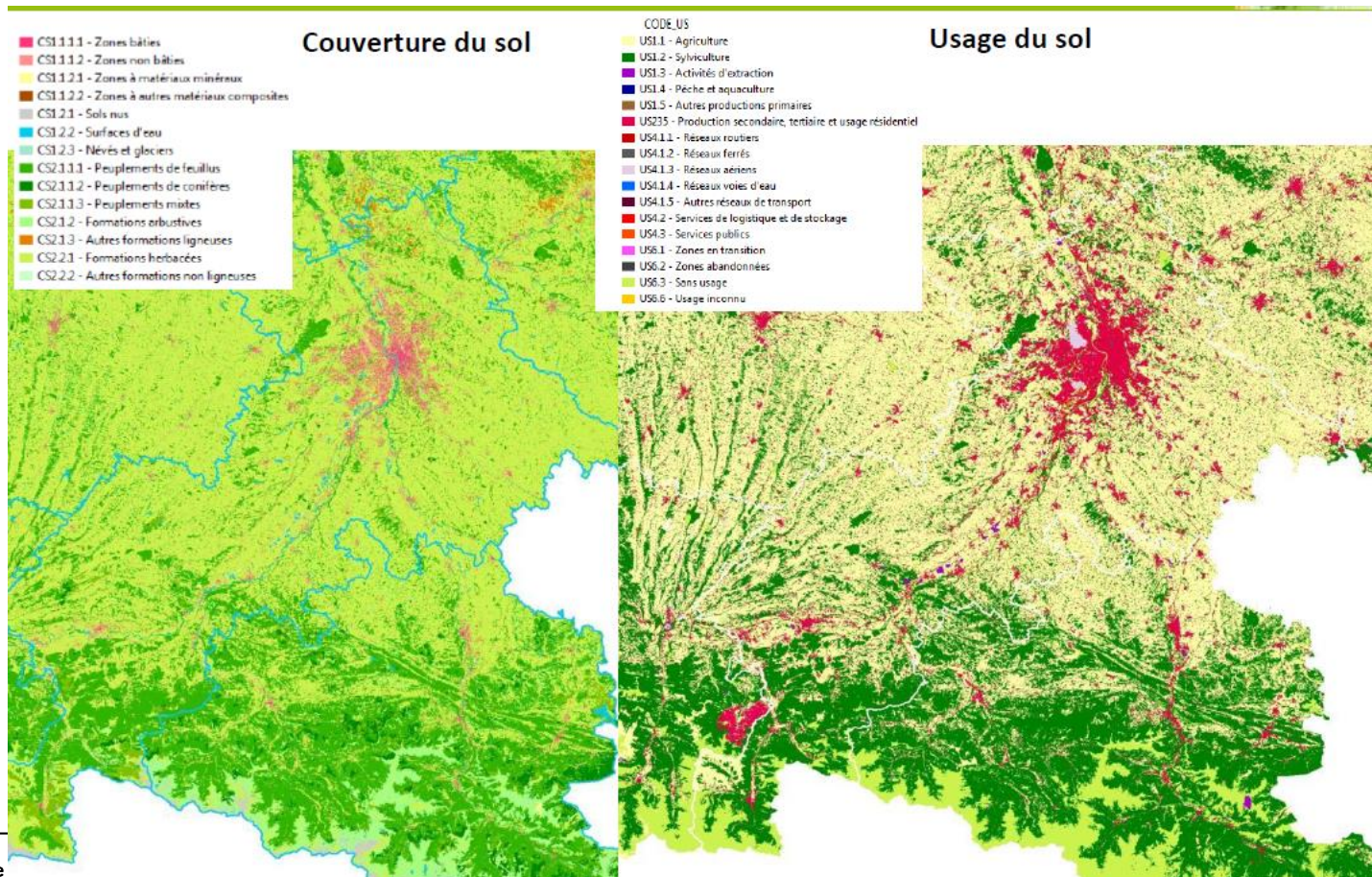
Produit conforme aux préconisations nationales du CNIG

- Deux dimensions principales, avec une nomenclature emboîtée :  
**Couverture** : que voit-on ? → vue « physionomique »  
**Usage** : à quoi ça sert ? → vue « anthropique »

- **Seuils surfaciques** et seuils de largeur selon les zones
- **Partition** du territoire (chaque point est renseigné)
- **Production millésimée** sur une référence image (cohérence temporelle)
- **La mise à jour** ne doit montrer que les vrais changements sur le terrain (pas de bruit de mesure ou d'interprétation)
- S'appuie sur une **ossature** (à partir des principaux réseaux) pour avoir une partition de référence du territoire, cohérente et continue

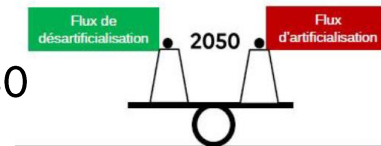
- Compatibilité **INSPIRE**
- Compatibilité **CNIG**

# Présentation de l'OCS GE

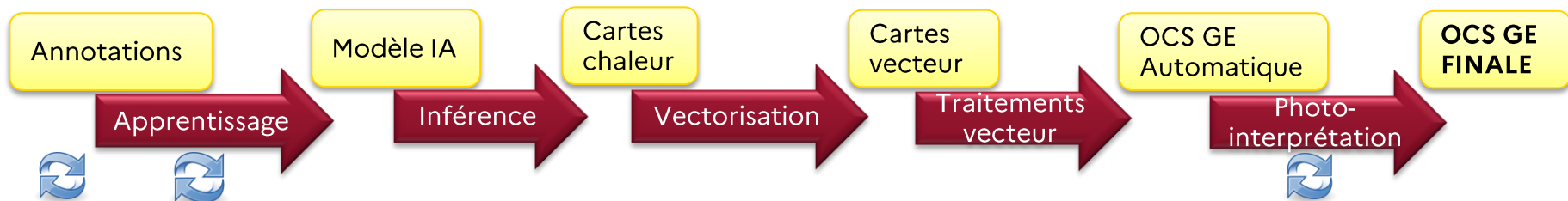


# Production dans le cadre du dispositif national de mesure de l'artificialisation des sols

- L'IGN produit de l'OCS GE depuis 2013
- **Contexte actuel** : objectif ZAN (« zéro artificialisation nette ») 2050
  - Inscrit dans la loi Climat & Résilience
  - Décret 2022 définit l'artificialisation
  - Besoin d'un outil homogène, cohérent avec cette définition, utilisable au niveau national comme au niveau local
- La DGALN a commandité la couverture du territoire national en OCS GE pour 2024, avec 2 millésimes pour chaque département. Données **open data**.
- **Objectif ambitieux**, grâce à l'introduction d'intelligence artificielle dans la chaîne (deep learning). Saut technologique dans les méthodes de production.
- **Déjà réalisé avec la nouvelle chaîne** : Gers (diffusion printemps 2022).  
**En cours** : Ile-et-Vilaine, Rhône, Var.



# La nouvelle chaîne de production



QUALITÉ

QUALITÉ

Evaluation **indirecte**  
(évaluation de l'outil)

Evaluation **directe** de la  
conformité du produit aux  
spécifications (post-prod)

QUALITÉ

Evaluation **à froid** (autre  
méthode, autre regard)

QUALITÉ

# La nouvelle chaîne de production



Annotations



Carte de chaleur



OCS GE

## 2. Evaluation du modèle IA

# L'IA en résumé ...

Annotations

Apprentissage supervisé  
(Gers puis France)

Modèle

Inférence  
(département)

Cartes de prédiction  
(= cartes de chaleur)

Orthophotographie  
(RVB, Infrarouge)  
Altimétrie



Stratégie d'apprentissage  
Calcul intensif sur GPU

Probabilité  
d'appartenance d'un  
pixel à une classe

## Architecture actuelle : U-Net

- Réseau de neurone à convolution (5 couches de convolution et 5 couches de déconvolution)
- Environ 30 millions de paramètres
- Une dizaine d'hyperparamètres

## Diffusion des ressources IA (Gers actuellement) :

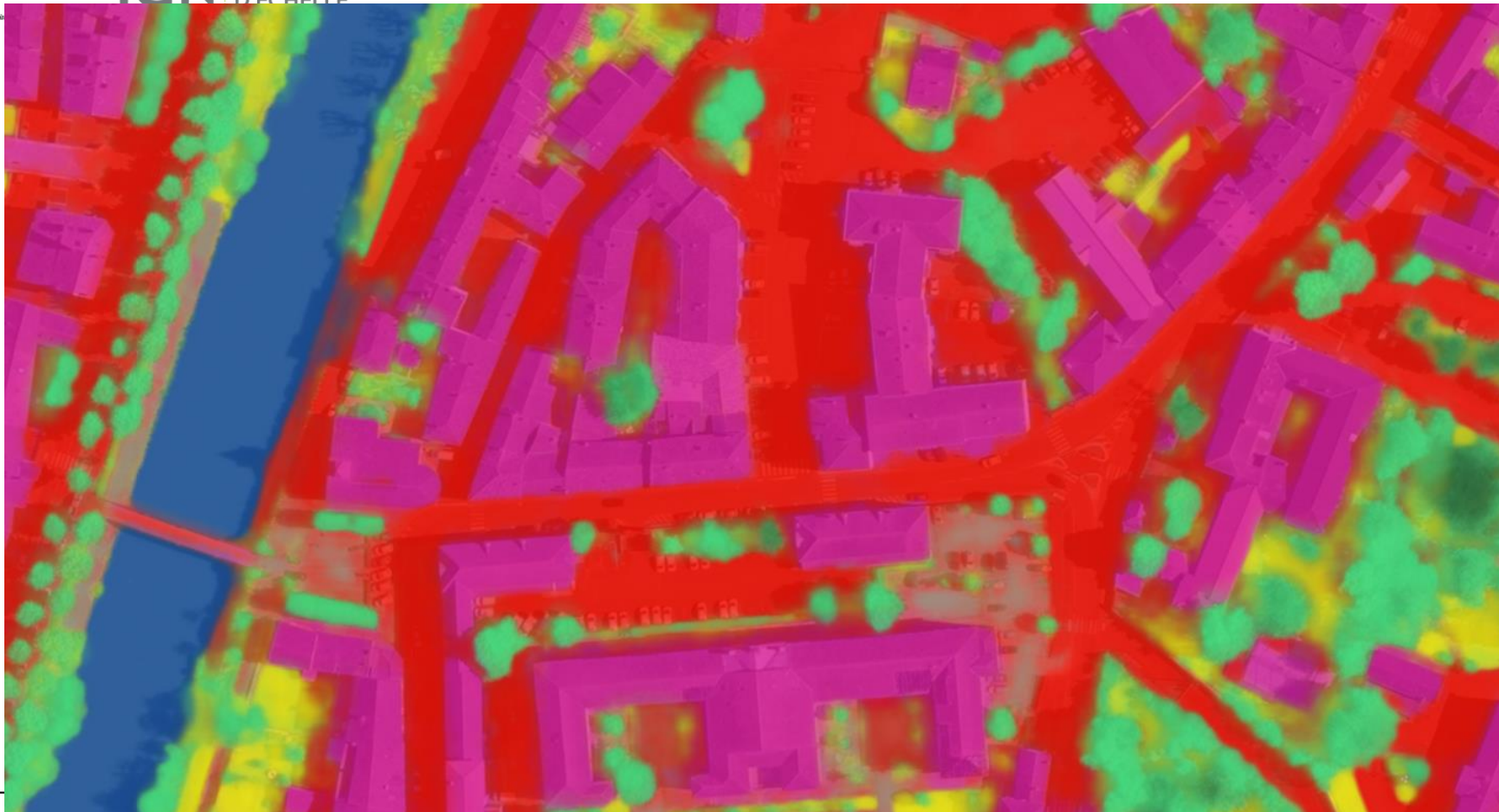
<https://geoservices.ign.fr/ressources-ia-de-couverture-du-sol>

## Regard des utilisateurs (annotations et cartes de chaleur).

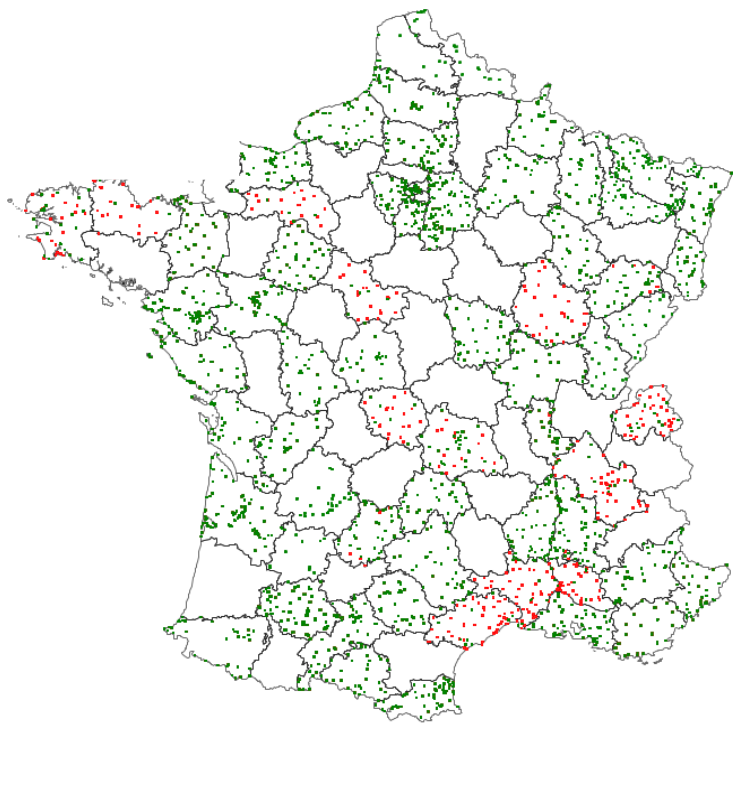
Ex. Agence Urbanisme Toulon

Ces travaux ont bénéficié d'un accès aux ressources en IA de l'IDRIS,  
au travers de l'allocation de ressources AD010113401 attribuée par GENCI.





# Production des annotations



*Vert : annotations utilisées pour modèle  
V1.0 (2050 km<sup>2</sup>)*

*Rouge : annotations complémentaires en cours  
de validation (cible ~2500 km<sup>2</sup>)*

***Plus de 1 million de vignettes 256x256 px***

**Choix des zones :** représentativité des classes, des paysages, des dates de prise de vue

# Contrôle qualité des annotations

## • Saisies par des prestataires, par lots.

- L'IGN fournit la segmentation géométrique, calculée par un logiciel dédié
- Les prestataires saisissent une valeur

## • Exigences de précision thématique

- Cf. tableau
- Nomenclature en annexe

## • Tirage aléatoire stratifié par classe

- ~100 objets par classe, selon les classes (compromis entre recherche de significativité statistique et charge à consacrer)
- Deux niveaux pour lisser la **variabilité** des photo-interprètes : si note négative d'un objet, alors second regard unifié (= « consolidation »)

## • Retour qualitatif effectué

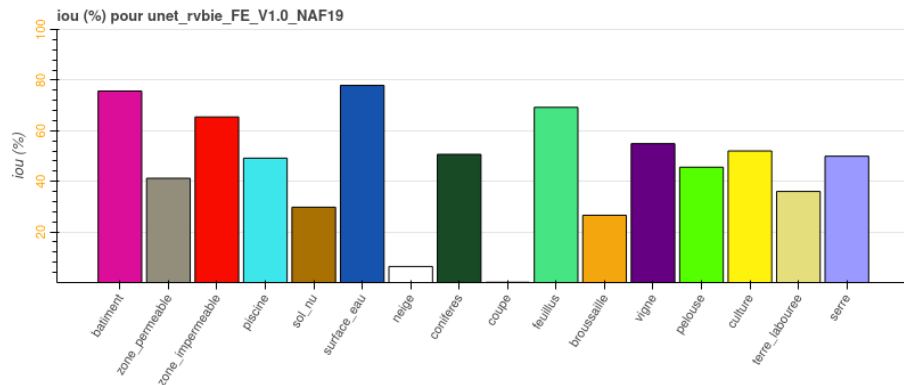
| Classe parent           | Classe de 2 <sup>ème</sup> niveau | Taux d'accord niveau parent (%) | Taux d'accord niveau 2 (%) |
|-------------------------|-----------------------------------|---------------------------------|----------------------------|
| Surface anthropisée     | bâtiment                          | 98                              | 95                         |
|                         | serre                             | 98                              | 95                         |
|                         | zone imperméable                  | 98                              | 95                         |
|                         | zone perméable                    | 90                              | 85                         |
|                         | piscine                           | 98                              | 95                         |
| Surface naturelle       | sol nu                            | 90                              | 85                         |
|                         | surface eau                       | 98                              | 95                         |
|                         | neige                             | 98                              | 95                         |
| Végétation ligneuse     | ligneux                           | 98                              | 95                         |
|                         | coupe                             | 90                              | 85                         |
|                         | feuillus                          | 98                              | 95                         |
|                         | conifères                         | 98                              | 95                         |
|                         | mixte                             | 98                              | 95                         |
|                         | broussaille                       | 98                              | 95                         |
|                         | vigne                             | 98                              | 95                         |
| Végétation non ligneuse | terre labourée                    | 98                              | 95                         |
|                         | culture                           | 98                              | 95                         |
|                         | pelouse                           | 98                              | 95                         |

# Les métriques IA

- **Méthode d'évaluation : validation croisée (K-Fold).** Elimine les biais du choix des zones d'annotation entre apprentissage et test pour calculer les métriques. Cf. annexes
- On a retenu les indicateurs classiques de la littérature IA et télédétection.
- **Indicateurs par classe :**
  - 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. Plus on se rapproche de 100% moins on a de sur-détection.  
(= user accuracy)
  - 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. Plus on se rapproche de 100% moins on a de sous-détection.  
(= producer accuracy, sensibilité)
  - F-score : moyenne harmonique de la précision et du rappel, indicateur populaire dans la communauté télédétection.
  - 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.
- **Indicateurs globaux**
  - Précision globale (overall accuracy) : taux de pixels bien classés. Influencé par les classes fortement représentées.
  - mIoU = moyenne des IoU des classes. Non pondéré.

# Exemple de résultat

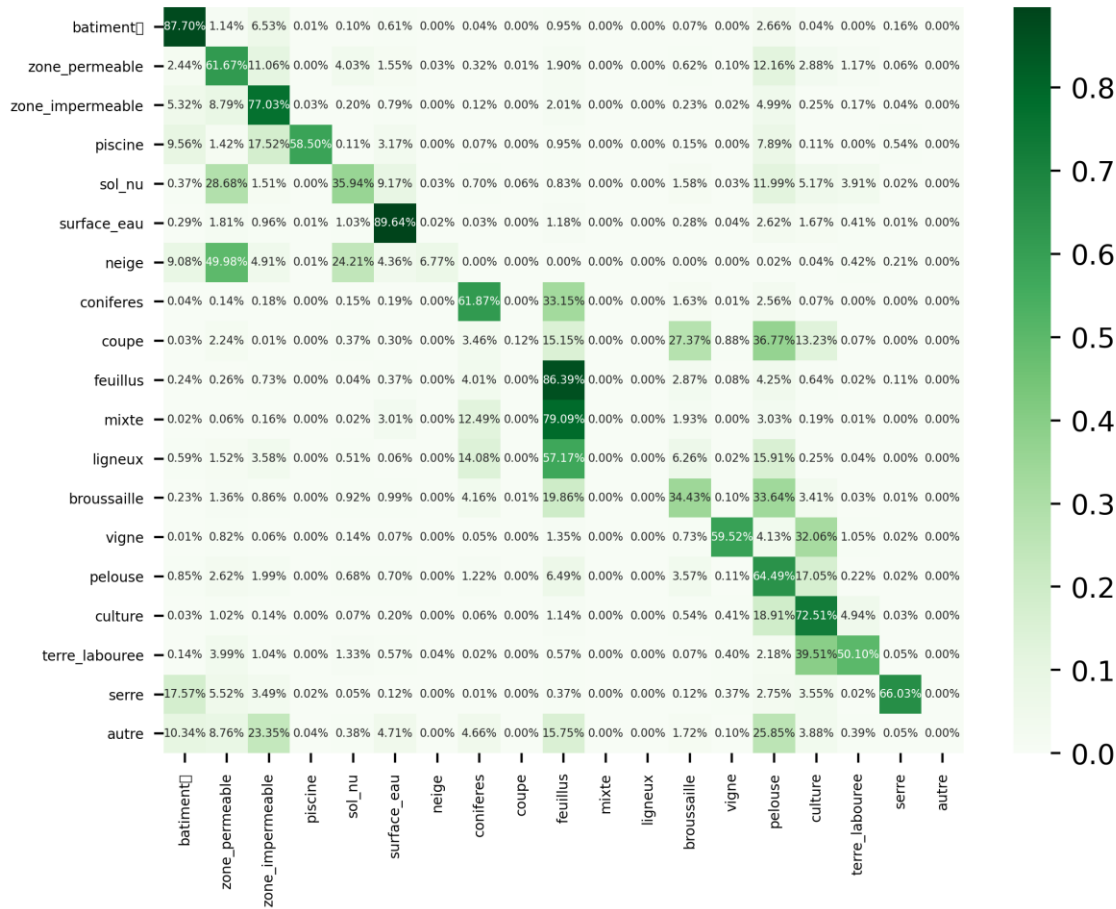
| classe             | Précision | Rappel | F-Score | IoU  |
|--------------------|-----------|--------|---------|------|
| 1-batiment         | 84.4      | 87.7   | 86.0    | 75.4 |
| 2-zone_permeable   | 55.2      | 61.7   | 58.3    | 41.1 |
| 3-zone_impermeable | 80.9      | 77.0   | 78.9    | 65.2 |
| 4-piscine          | 75.2      | 58.5   | 65.8    | 49.0 |
| 5-sol_nu           | 63.4      | 35.9   | 45.9    | 29.8 |
| 6-surface_eau      | 85.4      | 89.6   | 87.5    | 77.8 |
| 7-neige            | 50.8      | 6.8    | 12.0    | 6.4  |
| 8-coniferes        | 73.4      | 61.9   | 67.1    | 50.5 |
| 9-coupe            | 16.5      | 0.1    | 0.3     | 0.1  |
| 10-feuillus        | 77.4      | 86.4   | 81.6    | 69.0 |
| 13-broussaille     | 53.9      | 34.4   | 42.0    | 26.6 |
| 14-vigne           | 87.6      | 59.5   | 70.9    | 54.9 |
| 15-pelouse         | 60.7      | 64.5   | 62.5    | 45.5 |
| 16-culture         | 64.7      | 72.5   | 68.4    | 52.0 |
| 17-terre_labouree  | 56.1      | 50.1   | 52.9    | 36.0 |
| 18-serre           | 67.2      | 66.0   | 66.6    | 49.9 |



| Modèle                       | OA      | mIoU    |
|------------------------------|---------|---------|
| U-Net<br>RVBIE_FE_V1.0_NAF19 | 70.23 % | 45.57 % |

# Matrice de confusion

Référence



0.8  
0.7  
0.6  
0.5  
0.4  
0.3  
0.2  
0.1  
0.0

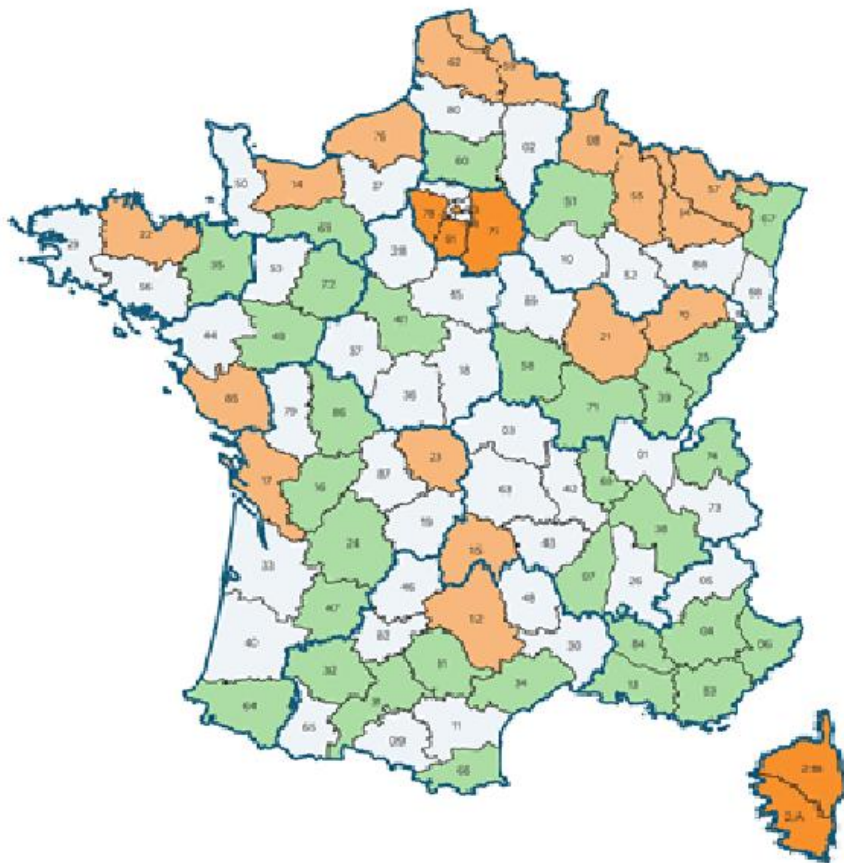
## Exemple de résultat

**Matrice de confusion normalisée**  
dans le sens des lignes (Référence).

La diagonale correspond au Rappel par classe

Exemple de lecture : « parmi tous les pixels étiquetés comme « bâtiment » dans les données de référence, 87.7% ont été classés comme « bâtiment » par le modèle, 1.14% comme « zone\_permeable », etc.

# Métrique globale / département

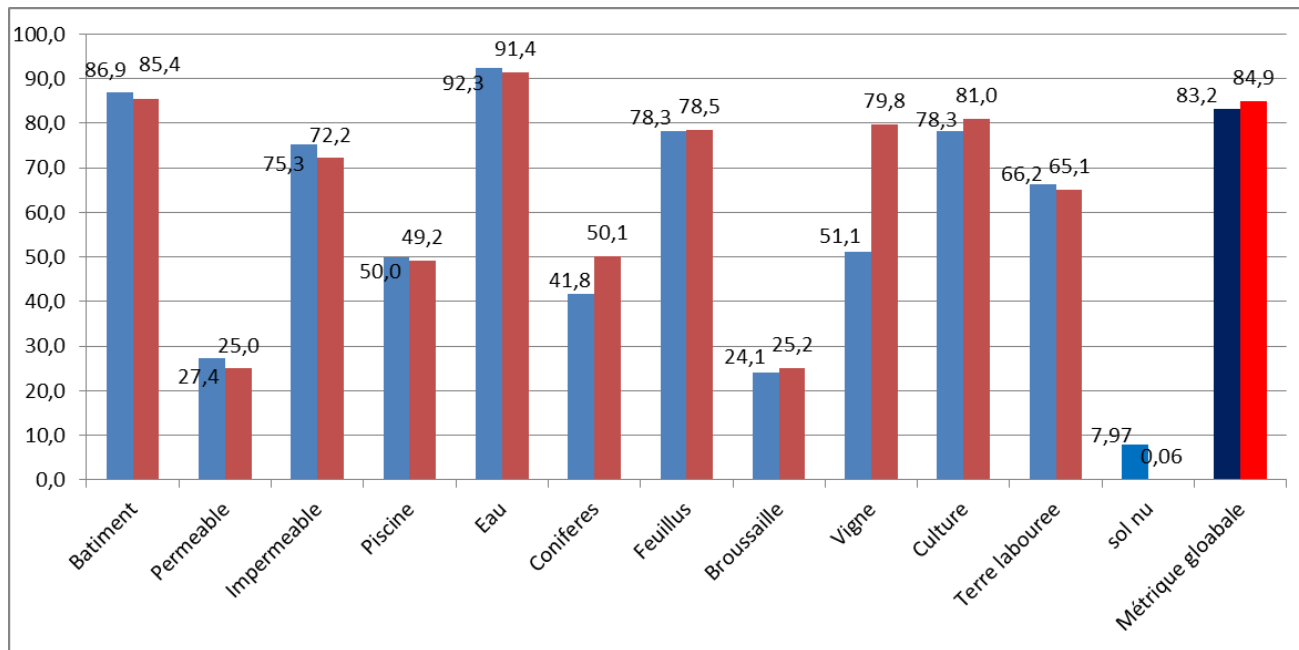


Calculé sur les  
départements avec  
annotations (53) et 2050  
km<sup>2</sup> (2500 prévus)  
Modèle FE V 1.0

Attention biaisé par la  
répartition des classes,  
inhomogène selon les  
départements

- A optimiser (< 60 %)
- Moyen (60-70 %)
- Bon (> 70 %)
- Pas de donnée

# Application du modèle FE sur le GERS (32)



Sol nus, pas assez d'exemple sur le Gers pour définir un métrique valide

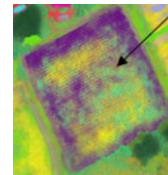
Meilleur pour la plupart des classes

Moins bon pour les vignes, conifères broussailles, et culture

OA - Pixels bien classés modèle « FE » : 83,2 %  
OA - Pixels bien classés modèle « Gers » : 84,9 %

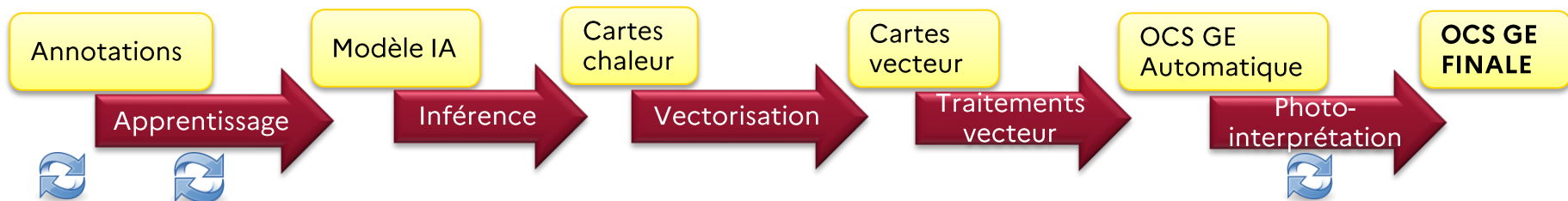
■ modèle national appliqué au Gers  
■ modèle départemental Gers

- L'IA fonctionne bien, et permet de dérouler le reste de la chaîne
- Pistes d'améliorations
  - Recherche de la **qualité des annotations**
  - Prise en compte des **retours utilisateurs**
  - **Procédure industrialisée de fourniture de métriques**, qui permettent de quantifier les modèles et de les comparer.
  - **Les métriques ne disent pas tout**
  - **Analyse qualitative nécessaire**, complémentaire au quantitatif. Ex. non exhaustif :
    - Mauvaise reconnaissance des patterns linéaires (cultures, vignes, verger)
    - Confusions sols nus / zones perméables
  - **Pas de « fétichisme de la métrique »** (au sens d'admiration sans réserve)  
L'IA n'est pas une fin en soi, mais un moyen.  
Ce qui compte, ce n'est pas l'outil, c'est le produit final.



# 3. Contrôles qualité en fin de production

# La chaîne de production



QUALITÉ

QUALITÉ

Evaluation **indirecte**  
(évaluation de l'outil)

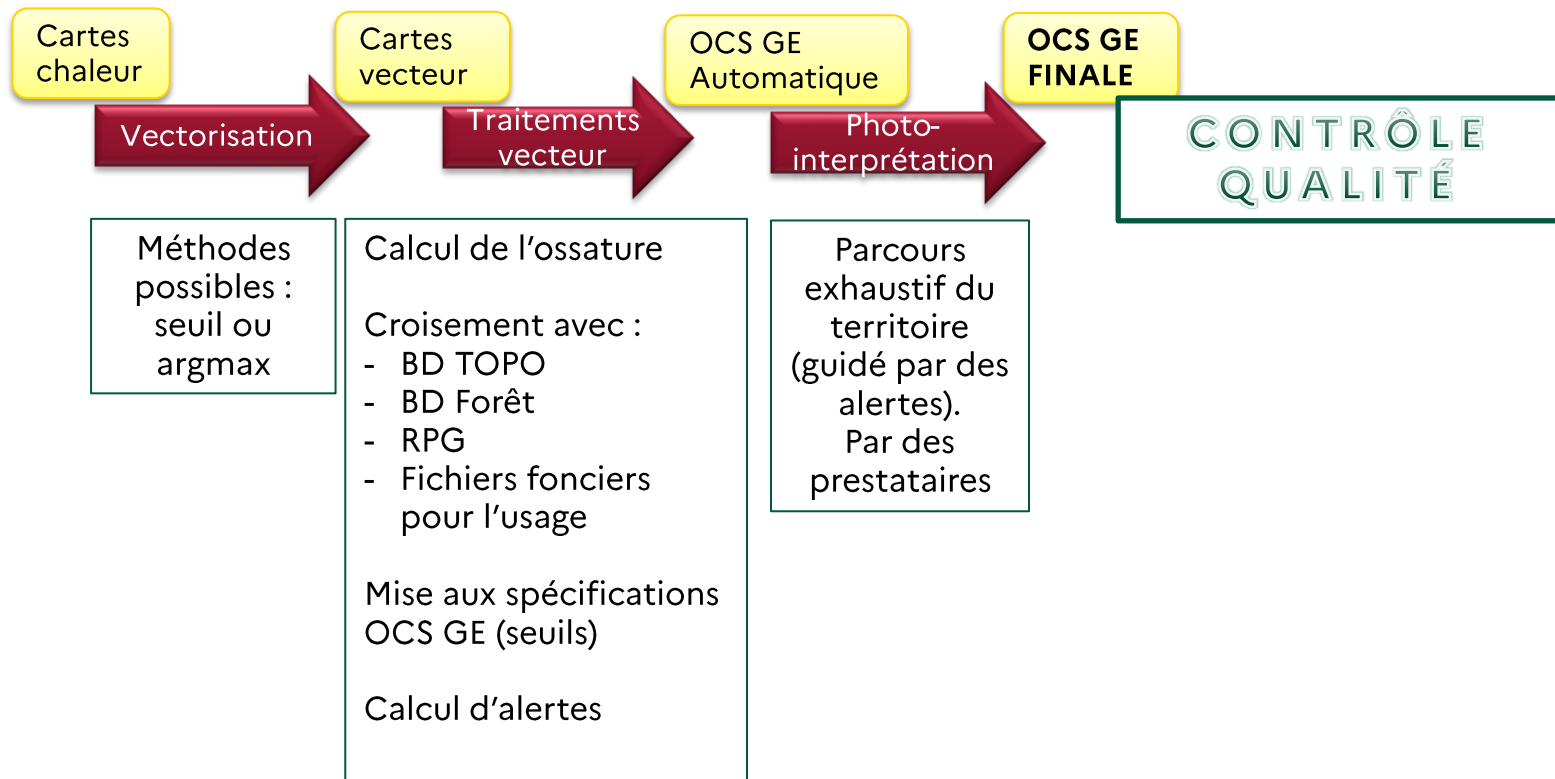
Evaluation **directe** de la  
conformité du produit aux  
spécifications (post-prod)

QUALITÉ

Evaluation **à froid** (autre  
méthode, autre regard)

QUALITÉ

# La chaîne de production, côté vecteur : millésime initial

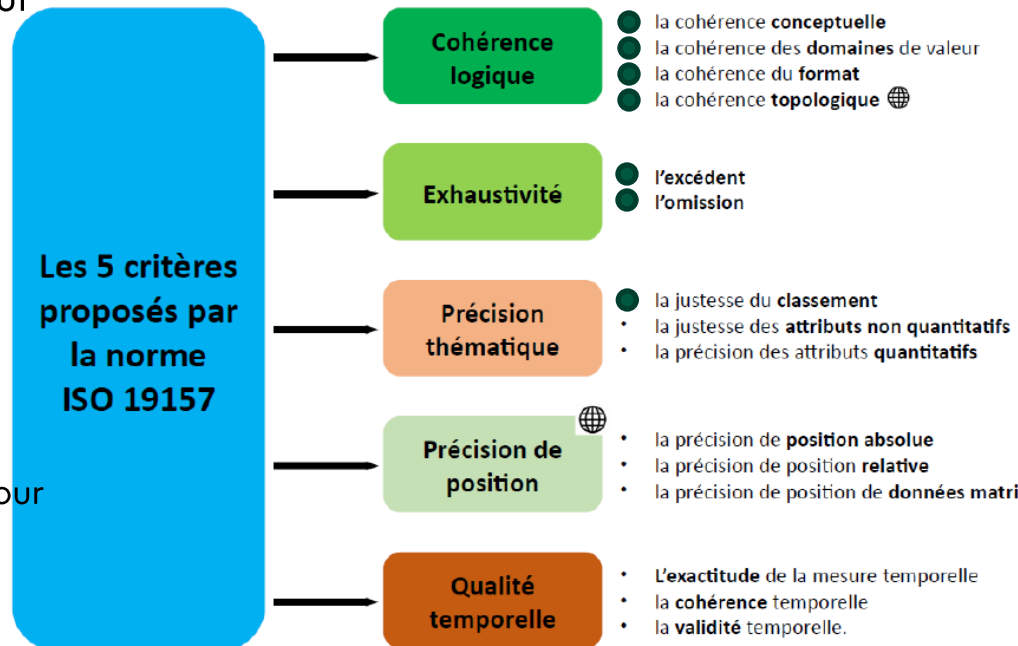


# Contrôle qualité sur l'OCS GE Finale

- Un prestataire traite un département (certains sont réalisés en interne IGN)  
pour le millésime initial et le millésime mis à jour

- Si on se rattache aux critères proposés  
par la norme ISO 19157, on contrôle :

- La cohérence logique pour les deux millésimes
- La précision thématique (justesse de classement)
  - Sur le millésime initial : tout le jeu
  - Sur le millésime mis à jour : les données mises à jour
- L'exhaustivité de la mise à jour



- Rapport sur le Gers : [lien](#)

# Contrôle qualité sur l'OCS GE Finale : cohérence logique

- **Sur les deux millésimes (initial et mise à jour)**
- **Cohérence de format** (shapefile, projection)
- **Cohérence conceptuelle**
  - Conformité au schéma de données
  - Pas de faces adjacentes de mêmes couverture et usage hors ossature
  - Respects des UMI (unités minimum d'intérêt ou d'interprétation)
    - Pas d'objets de surface < 200 m<sup>2</sup> sauf ossature
    - Pas d'objets de surface < 500 m<sup>2</sup> sauf zones bâties
    - Pas d'objets de surface < 2500 m<sup>2</sup> hors zone construite
    - Pas d'objet d'usage Sylviculture < 5000 m<sup>2</sup>
- **Cohérence des domaines de valeur**
  - Nomenclature couverture et usage respectée
  - Croisements couverture et usage autorisés
  - Autres champs (ossature, source, code origine, id origine)
- **Cohérence topologique**
  - Validité des géométries ; pas de géométries multiples
  - Pas de superposition ; pas de trous

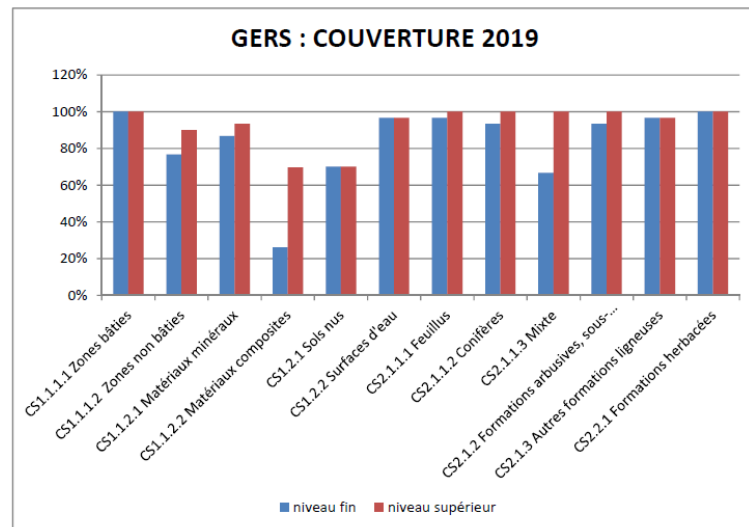
# Contrôle qualité sur l'OCS GE Finale : précision thématique millésime initial

- **Contrôle de la justesse de classement pour la couverture et l'usage**
- **Au préalable**, les objets de grande surface (largeur > 1 km) sont découpés par dichotomie
- **Tirage aléatoire stratifié par valeur de couverture et d'usage**
  - **Au moins** 30 objets par state (compromis à trouver entre significativité statistique et charge à consacrer)
  - Si population < 30 : contrôle exhaustif
- **Deux taux d'accord évalués :**
  - Appartenance à la classe la plus fine : exigence 85 %
  - Appartenance à la classe de premier niveau : exigence 95 %
- **Un polygone est considéré comme correct si au moins 85 % de sa surface est considérée comme bien codée**
- **En cas d'erreur, double regard pour lisser la variabilité (consolidation)**
- **Valeur correcte saisie pour dériver des matrices de confusion**
- **Raisonnement par test statistique, on veut à ce stade un résultat binaire (oui/non)**
  - Hypothèse H0 : au moins 85% (ou 95% selon le test) des objets sont conformes
  - On identifie les classes où l'on peut affirmer avec une confiance de 95% que le jeu n'est pas conforme

# Contrôle qualité sur l'OCS GE Finale : précision thématique millésime initial

## • Exemple de résultat : Gers 2019, couverture

| 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%          |



# Contrôle qualité sur l'OCS GE Finale : Limite de zone construite



Englobe les zones bâties et leur environnement proche.

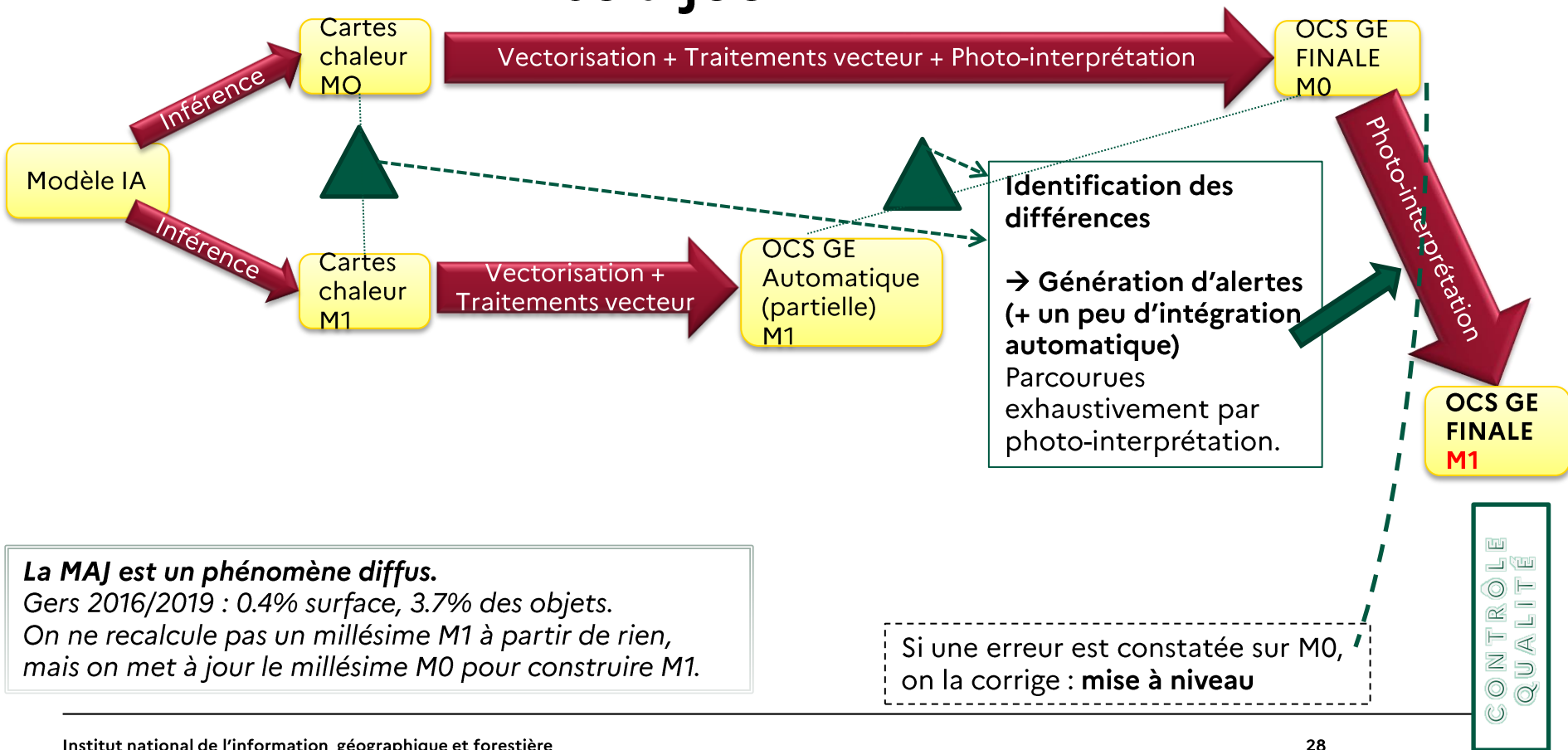
**Intérêt** : permet de focaliser sur des zones évolutives, donc à **enjeu** d'un point de vue aménagement du territoire et artificialisation.

Permet de distinguer les espaces construits, définis sur une base bâtie, des terrains naturels, agricoles, forestiers et espaces aménagés isolés non bâtis.

Cf. spécifications pour la définition exacte

- Tirage aléatoire de 200 limites de zones construites, découpées par tronçons de 60 m maxi
- Vérification binaire que la limite est correcte
- Exigence taux d'accord = 95 %

# La chaîne de production, côté vecteur : mise à jour

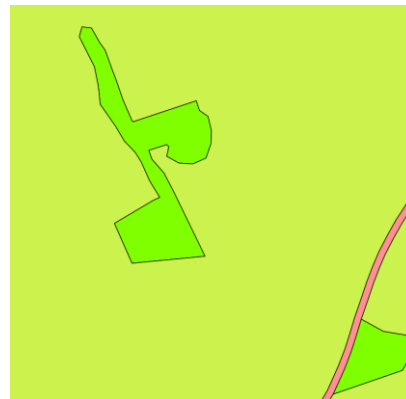
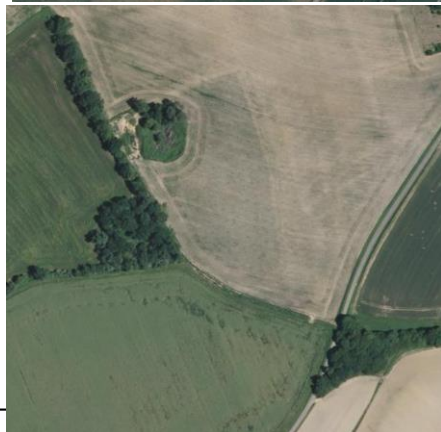


# La chaîne de production, côté vecteur : mise à jour

2016



2019



Stabilité de ce  
qui n'a pas  
bougé sur le  
terrain

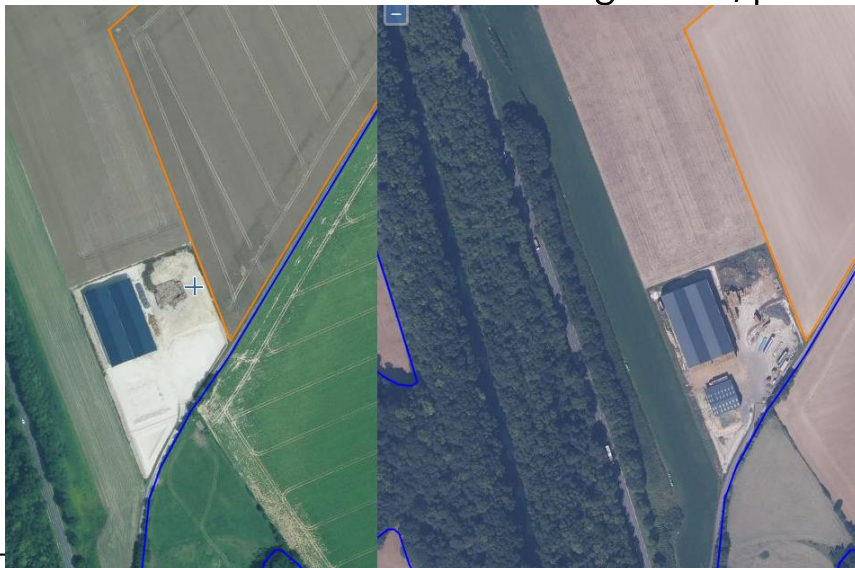
# Contrôle qualité sur l'OCS GE Finale : précision thématique millésime MAJ

- **Précision thématique du millésime INITIAL MIS A NIVEAU** (c'e-à-d corrigé des erreurs sur les objets mis à jour) :
  - 200 lignes de zone construite, touchant des objets mis à niveau
  - 300 objets mis à niveau
- **Précision thématique du millésime MIS A JOUR**
  - 200 lignes de zone construite, touchant des objets mis à jour
  - 100 objets en transition (US 6.1)
  - Parmi les objets mis à jour : tirage stratifié par valeur de couverture et d'usage (minimum 30 par strate)
  - Statistique différente de l'ini : taux d'accord global pondéré par le nombre d'objets de chaque strate. On attend 95 %.

# Contrôle qualité sur l'OCS GE Finale : exhaustivité de la mise à jour

## • Deux méthodes

- Tirage stratifié par valeur de couverture et d'usage pour voir si pas d'oubli de mise à jour, et statistique globale
- Gers : contrôle exhaustif sur une zone représentative du territoire (~5,7 % du Gers) par un logiciel dédié, en aveugle  
→ identification des zone de changement, puis comparaison au différentiel OCS GE.



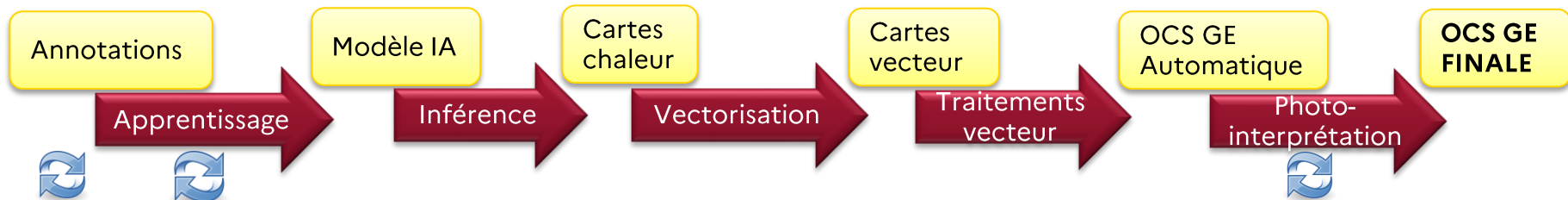
## Exemple de résultat

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

- **Analyse qualitative complémentaire + analyses ciblées** (ex. trous dans la zone construite, US 6.6, etc.)
- **Avantages :**
  - « Fait le job », vérifie la conformité du produit à ses spécifications, ni plus, ni moins
  - Méthode rôdée, utilisée pour l'OCS GE « classique » depuis le début de sa production
  - Bien intégrée aux outils de production, permet de travailler en délais contraints
- **Inconvénients :**
  - Pas d'évaluation de la qualité géométrique
  - Pas de stratification géographique
  - Pas d'évaluation des sous-détections
  - Des chiffres à la surface et non en nombre d'objets, pas intuitif
  - Pas de contrôle terrain
  - Mais tout ça n'est pas faisable dans le flux de production en délais contraints
  - Méthode complémentaire a posteriori, sur un nombre limité de départements

# 4. Qualification à froid

# La chaîne de production



QUALITÉ

QUALITÉ

Evaluation **indirecte**  
(évaluation de l'outil)

Evaluation **directe** de la  
conformité du produit aux  
spécifications (post-prod)

QUALITÉ

Evaluation **à froid** (autre  
méthode, autre regard)

QUALITÉ

- **Département contrôle qualité à l'IGN :**

- Effectue des contrôles qualité sur différentes productions IGN
- Indépendance des équipes de production
- Méthodologie complémentaire à celle des contrôles de fin de production
- Autres agents → nouveau regard
- Déplacement sur le terrain

- **Méthodes et outils développés pour l'OCS GE « classique »**

- Contrôle par département
- 7 contrôles départementaux réalisés entre 2015 et 2020
- Pause car mise au point du nouveau processus OCS
- Un nouveau contrôle sur un département produit avec la nouvelle chaîne en 2023, même méthode

# Echantillonnage

- **Strates constituées selon des critères de taille de zone construite**

- Strate 0 : pas de zone construite → paysage rural et de forêts
- Strate A :  $200 \text{ m}^2 < \text{surface Z.C.} < 1\,000 \text{ m}^2$  → paysage rural agricole
- Strate B :  $1\,000 \text{ m}^2 < \text{surface Z.C.} < 100\,000 \text{ m}^2$  → paysage de village
- Strate C :  $\text{Z.C.} > 100\,000 \text{ m}^2$  → paysage de grandes villes

- **Échantillonnage dans ces strates (tirage aléatoire)**

- **Contrôle géométrique** : 80 lignes de 50 m de Z.C., dans les strates A,B,C (donc 240 lignes en tout)
- **Contrôle de précision thématique** : 75 carrés de 250 x 250 m, dans les strates 0,A,B,C (donc 300 carrés)

- **Pondération des résultats** en fonction de la répartition de la population du département

# Précision géométrique de la Z.C.

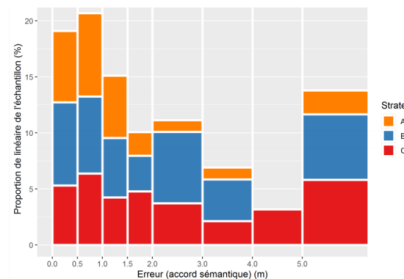
- **Objectif : vérifier la précision géométrique de la limite de zone construite**
- **Méthode**
  - 3 x 80 lignes de 50 m tirées dans les strates
  - Saisie d'une ligne équivalente sur ortho
  - Mesure d'écart de distance avec les données OCS GE
  - Les cas douteux ne sont pas comptabilisés
  - **Pondération du résultat final par la longueur de la population de chaque strate**

## LEGENDE :

- Ligne OCS initiale de l'échantillon 
- Ligne OCS Référence associée 
- Objet OCS 



## Exemple de résultat (OCS GE classique)



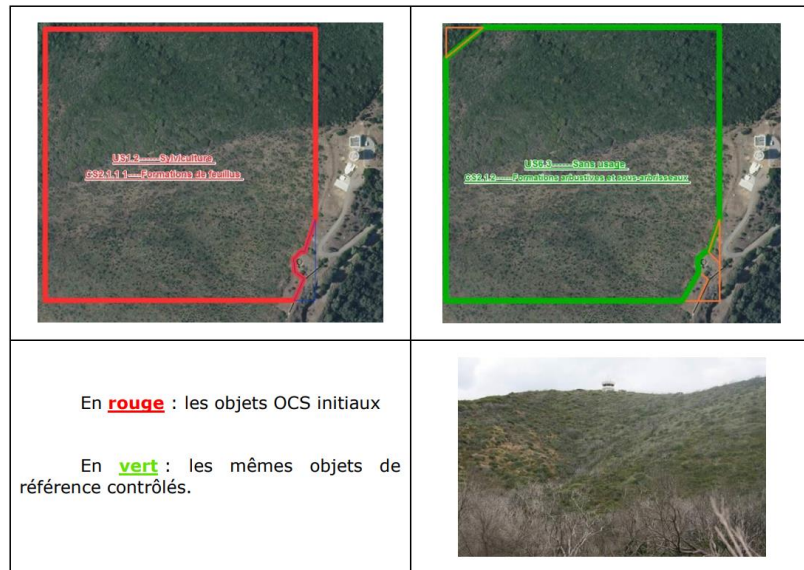
Graphique de répartition des écarts en accord sémantique

|                                                                                       |                                                                       |
|---------------------------------------------------------------------------------------|-----------------------------------------------------------------------|
| ÉCHANTILLON GÉOMÉTRIQUE DE DÉPART                                                     | Nombre de lignes : 240<br>soit une longueur totale de <b>11 993 m</b> |
| DONNÉES DE L'ÉCHANTILLON GÉOMÉTRIQUE<br>CODÉES LORS DU CQ EN « ACCORD<br>SÉMANTIQUE » | <b>69 %</b><br>soit une longueur totale de <b>8 303 m</b>             |
| Écart Moyen (m) : 4,4 m                                                               | Écart Type (m) : 1,1 m                                                |

# Précision thématique

Exemple de zone contrôlée sur le terrain

- **Objectif** : estimer le taux d'accord des champs couverture et usage
- **Méthode**
  - 4 x 75 carrés de 250x250 m tirés dans les strates
  - Au bureau : saisie des valeurs de couverture et d'usage des objets de l'échantillon, et redécoupage géométrique si nécessaire
  - En cas de doute : **passage terrain**
  - Calcul du taux d'accord, par valeur, et total + matrice de confusion
  - **Pondération du résultat final par la surface de la population de chaque strate**



Exemple de  
résultat (OCS  
GE classique)

| POPULATION OCS GE DE LA VENDÉE |                  |                     |                     |                          |                        |
|--------------------------------|------------------|---------------------|---------------------|--------------------------|------------------------|
| STRATES                        | Nombre de carrés | Surface totale (ha) | Taux d'accord USAGE | Taux d'accord Couverture | Taux d'accord CS et US |
| Strate 0                       | 62 811           | 381 222             | 97 %                | 99 %                     | 97 %                   |
| Strate A                       | 2 666            | 16 559              | 93 %                | 95 %                     | 90 %                   |
| Strate B                       | 29 926           | 186 133             | 92 %                | 95 %                     | 90 %                   |
| Strate C                       | 18 055           | 112 384             | 82 %                | 92 %                     | 77 %                   |
| <b>TOTAL POPULATION</b>        | <b>113 458</b>   | <b>696 298</b>      | <b>93 %</b>         | <b>97 %</b>              | <b>92 %</b>            |

- **Au-delà des chiffres, des retours utiles pour les équipes de production.**

**Exemples** (sur OCS « classique », non IA) :

- Pb de détection des campings arborés, de l'usage logistique
- RETEX sur la compréhension des spécifications (vergers, usage des surfaces d'eau)
- Etc.

- **Avantages :**

- Evaluation de la qualité géométrique
- Stratification géographique
- Evaluation des sous-détections
- Des chiffres à la surface et non en nombre d'objets
- Contrôle terrain

- **Inconvénients :**

- Chronophage ; impossible de contrôler exhaustivement tous les départements
- Pas de contrôle de la mise à jour : **évolution du processus nécessaire quand la France aura été couverte par le millésime initial**

# 5. Conclusion

- Des procédures à différentes étapes de la chaîne, qui permettent de **sécuriser la production**
- Une **capitalisation des développements** déjà en place pour l'OCS GE « classique »
- Un soin particulier pour **l'évaluation du modèle IA**, en ayant bien conscience que c'est une **évaluation indirecte** :
  - L'évaluation IA est indépendante de la qualité du produit final OCS GE
  - IA : évaluation exhaustive sur des pixels, avec des effets de bord
  - OCS GE : évaluation statistique sur des objets vecteur, avec des effets de seuils et une variabilité
  - Objectif à terme : mieux mesurer la corrélation entre métriques IA et OCS GE (besoin de plus de recul)
- **A l'écoute des utilisateurs** des départements concernés par les productions planifiées (*qualité perçue*) :
  - Sur les annotations
  - Sur les cartes de chaleur et l'OCS tout automatique, en avance de phase
  - Versions intermédiaires photo-interprétées disponibles sur l'espace collaboratif

# MERCI DE VOTRE ATTENTION

# ANNEXES

# Nomenclature des annotations

| NAF19 | label               | Classe OCS GE                          |  |
|-------|---------------------|----------------------------------------|--|
| 1     | Batiment            | Zones bâties                           |  |
| 2     | Zone perméable      | Zones à matériaux minéraux             |  |
| 3     | Zone imperméable    | Zones non bâties                       |  |
| 4     | Piscine             | Non exploité (en réserve pour l'usage) |  |
| 5     | Sol nu              | Sols nus                               |  |
| 6     | Surface d'eau       | Surfaces d'eau                         |  |
| 7     | Neige               | Névés et glacier                       |  |
| 8     | Conifères           | Formations arborées                    |  |
| 9     | Coupe               |                                        |  |
| 10    | Feuillus            |                                        |  |
| 11*   | Mixte               | Non appris                             |  |
| 12*   | Ligneux             | Non appris                             |  |
| 13    | Broussaille         | Formation arbustives, sous-arbrisseaux |  |
| 14    | Vigne               | Autres formations ligneuses            |  |
| 15    | Pelouse             | Formations herbacées                   |  |
| 16    | Culture             |                                        |  |
| 17    | Terre labourée      |                                        |  |
| 18    | Serre               | Zones bâties                           |  |
| 19*   | Autres (non annoté) | Non appris                             |  |

# Comment évaluer le modèle

- Pour un apprentissage, on répartit des annotations en trois jeux : **train (apprentissage) / val (validation) / test**
  - **train/val** : sert à entraîner le modèle = régler plusieurs millions de paramètres d'un réseau de neurones en N itérations (« époque »). On réserve 80 % des annotations.
    - *train (apprentissage)* : pendant une itération, sert à ajuster les paramètres. 60% des annotations.
    - *val (validation)* : à la fin d'une itération, sert à valider le paramétrage sur d'autres annotations. Permet de faire converger le réseau et choisir la meilleure époque. 20 % des annotations.
  - **test** : ne sert pas à ajuster le réseau de neurones, mais à l'évaluer sous sa forme finale, sur d'autres annotations. 20% des annotations.
- **Biais** dû à cette répartition. Pour l'éviter, utilisation d'une méthode maîtrisée et éprouvée par la littérature : **la validation croisée**.

# Comment évaluer le modèle

- **Méthode de validation croisée « K Fold »** : pour ne pas avoir d'effet aléatoire dû à la répartition des annotations train/val/test, consiste à entraîner et évaluer 5 modèles. Cette méthode permet donc de produire des **indicateurs de qualité sur tout l'ensemble du jeu de données annotées**. Elle est bien éprouvée par la littérature.

| Modèle | FOLD-1 | FOLD-2 | FOLD-3 | FOLD-4 | FOLD-5 |
|--------|--------|--------|--------|--------|--------|
| SET1   | TEST   | VALID  | TRAIN  | TRAIN  | TRAIN  |
| SET2   | TRAIN  | TEST   | VALID  | TRAIN  | TRAIN  |
| SET3   | TRAIN  | TRAIN  | TEST   | VALID  | TRAIN  |
| SET4   | TRAIN  | TRAIN  | TRAIN  | TEST   | VALID  |
| SET5   | VALID  | TRAIN  | TRAIN  | TRAIN  | TEST   |

