



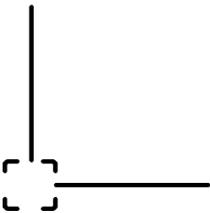
# **COSIA**

Version 1.0

Couverture du Sol par  
Intelligence Artificielle

Descriptif de contenu

Date du document : Avril 2023  
Révision : Mai 2025



# Sommaire

<b>Sommaire</b> .....	2
<b>1. Présentation de CoSIA</b> .....	3
1.1 Définition et contenu .....	3
1.2 Usages .....	3
1.3 Millésime et mise à jour .....	3
<b>2. Caractéristiques techniques</b> .....	4
2.1 Sources des données .....	4
2.2 Extension géographique .....	4
2.3 Références géodésiques .....	4
2.4 Production de CoSIA .....	4
2.5 Paramètres de qualité .....	4
<b>Annexe 1 – Qu'est-ce que CoSIA ?</b> .....	5
<b>Annexe 2 – Quelles sont les couvertures décrites par CoSIA ?</b> .....	6
<b>Annexe 3 – À partir de quelles données est produite CoSIA ?</b> .....	9
<b>Annexe 4 – Cartes des millésimes</b> .....	10
<b>Annexe 5 – Comment sont produites les cartes CoSIA ?</b> .....	11
<b>Annexe 6 – Pourquoi certaines données sont-elles erronées ?</b> .....	13
<b>Annexe 7 – À quels métriques se référer ?</b> .....	14
<b>Annexe 8 – Métriques de qualité par classe</b> .....	15
<b>Lexique</b> .....	19

# 1. Présentation de CoSIA

## 1.1 Définition et contenu

CoSIA (Couverture du Sol par Intelligence Artificielle) est un produit issu de la base de données OCS GE (voir également l'[Annexe 1 – Qu'est-ce que CoSIA ?](#))

Cette base de données décrit la couverture du sol selon 15 classes :

- Bâtiment
- Zone imperméable, Zone perméable
- Piscine, Surface eau, Neige
- Serre, Sol nu
- Conifère, Feuillu, Broussaille, Pelouse, Culture, Terre labourée, Vigne

L'ensemble de ces classes est décrit en détail à l'[Annexe 2 – Quelles sont les couvertures décrites par CoSIA ?](#)

## 1.2 Usages

La résolution spatiale et la finesse sémantique de cette base de données peut servir d'habillage à divers représentations cartographiques, faciliter la production de bases de données thématiques comme la réalisation de Mode d'occupation des sol (MOS) ou OCS GE ++ (plus détaillé) et servir comme donnée d'entrée au suivi de politiques publiques autour de la végétation en ville, des haies et bocages ou encore des trames vertes et bleues.

## 1.3 Millésime et mise à jour

CoSIA est actuellement disponible sur deux millésimes de prise de vues aériennes (BD ORTHO®).

Des cartes de disponibilité des millésimes par département ayant servi à calculer le produit CoSIA sont fournies en [Annexe 4 – Cartes des millésimes](#).

Vous trouverez dans les liens ci-dessous des cartes indiquant le millésime précis disponible. CoSIA sera calculé à chaque nouveau millésime de prise de vues aériennes.

## 2. Caractéristiques techniques

### 2.1 Sources des données

Les données utilisées pour la constitution du produit CoSIA, aussi appelées « données entrantes », sont issues de différentes sources (BD ORTHO®, MNT, ...), décrites en détail à l'**Annexe 3 – À partir de quelles données est produite CoSIA ?**

### 2.2 Extension géographique

CoSIA est produite pour tout le territoire français (métropole et DROM) à grande résolution

### 2.3 Références géodésiques

Elles sont fixées par l'arrêté du 5 mars 2019 :

[https://www.legifrance.gouv.fr/affichTexteArticle.do;jsessionid=6204032AC96689C6D5C43DB3320786C5.tplgfr33s\\_2?idArticle=JORFARTI000038203607&cidTexte=JORFTEXT000038203601&dateTexte=29990101&categorieLien=id](https://www.legifrance.gouv.fr/affichTexteArticle.do;jsessionid=6204032AC96689C6D5C43DB3320786C5.tplgfr33s_2?idArticle=JORFARTI000038203607&cidTexte=JORFTEXT000038203601&dateTexte=29990101&categorieLien=id)

### 2.4 Production de CoSIA

Les étapes de production et de simplification du produit CoSIA sont décrites en détail à l'**Annexe 5 – Comment est produite CoSIA ?**

### 2.5 Paramètres de qualité

Ces données sont brutes et obtenues par des processus d'intelligence artificielle (IA) notamment des méthodes d'apprentissage profond (*deep learning*). Elles représentent des informations dites de "prédiction". Elles sont obtenues à partir d'un modèle numérique d'IA qui estime statistiquement pour chaque pixel son appartenance à une classe.

Ces données ne reflètent donc pas de manière exhaustive la réalité du terrain.

Il existe des marges d'erreurs qui sont documentées dans l'**Annexe 6 – Pourquoi certaines données sont-elles erronées ?**

Les métriques de qualité sont obtenues à partir d'un jeu de données représentatif constitué de 20% des données d'annotations, les 80% restant sont utilisés pour l'entraînement et la validation du modèle. La définition des métriques réalisées sont détaillées dans l'**Annexe 7 – À quels métriques se référer ?**

Les résultats moyens obtenus sur l'ensemble des classes sont les suivants :

- Rappel : 81,0%
- Précision : 79,5%
- IoU : 68,1%

Les métriques de qualité sont détaillées en **Annexe 8 – Métriques de qualité par classe.**

## Annexe 1 – Qu'est-ce que CoSIA ?

Les cartes CoSIA décrivent la couverture du sol, soit la nature du sol, selon 15 classes (bâtiment, surface d'eau, conifère, culture, broussaille, ...). Cette description du sol est produite pour tout le territoire français (métropole et DROM-COM) avec une haute résolution de 20 cm par pixel.

Les cartes CoSIA sont un produit de l'IGN qui interviennent actuellement dans la conception de l'OCS GE. Leur résolution spatiale et leur finesse sémantique peuvent également aider dans la production d'autres cartographies et au calcul d'autres indicateurs comme la végétation en ville, les haies & bocages, les trames vertes & bleues ou encore intervenir dans la réalisation de MOS locaux ou d'un OCS GE plus fin.

Pour produire ces cartes, des processus d'intelligence artificielle sont utilisés, dont des méthodes d'apprentissage profond (*deep learning*). Ces cartes sont alors dites de "prédiction" car elles sont obtenues à partir d'un modèle numérique d'IA qui estime statistiquement pour chaque pixel son appartenance à une classe, et peuvent ne pas refléter de manière exhaustive la réalité du terrain. Il existe des marges d'erreurs qui sont référencées pour chaque classe.



### Légende

- Bâtiment
- Zone imperméable
- Zone perméable
- Piscine
- Serre
- Sol nu
- Surface eau
- Neige
- Conifère
- Feuillu
- Broussaille
- Pelouse
- Culture
- Terre labourée
- Vigne
- Autre

## Annexe 2 – Quelles sont les couvertures décrites par CoSIA ?

### Nomenclature

Les cartes CoSIA sont composées de 15 classes : bâtiment, zone imperméable, zone perméable, piscine, serre, sol nu, surface d'eau, neige, conifère, feuillus, broussaille, pelouse, culture, terre labourée, vigne et autre. Elles sont définies et illustrées dans la table ci-dessous.

CLASSE	DESCRIPTION	ILLUSTRATION
 <b>Bâtiment</b>	Bâtiment ou autres types de constructions. <u>Exemples</u> : tours, châteaux d'eau, silo, auvents, ... <u>Attention</u> : un simple mur n'est pas considéré comme un bâtiment.	
 <b>Zone imperméable</b>	Zone non construite, munie d'un revêtement la rendant imperméable (asphalte, béton, ...). <u>Exemples</u> : route, terrain de sport revêtu, parking, ...	
 <b>Zone perméable</b>	Terrain stabilisé et compacté, partiellement ou totalement perméable, et recouvert de matériaux minéraux (pierres, terre, graviers). <u>Exemples</u> : voies ferrées, chemins, carrières, salines, chantiers, enrochements rocheux, cours de ferme, cimetières (sauf zones goudronnées ou végétalisées), terrains de sports non revêtus, ...	
 <b>Piscine</b>	Bassin de piscine non couverte.	

● **Serre**

Construction, pérenne ou non, en verre ou en plastique, translucide, close ou couverte à vocation le plus souvent agricole.



● **Sol nu**

Zone naturelle non végétalisée.  
Exemples : sable, galets, rochers, lapiaz, ...  
Peuvent aussi être classées en sol nu, d'anciennes zones artificialisées retournant à la nature (carrières).



● **Surface eau**

Surface naturelle couvertes d'eau au moment de la prise de vues aériennes.



● **Neige**

Surface couverte de neige ou de glace au moment de la prise de vues aériennes.



● **Conifère**

Peuplement de conifères ou conifères isolés.



● **Feuilleu**

Peuplement de feuillus ou feuillus isolés.



● **Broussaille**

Terrain couvert d'arbustes et de sous-arbrisseaux.

Exemples : landes montagnardes, garrigues, maquis non boisés, terrains en friches, landes marécageuses, ...



● **Pelouse**

Surface non agricole couverte de végétation herbacée.

Exemples : pelouse naturelle, d'agrément, terrain de sport, clairière, ...



● **Culture**

Zone herbacée à vocation agricole incluant les cultures et les prairies.



● **Terre labourée**

Zone de culture sans végétation au moment de la prise de vues aériennes.



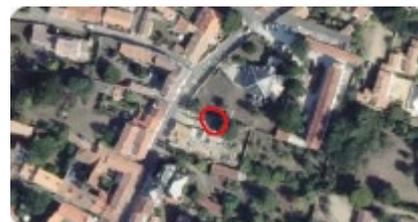
● **Vigne**

Plantations de vignes.



● **Autre**

Zone non classée (classe inconnue, ombre dense, ...).



## Annexe 3 – À partir de quelles données est produite CoSIA ?

### Données entrantes

Les cartes CoSIA sont produites à partir de trois bases de données de l'IGN: la BD ORTHO®, le MNT RGE-ALTI® et le MNS ainsi qu'à partir d'annotations manuelles.



### Annotations

L'IGN a fait produire 2500 km<sup>2</sup> d'annotations manuelles à travers 63 départements. À partir des prises de vues aériennes, chaque pixel est annoté et appartient à l'une des 15 classes de couverture du sol.

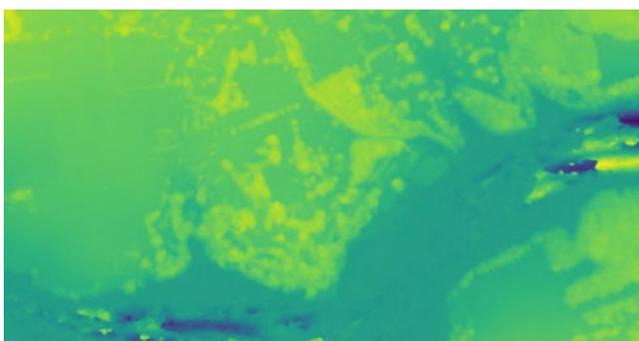
<https://geoservices.ign.fr/ressources-ia-de-couverture-du-sol>



### Données Ortho

La BD ORTHO® 20 cm (RVB & IRC) est utilisée. Elle est produite tous les trois ans (1/3 des départements français par an) à partir des prises de vues aériennes opérées par l'IGN. Les canaux rouge, vert et bleu sont utilisés ainsi que l'infrarouge pour détecter la végétation.

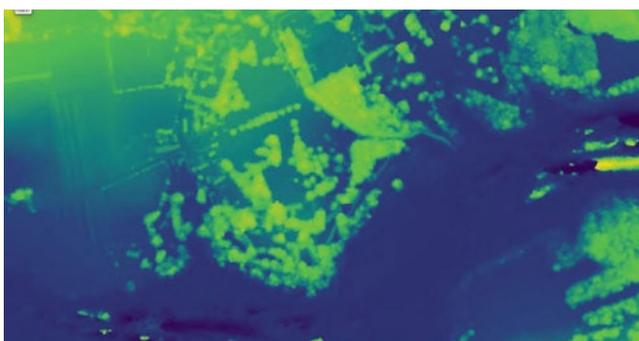
<https://geoservices.ign.fr/bdortho>



### MNS

Le modèle numérique de surface (**MNS**) maillé décrit le relief du territoire : la forme du sol et du sursol. En plus du terrain, est fournie l'altitude de l'ensemble des éléments au-dessus du sol comme la végétation, les bâtiments ou les ouvrages d'art.

<https://geoservices.ign.fr/rgealti>



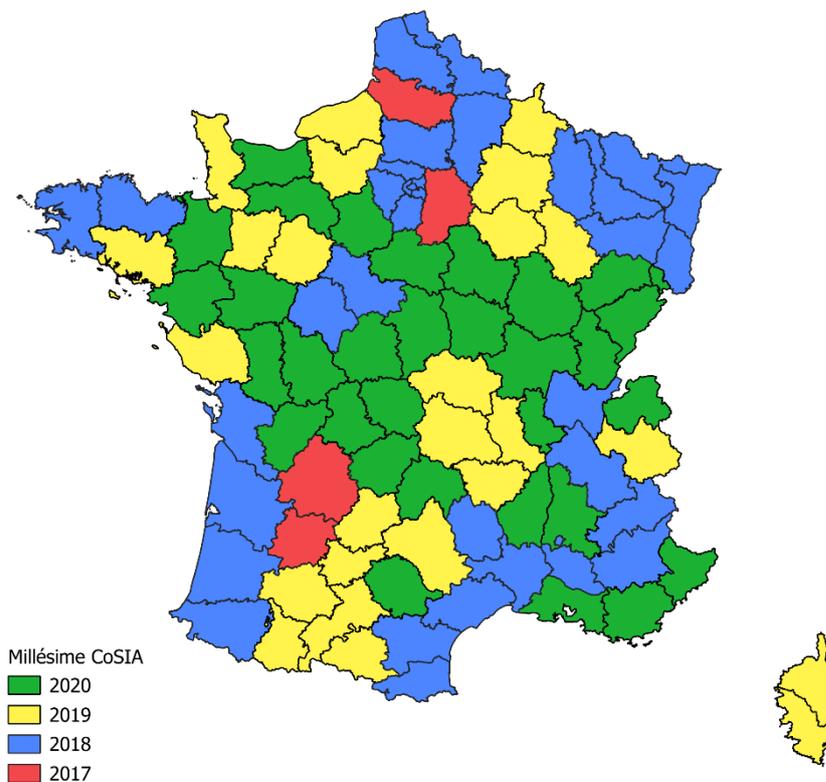
### MNT RGE ALTI®

Le modèle numérique de terrain (**MNT**) maillé décrit le relief du territoire : la forme et l'altitude normale de la surface du sol.

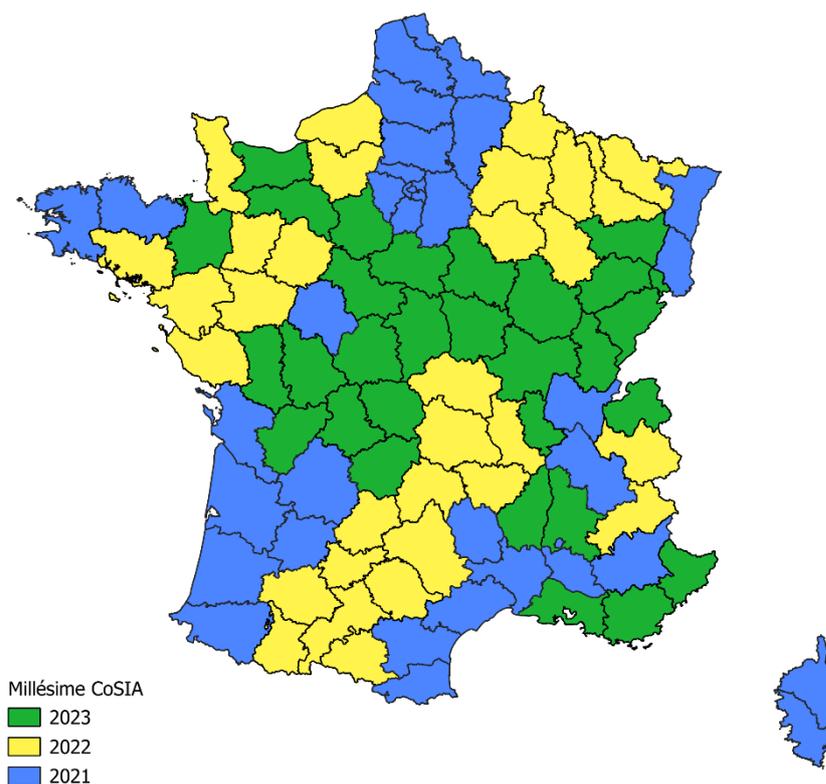
<https://geoservices.ign.fr/rgealti>

## Annexe 4 – Cartes des millésimes

### Millésimes utilisés pour le calcul du CoSIA. Millésime 2017-2020



### Millésimes utilisés pour le calcul du CoSIA. Millésime 2021-2023



# Annexe 5 – Comment sont produites les cartes CoSIA ?

La production de COSIA en 8 étapes.

1

## Prises de vues aériennes

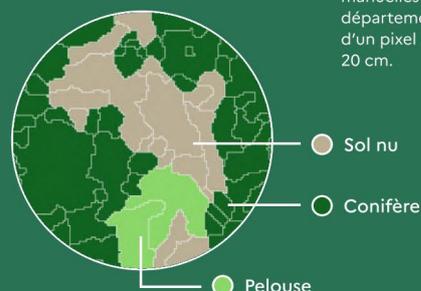
La production de CoSIA se fait par département et à partir des données ortho issues des prises de vues aériennes (un tiers des départements sont couverts chaque année).



2

## Annotation du territoire

À partir des données ortho, on crée des groupes de pixels dont les couleurs sont proches (on parle de segmentation d'images) et l'on assigne manuellement à chaque groupe de pixel l'une des 16 classes de couverture du sol.



Q L'IGN a fait produire 2500 km<sup>2</sup> d'annotations manuelles à travers 63 départements. La taille d'un pixel est de 20 cm sur 20 cm.

3



## Génération d'un jeu de données

Les annotations au format raster sont regroupées avec la BD Ortho RVB IR, le MNT et le MNS pour obtenir un jeu de données. On a alors pour chaque pixel un certain nombre d'informations.

4

## Conception d'un modèle d'intelligence artificielle

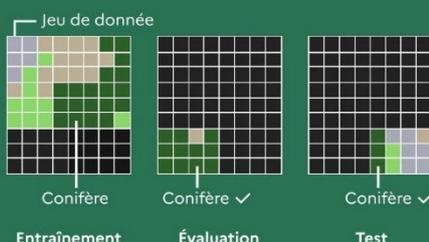
On crée un modèle numérique (un ensemble de règles) et on utilise pour cela l'intelligence artificielle et des méthodes d'apprentissage profond (deep learning). Ces technologies servent à reconnaître automatiquement des objets sur un jeu de données (ici bâtiment, arbre, plan d'eau...).

5

## Entraînement du modèle

Le modèle d'intelligence artificielle prédit pour chaque pixel du jeu de données la probabilité qu'il appartienne à chacune des classes. Le modèle s'entraîne : il teste et affine un très grand nombre de paramètres jusqu'à ce qu'il produise un résultat le plus proche des annotations.

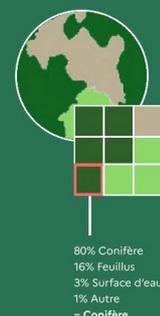
Q Pour l'entraînement du modèle, le jeu de données est divisé en trois parties. Le modèle s'entraîne sur la partie 1. Il est ensuite appliqué sur la partie 2 pour évaluer les paramètres choisis et est enfin testé sur la partie 3 pour s'assurer de sa performance.



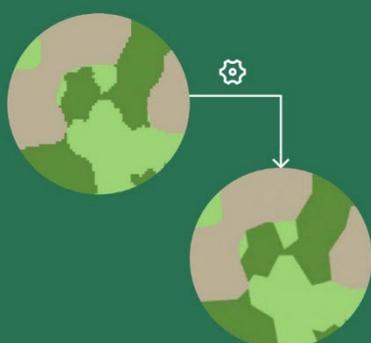
6

## Inférence France entière

Le modèle, correctement paramétré, est appliqué (ou inféré) sur l'ensemble de la France. Le pourcentage le plus élevé détermine dans chaque cas la classe du pixel.



7



## Vectorisation

En data-science, les données obtenues grâce au modèle sont appelées cartes de de prédiction (car le modèle produit une estimation). Ces données, sous format raster, sont ensuite simplifiées et vectorisées (suppression des pixels isolés et lissage des contours des polygones). Cette simplification réduit fortement le poids de la donnée et offre une meilleure usabilité.

8

## Mise à disposition

CoSIA est disponible depuis les sites de l'IGN. La donnée millésimée est mise à jour tous les trois ans pour un même département.



## Q Zoom sur la simplification

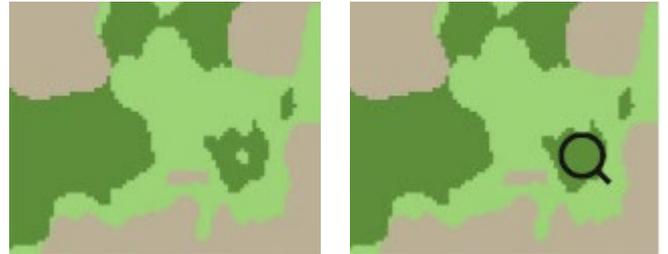
### Pourquoi simplifier la donnée ?

Pour répondre aux usages métiers, la donnée est vectorisée. Mais la résolution initiale des pixels de 20 cm par 20 cm alourdit considérablement la donnée. On procède alors à une simplification pour que son poids reste

raisonnable et qu'elle puisse facilement être exportée et manipulée au sein de logiciels SIG. Par exemple, la métropole de Tours sans simplification est d'environ 1.6 Gb. Avec le procédé de simplification, on réduit son poids à 186 Mb, soit par 8.

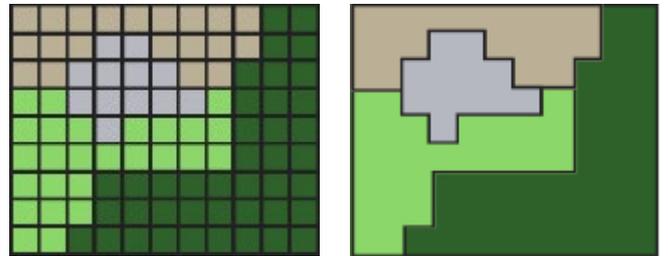
#### 1 Suppression de pixels

On supprime les pixels isolés et les petits groupes de pixels par méthode Sieve. Seules les petites surfaces, correspondant à des objets inférieurs à 1m, sont supprimées.



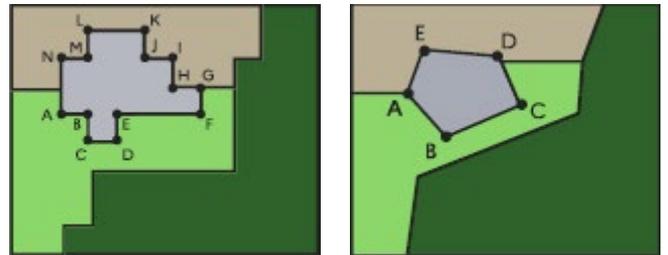
#### 2 Vectorisation de la donnée

Les données sont vectorisées. Les polygones suivent les formes des pixels et ont une forme "d'escalier". Cette opération participe à alourdir la donnée.



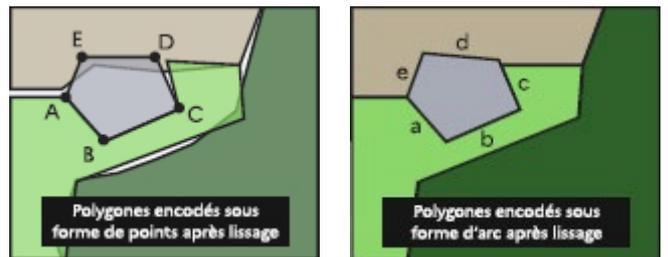
#### 3 Simplification des polygones

Les polygones sont lissés en suivant l'algorithme de Douglas-Peucker avec un coefficient de 1 m. La valeur indique approximativement la distance maximale qu'un contour peut changer après lissage.



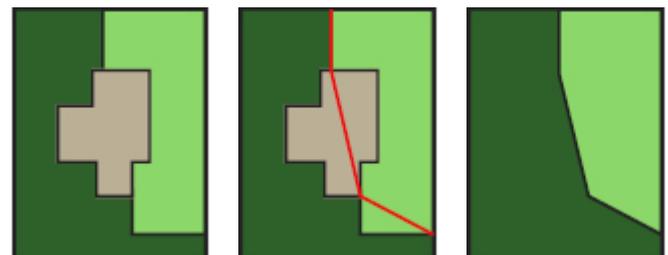
#### + Cohérence Topologique

Pour garder une cohérence topologique entre les polygones adjacents lors du lissage et éviter des trous ou des recouvrements, les polygones sont encodés sous une suite d'arcs au lieu de points. Puis, après lissage, ils sont de nouveau encodés sous une suite de points standards pour faciliter leur manipulation.



#### + Effets de bord

On note parfois dans la méthode utilisée, des effets de bords. Des polygones avec une géométrie particulière et une largeur étroite (inférieur à deux fois le coefficient de lissage, soit environ 1m) sont supprimés ou il existe des recouvrements. Mais ces cas restent rares.



## Annexe 6 – Pourquoi certaines données sont-elles erronées ?

### 5 paramètres à prendre en compte

pour mieux comprendre les technologies d'intelligence artificielles et de *deep learning* et leurs limites.

#### 1 Les données d'entrée

Les modèles d'intelligence artificielle sont des modèles qui appliquent des traitements à une donnée. Si l'on a une donnée entrante qui possède des biais ou des inexactitudes, ceux-ci se retrouveront au sein des données sortantes. S'il y a des erreurs dans les annotations manuelles ou les bases de données entrantes comme le MNS, celles-ci induiront des résultats erronés dans CoSIA.

#### 2 La caractérisation des données

Pour le modèle, certaines classes sont plus faciles que d'autres à interpréter par leur couleur, altitude, géométrie ou voisinage. Les bâtiments ou les piscines sont par exemple facilement reconnaissables par leur couleur, altitude et géométrie. En revanche, les cultures, broussailles et pelouses peuvent être plus difficiles à distinguer par leurs géométries plus organiques et leurs couleurs similaires.

#### 3 La diversité des données

Certaines classes sont plus représentées sur l'ensemble du territoire comme les bâtiments ou les cultures. Le modèle aura pu identifier une diversité de représentations de ces classes. Au contraire, le modèle aura été moins exposé aux neiges ou aux serres et aura plus de mal à classifier les pixels s'y apparentant.

#### 4 Le classement par vue aérienne

Une autre problématique liée au classement est la nature de l'image. Si un parking est recouvert par de la végétation, par exemple, les feuilles d'un arbre, sa surface sera réduite ou morcelée. Le modèle n'est entraîné que pour détecter ce qui est visible sur les images. Les cartes de couverture du sol sont donc à utiliser avec d'autres bases de données pour obtenir des cartographies cohérentes.

#### 5 Les zones sensibles

Enfin, certaines zones considérées comme sensibles tels que les aéroports ou les bases militaires sont floutées dans les données entrantes. Le modèle ne peut prédire pour ces zones des données correctes.

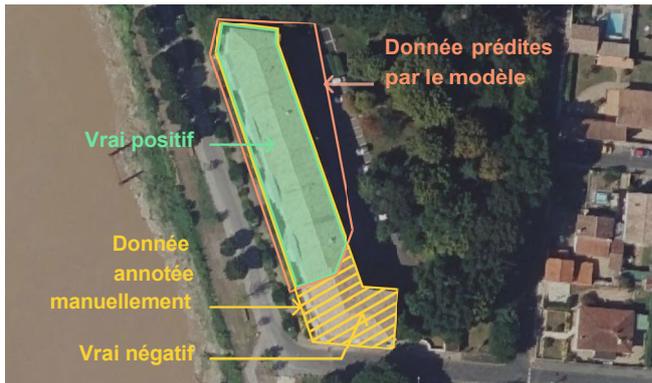


Exemple d'une route morcelée

## Annexe 7 – À quels métriques se référer ?

### Métriques

Les métriques disponibles mesurent la performance du modèle et sont calculés sur les annotations réparties sur la France entière. Ils ne prévalent donc pas pour un département en particulier.

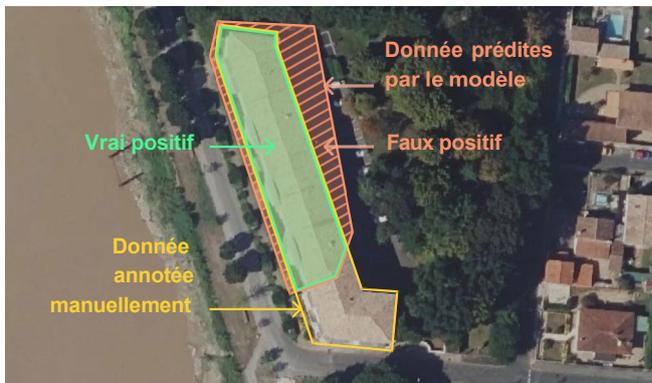


### Rappel

(ou sous-détection)

Ce métrique indique le nombre de pixels bien classés sur l'ensemble des pixels qui ont été annotés.

$$\frac{\text{Vrai positif}}{\text{Vrai positif} + \text{Vrai négatif}}$$

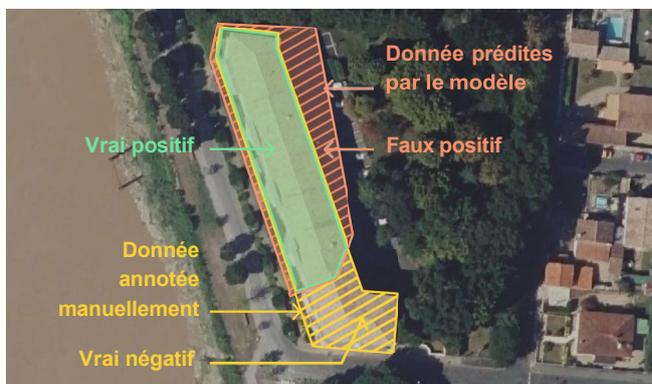


### Précision

(ou sur-détection)

Ce métrique indique le nombre de pixels bien classés sur l'ensemble des pixels prédits par le modèle.

$$\frac{\text{Vrai positif}}{\text{Vrai positif} + \text{Faux positif}}$$



### IoU

Intersection over Union

Ce métrique indique le nombre de pixels bien classés sur l'ensemble des pixels qui ont été annotés et des pixels prédits par le modèle.

$$\frac{\text{Vrai positif}}{\text{Vrai positif} + \text{Vrai négatif} + \text{Faux positif}}$$

## Annexe 8 – Métriques de qualité par classe

### Rappel

France entière - Millésimes : 2017- 2023

CLASSE	CONCLUSIONS GÉNÉRALES
 <b>Bâtiment</b>	Parmi tous les pixels annotés comme bâtiment dans les annotations de référence : <b>92 %</b> ont été prédits par le modèle comme bâtiment.
 <b>Zone imperméable</b>	Parmi tous les pixels annotés comme zone imperméable dans les annotations de référence : <b>87,7 %</b> ont été prédits par le modèle comme zone imperméable.
 <b>Zone perméable</b>	Parmi tous les pixels annotés comme zone perméable dans les annotations de référence : <b>73,1%</b> ont été prédits par le modèle comme zone perméable.
 <b>Piscine</b>	Parmi tous les pixels annotés comme piscine dans les annotations de référence : <b>78,2 %</b> ont été prédits par le modèle comme piscine.
 <b>Serre</b>	Parmi tous les pixels annotés comme serre dans les annotations de référence : <b>75,5 %</b> ont été prédits par le modèle comme serre.
 <b>Sol nu</b>	Parmi tous les pixels annotés comme sol nu dans les annotations de référence : <b>85,7%</b> ont été prédits par le modèle comme sol nu.
 <b>Surface eau</b>	Parmi tous les pixels annotés comme surface d'eau dans les annotations de référence : <b>94,9 %</b> ont été prédits par le modèle comme surface d'eau.
 <b>Neige</b>	Parmi tous les pixels annotés comme neige dans les annotations de référence : <b>89,3 %</b> ont été prédits par le modèle comme neige.
 <b>Conifère</b>	Parmi tous les pixels annotés comme conifère dans les annotations de référence : <b>79,0 %</b> ont été prédits par le modèle comme conifère.

 <b>Feuille</b>	Parmi tous les pixels annotés comme feuillu dans les annotations de référence : <b>87,7 %</b> ont été prédits par le modèle comme feuillu.
 <b>Broussaille</b>	Parmi tous les pixels annotés comme broussaille dans les annotations de référence : <b>55,9 %</b> ont été prédits par le modèle comme broussaille.
 <b>Pelouse</b>	Parmi tous les pixels annotés comme pelouse dans les annotations de référence : <b>80,6 %</b> ont été prédits par le modèle comme pelouse.
 <b>Culture</b>	Parmi tous les pixels annotés comme culture dans les annotations de référence : <b>71,0 %</b> ont été prédits par le modèle comme culture.
 <b>Terre labourée</b>	Parmi tous les pixels annotés comme terre labourée dans les annotations de référence : <b>70,1 %</b> par le modèle comme terre labourée.
 <b>Vigne</b>	Parmi tous les pixels annotés comme vigne dans les annotations de référence : <b>94,6%</b> par le modèle comme vigne.

## Précision

France entière Millésimes : 2017- 2023

 CLASSE	CONCLUSIONS GÉNÉRALES
 <b>Bâtiment</b>	Parmi tous les pixels prédits comme bâtiment par le modèle : <b>91,4 %</b> sont annotés comme bâtiment dans les annotations de référence.
 <b>Zone imperméable</b>	Parmi tous les pixels prédits comme zone imperméable par le modèle : <b>84,7 %</b> sont annotés comme zone imperméable dans les annotations de référence.
 <b>Zone perméable</b>	Parmi tous les pixels prédits comme zone perméable par le modèle : <b>78,5 %</b> sont annotés comme zone perméable dans les annotations de référence.
 <b>Piscine</b>	Parmi tous les pixels prédits comme piscine par le modèle : <b>71,7 %</b> sont annotés comme piscine dans les annotations de référence.

 <b>Serre</b>	Parmi tous les pixels prédits comme serre par le modèle : <b>61,6 %</b> sont annotés comme serre dans les annotations de référence.
 <b>Sol nu</b>	Parmi tous les pixels prédits comme sol nu par le modèle : <b>85,1 %</b> sont annotés comme sol nu dans les annotations de référence.
 <b>Surface eau</b>	Parmi tous les pixels prédits comme surface d'eau par le modèle : <b>96,1 %</b> sont annotés comme surface d'eau dans les annotations de référence.
 <b>Neige</b>	Parmi tous les pixels annotés comme neige dans les annotations de référence : <b>86,6 %</b> sont annotés comme neige dans les annotations de référence.
 <b>Conifère</b>	Parmi tous les pixels prédits comme conifère par le modèle : <b>82,0 %</b> sont annotés comme conifère dans les annotations de référence.
 <b>Feuillu</b>	Parmi tous les pixels prédits comme feuillu par le modèle : <b>86,6 %</b> sont annotés comme feuillu dans les annotations de référence.
 <b>Broussaille</b>	Parmi tous les pixels prédits comme broussaille par le modèle : <b>58,1 %</b> sont annotés comme broussaille dans les annotations de référence.
 <b>Pelouse</b>	Parmi tous les pixels prédits comme pelouse par le modèle : <b>68,5 %</b> sont annotés comme pelouse dans les annotations de référence.
 <b>Culture</b>	Parmi tous les pixels prédits comme culture par le modèle : <b>85,4 %</b> sont annotés comme culture dans les annotations de référence .
 <b>Terre labourée</b>	Parmi tous les pixels prédits comme terre labourée par le modèle : <b>61,9 %</b> sont annotés comme terre labourée dans les annotations de référence.
 <b>Vigne</b>	Parmi tous les pixels prédits comme vigne par le modèle : <b>93,8 %</b> sont annotés comme vigne dans les annotations de référence.

# IoU

France entière Millésimes : 2017 - 2023

---

## Données du modèle

IoU par classe en %

CLASSE	%
Bâtiment	84,7
Zone imperméable	75,7
Zone perméable	60,9
Piscine	59,7
Serre	51,3
Sol nu	74,5
Surface eau	91,4
Neige	78,4
Conifère	67,3
Feuillu	77,2
Broussaille	39,9
Pelouse	58,8
Culture	63,3
Terre labourée	48,9
Vigne	88,9

---

## Conclusions

- Le modèle prédit avec précision les classes « bâtiment », « zone imperméable », « surface d'eau », « vignes » et « feuillu ».
- Le modèle prédit avec plus de difficulté les classes « serre », « pelouse », « broussaille » et « terre labourée ».

# Lexique

Ces définitions proviennent du glossaire de la CNIL et sont disponibles à cette URL : <https://www.cnil.fr/fr/glossaire>.

## Annotation

L'annotation est le procédé par lequel les données sont décrites manuellement afin d'être caractérisées, par exemple en attribuant à une image de chien l'étiquette correspondante. On parle aussi de vérité terrain ou groundtruth.

## Apprentissage automatique (ou *Machine learning*)

L'apprentissage automatique (*machine learning* en anglais) est un champ d'étude de l'intelligence artificielle qui vise à donner aux machines la capacité d'« apprendre » à partir de données, via des modèles mathématiques. Plus précisément, il s'agit du procédé par lequel les informations pertinentes sont tirées d'un ensemble de données d'entraînement. Le but de cette phase est l'obtention des paramètres d'un modèle qui atteindront les meilleures performances, notamment lors de la réalisation de la tâche attribuée au modèle. Une fois l'apprentissage réalisé, le modèle pourra ensuite être déployé en production.

## Apprentissage profond (ou *Deep learning*)

L'apprentissage profond est un procédé d'apprentissage automatique (*machine learning*) utilisant des réseaux de neurones possédant plusieurs couches de neurones cachées. Ces algorithmes possédant de très nombreux paramètres, ils demandent un nombre très important de données afin d'être entraînés.

## Donnée d'entrée

Dans le domaine de l'intelligence artificielle, une donnée d'entrée est une donnée utilisée pour l'apprentissage automatique ou la prise de décision du système d'IA (en phase de production).

## Donnée de sortie

Dans le domaine de l'intelligence artificielle, une donnée de sortie est une valeur représentant tout ou partie de l'opération effectuée par le système d'IA à partir des données d'entrée.

## Échantillon

Dans le domaine de l'intelligence artificielle, l'échantillon est une fraction représentative d'une population ou d'un univers statistique.

## Entraînement

L'entraînement est le processus de l'apprentissage automatique pendant lequel le système d'intelligence artificielle construit un modèle à partir de données.

## Intelligence artificielle

L'intelligence artificielle est un procédé logique et automatisé reposant généralement sur un algorithme et en mesure de réaliser des tâches bien définies. Pour le Parlement européen, constitue une intelligence artificielle tout outil utilisé par une machine afin de « reproduire des comportements liés aux humains, tels que le raisonnement, la planification et la créativité ».

## Modèle

Le modèle d'IA est la construction mathématique générant une déduction ou une prédiction à partir de données d'entrée. Le modèle est estimé à partir de données annotées lors de la phase d'apprentissage (ou d'entraînement) du système d'IA.