

Étude sur les outils pour améliorer la production de données géospatiales

Fabien Amarger, Anaïs Cipel, Nathalie Noblecourt, Jehanne Portefaix et Pierre-Marie Brunet

29/06/2026 – PFIA APIA



1. Introduction

Produire des données d'entraînement : un défi clé en télédétection



L'observation de la Terre par imagerie satellite est un levier majeur pour répondre aux enjeux climatiques, socio-économiques et d'aménagement. La performance de cette filière dépend de données d'entraînement fiables, issues d'un processus de photo-interprétation (vectorisation et labellisation) exigeant.



Objectif de l'étude

Peut-on améliorer la productivité sans dégrader la qualité?

1

Augmenter la productivité

Accroître la productivité sans dégrader la qualité des annotations

2

Assurer le suivi

Garantir un suivi de production transparent et des statistiques fiables

3

Intégrer l'automatisation

Intégrer de façon pragmatique des modèles d'annotation automatisée

4

S'intégrer dans l'existant

S'insérer dans l'écosystème SIG et documentaire existant (projections, formats, ect...)

Productivité

Qualité





Etat de l'art



3. Etat de l'art : Critères de sélection

Critères de sélection des outils et modèles

C1

Adaptée aux données géospatiales THR :

prise en charge d'images géoréférencées (ex. GeoTIFF) et conservation des métadonnées spatiales.

C2

Gestion des projections (au minimum les références mondiales) :

affichage et gestion explicite des CRS/EPSG et capacité à manipuler des données multi-sources (raster/vecteur) sans perte de cohérence spatiale

C3

Interopérabilité avec des environnement SIG :

import/export de formats géospatiaux usuels (GeoJSON, Shapefile, GeoPackage) et/ou connecteurs vers des outils SIG (QGIS/ArcGIS).

C4

Intégration de données contextuelles :

possibilité d'afficher des couches exogènes (vecteurs, services web WMS/WMTS, MNT/DSM si nécessaire) pour assister la photo-interprétation.

C5

Gestion collaborative de la production :

gestion des rôles, assignation de tâches, mécanismes de revue/commentaires et gestion des retours (rework) avec une traçabilité des décisions.

C6

Traçabilité :

journalisation des actions (qui/quoi/quand), versionnement ou historisation des exports, et capacité à relier une annotation à son contexte de production.

C7

Contraintes de déploiement et de licence :

compatibilité avec les exigences de sécurité (on-premise et/ou cloud selon le contexte), transparence du modèle de licence et maîtrise des coûts de possession (TCO).



3. Etat de l'art : Sélection des outils/modèles

Afin de **limiter la subjectivité**, les preuves utilisées pour qualifier les statuts (C/P/NC/ND) proviennent principalement de la **documentation officielle** des outils/modèles, des **dépôts open source** et des **retours techniques** disponibles publiquement.

Chaque critère a été évalué selon la grille suivante :

- C = conforme
- P = partiel (possible avec adaptation / connecteur / prétraitement / en option)
- NC = non conforme
- ND = non déterminé

Outil	C1	C2	C3	C4	C5	C6	C7	Retenu ?
BRIMA	NC	ND	ND	ND	ND	ND	ND	Non
CVAT + SAM	P	P	P	NC	P	ND	C	Oui
Dataloop	ND	ND	ND	ND	ND	ND	P	Non
FlyPix AI	NC	NC	ND	ND	ND	ND	ND	Non
Geo4i	C	C	C	ND	ND	C	ND	Non
GeoAI (Esri)	C	C	C	C	P	P	P	Non
IMERIT Annotator	ND	ND	ND	ND	P	ND	NC	Non
KalDA	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Non
Kili Technology	P	ND	P	ND	C	P	P	Oui
Label Studio	P	NC	P	NC	P	ND	C	Non
Labelbox	P	ND	P	ND	C	ND	P	Non
LabelImg	NC	NC	NC	NC	NC	NC	C	Non
LabelMe	NC	NC	NC	NC	NC	NC	C	Non
Label-Studio	NC	NC	P	NC	C	P	C	Non
Makesense	NC	NC	NC	NC	NC	NC	P	Non
Mapflow.ai	P	P	P	NC	NC	NC	P	Oui
Pixano	NC	NC	NC	NC	ND	ND	P	Non
RectLabel	NC	NC	NC	NC	NC	NC	P	Non
Roboflow	NC	NC	ND	ND	ND	ND	ND	Non
Snail	P	NC	NC	NC	NC	NC	C	Non
SuperAnnotate	P	NC	P	ND	ND	ND	P	Non
Supervisely	P	ND	P	ND	C	C	P	Oui
TagLab	NC	NC	ND	ND	ND	ND	ND	Non
VIA	NC	NC	NC	NC	NC	NC	P	Non
VoTT	NC	NC	NC	NC	NC	NC	P	Non





Expérimentations



4. Indicateurs et métriques

Productivité

- Temps d'annotation / objet / classe
- Taux de production / heure
- Temps des reprises / heure

Qualité

- Précision/Rappel des labels
- F1-score / classe
- Constance topologique

Suivi & gestion

- Évaluation utilisateur (questionnaire)
- Interopérabilité
- Ergonomie
- Gestion de projet



4. Méthodologie : Types de scènes

Le choix des différents types de scènes répond directement à l'expérience acquise lors des projets antérieurs (**AI4GEO**, **PNEO Cities**, **FROZEN**), où la diversité des territoires a montré une forte influence sur la productivité et la qualité des annotations.

Urbain dense

Forte complexité, densité d'objets



Paris

Péri-urbain

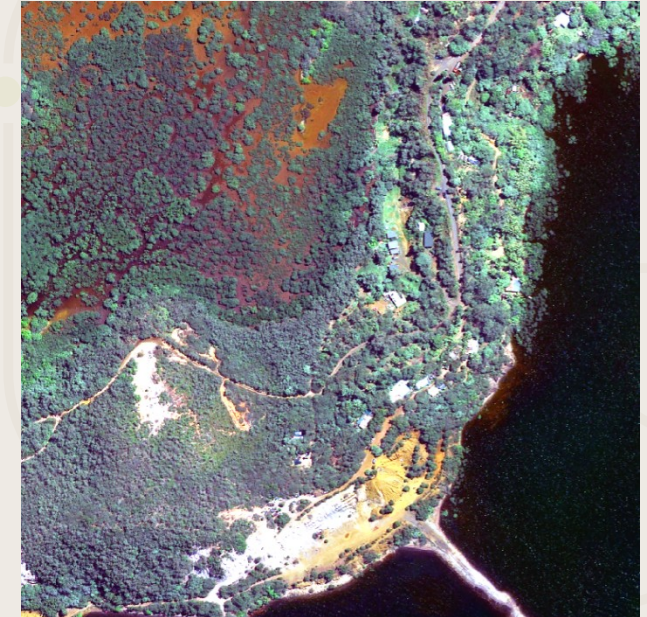
Mix habitats / agriculture.



Nice

Rural

Faible densité mais variabilité texturale.



Nouméa



4. Méthodologie : Protocole d'annotation

Le protocole d'annotation repose sur quatre approches complémentaires : Un schéma de classe unifié, une baseline manuelle, la comparaison des outils/modèles ainsi qu'une étude et comparaison des métriques

Standardisation

Schéma de classes unifié,
règles géométriques,
conventions.



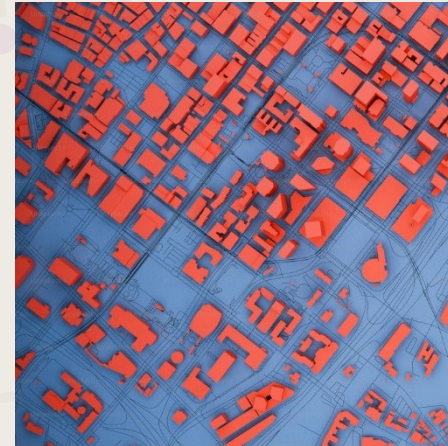
Baseline (manuelle)

annotation standardisée via
QGIS → comparaison de
référence.



Comparaison hybride des outils/modèles

pré-annotation automatisée +
post-édition humaine via QGIS
+ contrôle qualité.



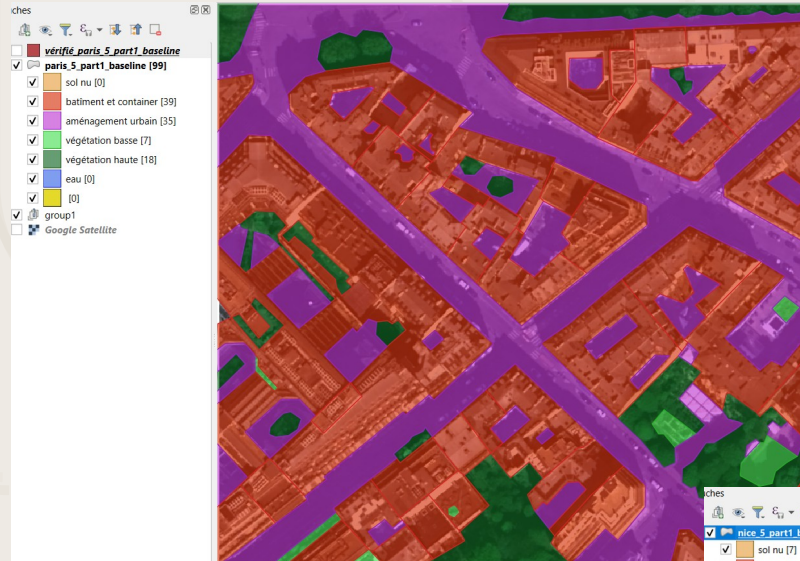
Etudes des métriques

Comparaison, étude et
analyses des métriques

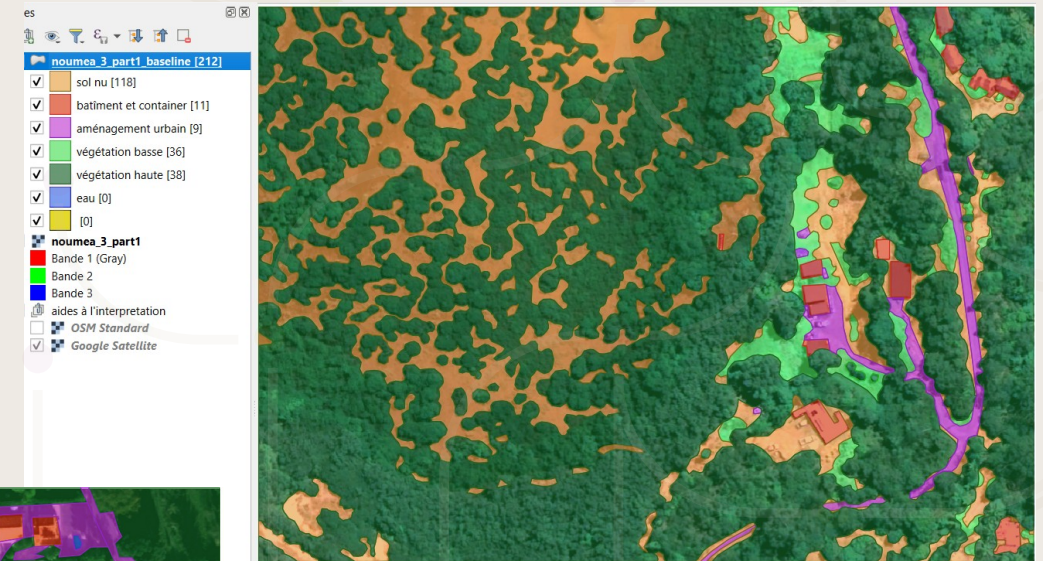


4. Baseline : Résultats

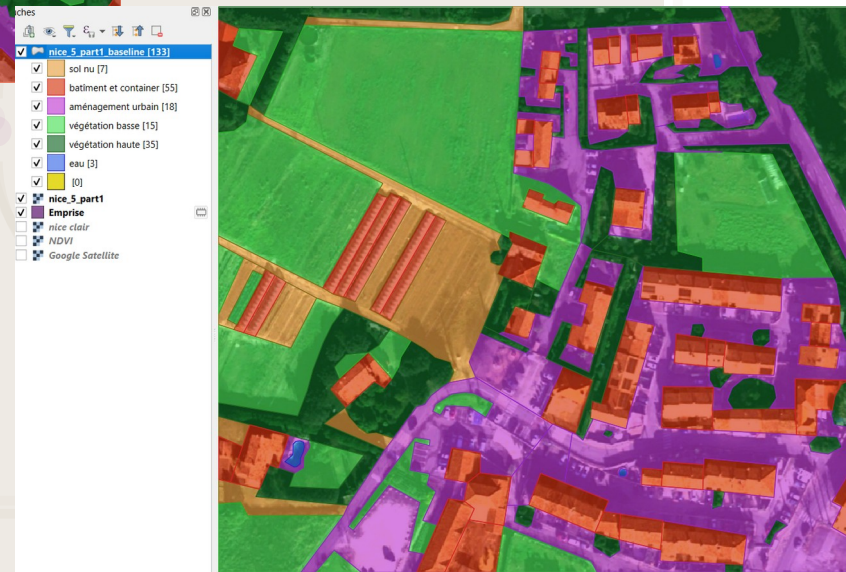
Urbain dense



Rural



Péri-urbain



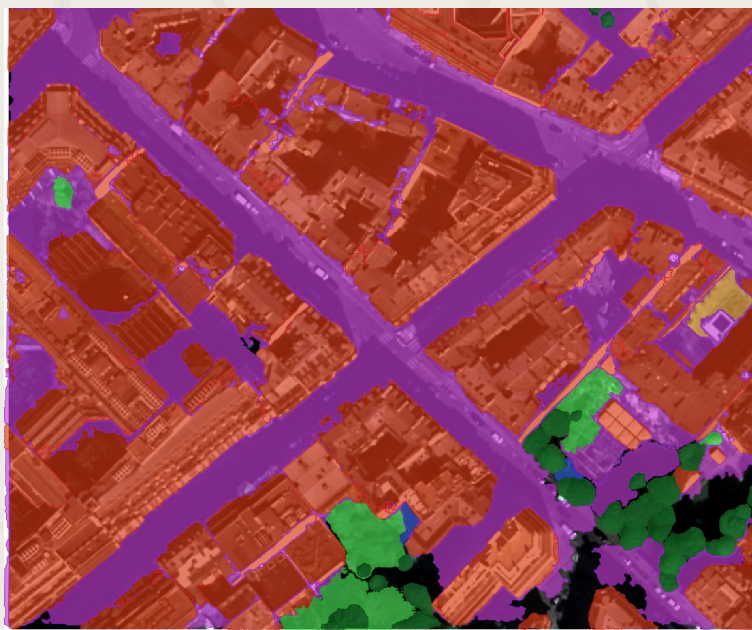
4. Méthode de calcul du temps de production



Baseline

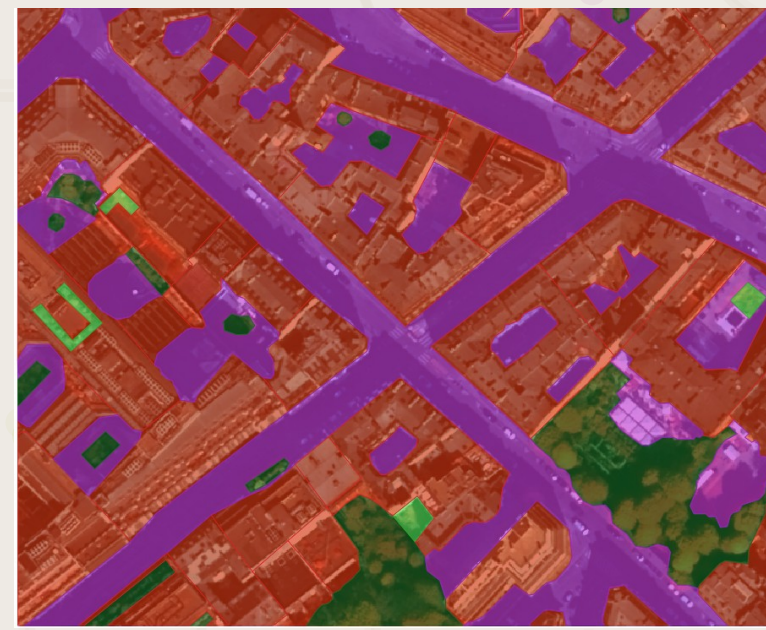
- Prod : 4h20
- CC : 3h

Total Baseline =
7h20



**RAW (Kili
technology)**

- Semi-automatisation: 1h38



Post-édition

- Correction erreurs : 2h25
- Post-édition : 2h15

Total Outil =
6h18

Temps total = RAW + Corrections erreurs + Post-édition / Temps Baseline



4. Indicateurs et métriques de comparaison

Qualité

La qualité est évaluée en comparant les annotations produites à la baseline, à partir de leur intersection géométrique.

Les métriques de **précision**, **rappel** et **F1-score** permettent de quantifier la qualité finale des annotations.

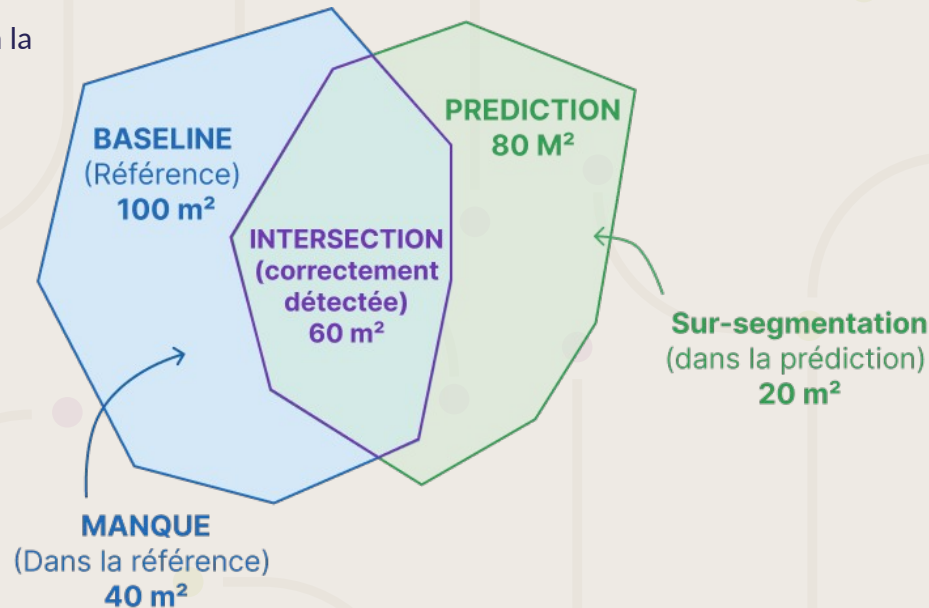
Calcul basé sur :

- Intersection géométrique (m²) entre baseline et résultat
- Géotraitements en model builder sur QGIS
- Calcul par classe

$$rappel = S_{inter} / S_{baseline}$$

$$precision = S_{inter} / S_{resultat}$$

$$F1 = 2 \times \frac{(precision \times rappel)}{(precision + rappel)}$$



RAPPEL

Surface correctement retrouvée

$$\frac{\text{Surface de référence totale}}{\text{Surface correctement retrouvée}} = \frac{60 \text{ m}^2}{100 \text{ m}^2} = 0,60 \rightarrow 60 \%$$



PRECISION

Surface correcte

$$\frac{\text{Surface prédite totale}}{\text{Surface correcte}} = \frac{60 \text{ m}^2}{80 \text{ m}^2} = 0,75 \rightarrow 75 \%$$



RAPPEL

Mesure la capacité à retrouver toute la géométrie de référence
Ici : 40 m² manquants → rappel de 60 %



PRECISION

Mesure de la part de la prédiction qui est correcte
Ici : 20 m² en trop → précision de 75 %





Résultat des expérimentations

Outils



5. Synthèse

Les performances des outils d'annotation automatisée dépendent fortement du contexte territorial et des caractéristiques des données.

- En **milieu urbain dense**, les sorties brutes sont fortement dégradées et nécessitent une correction importante, limitant les gains de productivité.
- En **contexte péri-urbain**, les résultats sont hétérogènes et les gains restent incertains.
- En **milieu rural**, les performances sont plus stables bien que certaines annotations soient incomplètes, mais les gains de productivité restent variables selon les classes et les outils.

	Comparaison baseline/RAW			Comparaison baseline/Post-édition			Temps Baseline			Temps Hybride		
	F1-Score Urbain	F1-Score Peri-Urbain	F1-Score Rural	F1-Score Urbain	F1-Score Peri-Urbain	F1-Score Rural	Urbain	Peri-Urbain	Rural	Urbain	Peri-Urbain	Rural
Kili Technology	0,53	0,64	0,59	0,85	0,83	0,73	7,33	6,00	4,66	6,30	16,03	12,57
CVAT+SAM	0,56	0,61	0,53	0,79	0,82	0,78				4,62	5,52	4,97
Supervisely	0,79	0,75	0,73	0,87	0,76	0,84				4,25	10,27	5,33
Mapflow.ai	0,63	0,28	0,50	0,79	0,70	0,70				3,89	10,75	4,85

Aucun outil ne se distingue de manière systématique : les performances dépendent principalement de l'interaction entre les outils, les données et le flux de production.

Le temps de correction domine le temps total → conditionne la performance du flux hybride



5. Résultats expérimentaux - Outils

Critères de gestion de projet et ergonomie

	Interopérabilité				Ergonomie				Gestion de projet		
	Formats des données	Projections	Compatibilité	Intégration de modèles	Prise en main	Interface	Facilité d'utilisation	Numérisation avancés	Suivi de l'avancement	Collaboration	Indicateurs
Kili Technology	4	3	2	1	4	2	3	3	4	5	3
CVAT+SAM	0	0	0	3	3	4	3	0	5	4	0
Supervisely	2,5	2	3	4	4	5	4	2	4	4	2
Mapflow.ai	3	2	4	0	3	3	4	5	5	0	0





Conclusions & Perspectives



6. Synthèse globale

L'automatisation n'est pas un gain en soi : tout dépend du coût de correction et du niveau de qualité attendu.

Rôle des outils vs modèles

- Les modèles seuls produisent des annotations incomplètes
- Les outils structurent le flux mais restent dépendants des modèles

Qualité & Productivité

- Les sorties brutes ne sont pas directement exploitables
- Le temps de correction est le facteur déterminant de la performance

Dépendance au contexte

- Contextes simples / classes structurées → **cas favorables**
- Contextes complexes / classes diffuses → **corrections lourdes**

Les outils et modèles ne permettent pas d'automatiser la production, mais peuvent apporter un gain dans certains cas, à condition que le coût de correction reste limité.



6. Perspectives

Limites de l'étude

- Augmenter le nombre d'images et de territoires
- Multiplier les annotateurs (réduction du biais)
- Comparaison inter-annotateurs
- Tester plus de classes / cas d'usage
- Évaluer sur des projets réels à plus grande échelle

Perspectives techniques

1. Concevoir un outil complémentaire à QGIS

- Plateforme connectée à QGIS (plugin / synchronisation)
- Ne pas remplacer QGIS → s'appuyer dessus

2. Structurer un workflow hybride

- Modèles → pré-annotation
- Outils → gestion / collaboration
- QGIS → qualité finale
- Intégrer plusieurs modèles interchangeables
- Découpler outils ↔ modèles

4. Amélioration des modèles

- Progrès rapides en vision par ordinateur
- Potentiel d'amélioration à moyen terme





Merci :)

CNES

pierre-marie.brunet@cnes.fr

olivier.queyrut@cnes.fr

DIGITANIE

jehanne.portefaix@digitanie.org

fabien.amarger@digitanie.org

nathalie.noblecourt@digitanie.org

anais.cipel@digitanie.org

5. Résultats expérimentaux - Outils

Territoire urbain dense

Baseline / RAW

	Bare ground			Building			Hardscape			Tall Vegetation			Low Vegetation			Water			Moyenne
	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1	
Kili Technology	0,00	0,00	0,00	0,90	0,85	0,87	0,76	0,75	0,75	0,35	0,96	0,51	0,03	0,06	0,04	1,00	1,00	1,00	0,53
CVAT+SAM	0,00	0,00	0,00	0,87	0,91	0,89	0,76	0,80	0,78	0,66	0,73	0,69	0,00	0,00	0,00	1,00	1,00	1,00	0,56
Supervisely	1,00	1,00	1,00	0,91	0,91	0,91	0,76	0,85	0,80	0,70	0,85	0,77	0,16	0,70	0,26	1,00	1,00	1,00	0,79
Mapflow.ai	1,00	1,00	1,00	0,89	0,93	0,91	0,09	0,96	0,16	0,69	0,78	0,73	0,00	0,00	0,00	1,00	1,00	1,00	0,63

Baseline / Post-édition

	Bare ground			Building			Hardscape			Tall Vegetation			Low Vegetation			Water			Moyenne
	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1	
Kili Technology	1,00	1,00	1,00	0,94	0,94	0,94	0,92	0,83	0,87	0,85	0,99	0,91	0,23	0,92	0,37	1,00	1,00	1,00	0,85
CVAT+SAM	1,00	1,00	1,00	0,92	0,94	0,93	0,90	0,80	0,85	0,80	0,82	0,81	0,09	0,64	0,16	1,00	1,00	1,00	0,79
Supervisely	1,00	1,00	1,00	0,93	0,95	0,94	0,92	0,87	0,89	0,91	0,85	0,88	0,34	0,90	0,49	1,00	1,00	1,00	0,87
Mapflow.ai	1,00	1,00	1,00	0,97	0,94	0,95	0,94	0,96	0,95	0,92	0,80	0,86	0,00	0,00	0,00	1,00	1,00	1,00	0,79

Temps

		nb erreurs geom	nb erreurs topo	Temps traitement	Temps post-édition		Temps total
					Tps corrections erreurs	Tps corrections seg/lab	
BASELINE		0	0	04:20		03:00	07:20
OUTILS	Kili Technol	0	130	01:38	02:25	02:15	06:18
	CVAT+SAM	0	130	00:12	02:15	02:10	04:37
	Supervisely	0	0	01:40	00:00	02:35	04:15
	Mapflow.ai	41	123	00:03	02:00	01:50	03:53

Synthèse - territoire urbain dense : Dans ce contexte, la complexité géométrique dégrade fortement la qualité des sorties brutes, générant de nombreuses erreurs de segmentation et de classification.

Le temps de correction devient alors dominant, annulant dans la plupart des cas les gains de productivité du flux hybride.



5. Résultats expérimentaux - Outils

Territoire péri-urbain

Baseline / RAW

	Bare ground			Building			Hardscape			Tall Vegetation			Low Vegetation			Water			Moyenne
	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1	
Kili Technology	0,79	0,45	0,57	0,78	0,85	0,81	0,62	0,81	0,70	0,41	0,84	0,55	0,46	0,82	0,59	0,57	0,64	0,60	0,64
CVAT+SAM	0,41	0,21	0,28	0,83	0,92	0,87	0,52	0,75	0,61	0,55	0,85	0,67	0,54	0,94	0,69	0,83	0,39	0,53	0,61
Supervisely	0,75	0,43	0,55	0,85	0,85	0,85	0,71	0,80	0,75	0,71	0,75	0,73	0,84	0,88	0,86	0,68	0,80	0,74	0,75
Mapflow.ai	0,00	0,00	0,00	0,80	0,93	0,86	0,11	0,63	0,19	0,59	0,66	0,62	0,00	0,00	0,00	0,00	0,00	0,00	0,28

Baseline / Post-édition

	Bare ground			Building			Hardscape			Tall Vegetation			Low Vegetation			Water			Moyenne
	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1	
Kili Technology	0,97	0,78	0,86	0,87	0,88	0,87	0,87	0,82	0,84	0,61	0,90	0,73	0,92	0,86	0,89	0,91	0,66	0,77	0,83
CVAT+SAM	0,85	0,73	0,79	0,95	0,93	0,94	0,92	0,84	0,88	0,83	0,87	0,85	0,85	0,97	0,91	0,83	0,44	0,58	0,82
Supervisely	0,75	0,45	0,56	0,85	0,86	0,85	0,71	0,81	0,76	0,74	0,79	0,76	0,90	0,90	0,90	0,68	0,80	0,74	0,76
Mapflow.ai	0,85	0,82	0,83	0,81	0,96	0,88	0,95	0,76	0,84	0,92	0,78	0,84	0,00	0,00	0,00	0,99	0,69	0,81	0,70

Temps

		nb erreurs geom	nb erreurs topo	Temps traitement	Temps post-édition		Temps total
					Tps corrections erreurs	Tps corrections seg/lab	
BASELINE		0	0	04:30		01:30	06:00
OUTILS	Kili Technolo	205	727	02:12	10:00	03:50	16:02
	CVAT+SAM	0	110	00:26	02:55	02:10	05:31
	Supervisely	0	1	02:00	00:01	08:15	10:16
	Mapflow.ai	47	119	01:45	02:30	06:30	10:45

Synthèse - territoire péri-urbain : Dans ce contexte, la qualité des sorties brutes est améliorée mais reste hétérogène, notamment dans les zones de transition où les confusions entre classes sont fréquentes.

Les gains de productivité sont variables et dépendent du volume de corrections, qui peut compenser voire annuler le bénéfice de l'automatisation.



5. Résultats expérimentaux - Outils

Territoire rural

Baseline / RAW

	Bare ground			Building			Hardscape			Tall Vegetation			Low Vegetation			Water			Moyenne
	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1	
Kili Technology	0,11	0,60	0,19	0,82	0,81	0,81	0,39	0,75	0,51	0,86	0,87	0,86	0,19	0,19	0,19	1,00	1,00	1,00	0,59
CVAT+SAM	0,10	0,78	0,18	0,71	0,92	0,80	0,45	0,39	0,42	0,65	0,86	0,74	0,06	0,01	0,02	1,00	1,00	1,00	0,53
Supervisely	0,66	0,71	0,68	0,80	0,89	0,84	0,42	0,67	0,52	0,84	0,91	0,87	0,35	0,67	0,46	1,00	1,00	1,00	0,73
Mapflow.ai	0,00	0,00	0,00	0,73	0,92	0,81	0,25	0,54	0,34	0,95	0,78	0,86	0,00	0,00	0,00	1,00	1,00	1,00	0,50

Baseline / Post-édition

	Bare ground			Building			Hardscape			Tall Vegetation			Low Vegetation			Water			Moyenne
	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1	
Kili Technology	0,64	0,70	0,67	0,86	0,85	0,85	0,43	0,78	0,55	0,92	0,89	0,90	0,40	0,38	0,39	1,00	1,00	1,00	0,73
CVAT+SAM	0,50	0,86	0,63	0,96	0,94	0,95	0,72	0,45	0,55	0,97	0,90	0,93	0,60	0,61	0,60	1,00	1,00	1,00	0,78
Supervisely	0,72	0,77	0,74	0,87	0,90	0,88	0,73	0,78	0,75	0,94	0,91	0,92	0,60	0,89	0,72	1,00	1,00	1,00	0,84
Mapflow.ai	0,74	0,80	0,77	0,96	1,00	0,98	0,50	0,65	0,57	0,99	0,81	0,89	0,00	0,00	0,00	1,00	1,00	1,00	0,70

Temps

		nb erreurs geom	nb erreurs topo	Temps traitement	Temps post-édition		Temps total
					Tps corrections erreurs	Tps corrections seg/lab	
BASELINE		0	0	04:00		00:40	04:40
OUTILS	Kili Technology	5	373	03:24	06:50	02:20	12:34
	CVAT+SAM	0	68	00:13	01:10	03:35	04:58
	Supervisely	0	0	04:00	00:00	01:20	05:20
	Mapflow.ai	5	235	00:01	02:05	02:45	04:51

Synthèse - territoire rural : Dans ce contexte, la qualité des sorties brutes est globalement meilleure, mais reste incomplète et variable selon les classes.

Les gains de productivité ne sont pas systématiques et dépendent du volume de corrections, notamment lorsque certaines classes sont mal détectées ou absentes.



1. Déroulement de l'étude

