Mise à jour de l’attribut occupation des sols de la base de données cadastrales

LIEGEON Théo [1] , MERCIER Solane [1], SERY Ange [1], UOZUMI Misaki [1]

1. Master II Observation de la Terre et Géomatique – 3 Rue de l'Argonne, 67000 Strasbourg
solane.mercier@etu.unistra.fr, misaki.uozumi@etu.unistra.fr, ange.sery@etu.unistra.fr, theo.liegeon@etu.unistra.fr

RESUME

Bien que le système cadastral français constitue un maillage territorial exhaustif, son attribut d'Occupation des Sols (OCS) pâtit d'une actualisation déclarative, induisant un décalage sémantique avec la réalité physique. Pour pallier cette lacune, cette étude déploie une architecture automatisée de fusion multi-sources — couplant données satellitaires (OSO, COSIA) et référentiels vectoriels experts (RPG, BD TOPO, BD FORÊT) — visant à fiabiliser l'OCS à l'échelle parcellaire. Deux approches méthodologiques sont confrontées : une fusion décisionnelle par scoring (Méthode 1) et un classifieur hiérarchique (Méthode 2). Les résultats objectivent la supériorité de la première méthode pour caractériser la couverture du sol (Land Cover), grâce à sa robustesse face aux biais morphométriques, tandis que la seconde privilégie une lecture fonctionnelle (Land Use). En exploitant l'asymétrie fonctionnelle des référentiels et en érigeant l'incertitude comme une métrique de qualification explicite, ce protocole pose les premier pas d'une actualisation cadastrale robuste et reproductible.

MOTS-CLÉS : Cadastre, OCS, Multi-source

1. Introduction

Créé au XIX[e] siècle, le système cadastral français constitue un référentiel spatial et administratif exhaustif à l’échelle nationale, comptabilisant près de 106 millions d'unités foncières en 2019 (Bonnefond M. et Desrousseaux M., 2021). Originellement conçu comme un instrument fiscal et foncier, le Plan Cadastral Informatisé (PCI) connaît, depuis son ouverture en *open data* en 2017, une diversification majeure de ses usages. Sa mobilisation s'avère désormais incontournable pour le déploiement des politiques publiques, particulièrement via son croisement avec d'autres référentiels géospatiaux (Bonnefond M. et Desrousseaux M., 2021). Dans cette perspective, la variable « Occupation du Sol » (OCS) constitue un descripteur critique pour l'observation et la modélisation des dynamiques territoriales, alimentant les dispositifs d'évaluation transversaux aux enjeux écologiques, urbanistiques et de prévention des risques.

Toutefois, la diversification et la hausse des usages de la données cadastrale par les acteurs publics et privés exacerbe ses vulnérabilités structurelles, tenant tant aux contraintes juridiques qu'à sa précision sémantique et sa fréquence d'actualisation. La fiabilité de l’attribut OCS est donc aujourd'hui jugée insuffisante (Cerema, s.d., consulté en 2026). Administrée par la Direction Générale des Finances Publiques (DGFiP), sa mise à jour repose principalement sur un système déclaratif propriétaire. Cette inertie procédurale qui s’avère souvent long et complexe, provoque un décalage temporel par rapport à la réalité du terrain (Chauvin M-C., 2020).

Dans ce contexte, il est intéressant de tirer pleinement profit des avancées dans le domaine de la géomatique et de la télédétection pour la mise à jour de l'attribut OCS du cadastre, comme Kasser M. (2008) l'a mentionné dans son étude. L'essor récent de données spatiales en *open source* offre un changement d'échelle inédit en termes de résolution spatiale, de couverture territoriale et de récurrence temporelle. Ce tremplin d'informations intègre tant les produits de télédétection (Corine Land Cover, OSO, COSIA) que des données thématiques considérés comme des références nationales (LiDAR HD, BD TOPO, RPG).

De nombreux travaux démontrent que l'exploitation conjointe de plusieurs sources constitue un levier majeur pour améliorer la qualité et la fiabilité des données d'OCS : c'est l'approche multi-source. Cette méthode permet de produire des cartes intégrées plus précises et plus cohérentes à différentes échelles (Batista F. e S. et al., 2013). Par ailleurs, la prise en compte de la dimension multi-temporelle apparaît comme un enjeu central permettant non seulement d'améliorer la détection des changements, mais aussi d'évaluer la stabilité et la cohérence temporelle des classifications produites (Verburg P. et al., 2011). Une part importante de cette littérature s'appuie sur des données raster, notamment satellitaires. Cependant, plusieurs études proposent des méthodologies de fusion de couches LULC (Land use and land cover) combinant des données raster avec des données thématiques vectorielles (Kalenga Tshingomba U. et al., 2020 ; Thierry C. et al., 2020). En outre, les études menées par le Cerema (2018), notamment dans le cadre de l'OCSGE, proposent des méthodologies reposant sur le croisement de données vectorielles (fichiers fonciers, BD TOPO®, RPG, etc.) à l'aide d'arbres de décision et de règles expertes, afin d'enrichir ou de fiabiliser l'information d'usage du sol à l'échelle parcellaire. Toutefois, ces approches nécessitent de mettre en cohérence des différences entre sources en termes de choix méthodologiques, de précision, de temporalité et de nomenclatures (Verburg P. et al., 2011).

Pour formaliser la fusion de jeux de données, plusieurs cadres théoriques sont mobilisés. La théorie de Dempster-Shafer par exemple, est utilisée pour gérer l'incertitude et les conflits entre sources, en attribuant des degrés de croyance aux différentes classes issues de jeux de données distincts (Liu L. et al., 2021). De même, les ensembles flous offrent un cadre pertinent pour représenter la gradualité et l'ambiguïté inhérentes à certaines catégories d'usage du sol, notamment dans les zones de transition ou à forte dynamique (Verburg P. et al., 2011). Les approches multi-sources ou multi-temporelles sont également associées à des scores de confiance ou d'incertitude, permettant de quantifier la qualité de l'information produite (Liu L. et al., 2021 ; Pérez-Hoyos A. et al., 2012 ; Verburg P. et al., 2011).

De ce fait, cet article répondra à la problématique suivante : ***Dans quelle mesure le croisement de sources hétérogènes d'OCS avec des données exogènes de référence nationale peut-il fiabiliser l'attribut « OCS » du cadastre, et selon quelles marges d'incertitude ?*** L'hypothèse principale repose sur l'intégration de la dimension temporelle et la combinaison des données multi-sources, permettant de dépasser les limites des approches mono-source et d'améliorer la cohérence et la robustesse de la qualification de l'OCS. Dans cette perspective, cet article a pour premier objectif de concevoir une méthodologie attribuant une classe unique d'OCS et un indice de confiance à la parcelle cadastrale. La prise en compte de l'automatisation et de l'interopérabilité vise une meilleure reproductibilité sur l'ensemble du territoire français. Il s'agit également de rendre les résultats obtenus visibles et interprétables sous forme cartographique.

Si les travaux cités constituent une référence opérationnelle majeure pour les démarches d'actualisation de l'OCS à partir de données existantes, il n'existe à notre connaissance aucune documentation visant à produire une information exhaustive concernant l'OCS à l'échelle parcellaire, ni de méthodologie permettant d'automatiser le processus. C'est la raison pour laquelle la méthodologie que nous proposons dans cet article est exploratoire. Après avoir présenté les choix de la zone d'étude et des données, nous exposerons la méthodologie construite. Par la suite, nous présenterons les résultats obtenus. Enfin, une discussion critique mettra en lumière les interprétations des résultats, les apports, les limites et les perspectives.

2. Présentation du site d'étude et des données

2.1 Le site d'étude

Le choix du site d'étude s'inscrit dans une démarche méthodologique reproductible visant à mettre à jour de manière automatique les informations sur l'ensemble du territoire français. Dans le cadre d'un projet commandé par le Ministère de l'Agriculture et la DGFIP, le département de l'Aude a été retenu comme territoire d'expérimentation. Situé en région Occitanie, dans le sud de la France, le

département de l'Aude couvre une superficie de 6 139 km² et compte 433 communes. Le territoire est dominé par des espaces ruraux et globalement peu urbanisés (Ministères de la Transition Écologique et de la Cohésion des Territoires 2018). L'occupation du sol (source : Corine Land Cover en 2018) met en évidence une prédominance des forêts (48 %) dans le sud du département et des terres agricoles (47 %), que l'on retrouve davantage au nord-ouest, au centre et au nord-est du territoire, le long de l'Aude (Annexe 1). Le littoral est, de son côté, caractérisé par la présence de nombreux étangs et lagunes (2 %). Les principales zones urbanisées (3 %) se situent en plaine, dans le sillon audois, autour des principales villes du département (Castelnaudary, Carcassonne et Narbonne) et sur l'ensemble du littoral (Le Roux B. et Albouy S., 2024).

Considérant le nombre important de parcelles ainsi que l'importance des données mobilisées, il est essentiel de tester la méthodologie sur des échantillons dans un premier temps. L'analyse repose donc sur cinq échantillons (dalles) choisis depuis une grille spatiale de résolution 2 × 2 km, s'appuyant sur le cadre méthodologique de LUCAS (Land Use/Cover Area frame Survey) (D'Andrimont R., et al., 2020). Cette approche produit une carte de l'OCS plus fiable sur l'ensemble du territoire européen. Elle se base sur ce type de grille, dont la résolution est pertinente pour observer les parcelles cadastrales, y compris celles de petite taille.

Compte tenu de l'hétérogénéité et de la distribution de l'OCS sur l'ensemble du territoire, le choix s'est porté sur cinq dalles stratégiques, sélectionnées selon deux grands critères complémentaires (Figure 1) :

1. La diversité des classes OCS (pour OSO) présentes dans les parcelles pour une dalle, couvrant le type agricole, artificialisé, boisé, forestier et aquatique.
2. Les caractéristiques géométriques des parcelles telles que la superficie, le périmètre et la circularité (Piombini A. et al., 2013), distinguant des formes simples ou plus complexes.

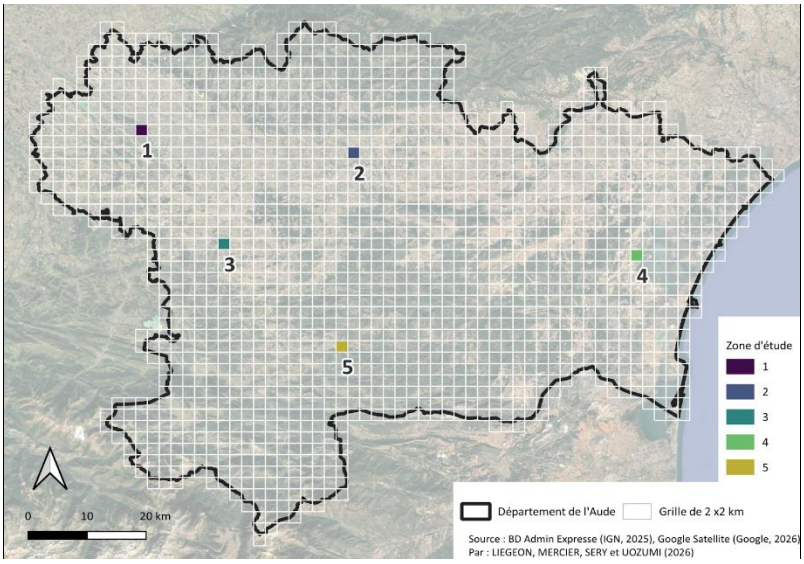


Figure 1 : Cinq dalles échantillons choisis dans le département de l'Aude.

Ces dalles ont été sélectionnées afin de capturer l'hétérogénéité des configurations morpho-spatiales et des gradients thématiques inhérents dans le département. L'Annexe 2 détaille les métriques quantitatives complémentaires, relatives aux propriétés topologiques et aux signatures sémantiques caractérisant chaque unité d'échantillonnage.

Cette approche est essentielle pour construire une méthodologie reproductible sur un grand territoire présentant des caractéristiques variées. Elle permet également de limiter les contraintes liées au volume de données et au temps de calcul.

2.2 Les données disponibles

La mise à jour de l'OCS sur le département de l'Aude nécessite l'exploitation de différentes sources de données complémentaires, adaptées à la diversité des contextes paysagers observés.

Le référentiel BD ADMIN circonscrit les délimitations administratives et constitue le socle géométrique servant à la génération de la grille de partitionnement spatial sur l'ensemble de l'emprise départementale. En complément, le produit Parcellaire Express (PCI) de l'IGN est mobilisé pour extraire la topologie fine et la structuration morphologique des unités foncières cadastrales. L'implémentation de la chaîne de traitement repose sur l'intégration de six référentiels spatiaux majeurs. La modélisation globale de l'Occupation des Sols (OCS) est assurée par l'exploitation conjointe des bases OSO (millésimes 2018-2023) et COSIA (2018, 2021, 2024). Ce socle généraliste est abondé par des bases de données vectorielles thématiques expertes : le Registre Parcellaire Graphique standard (RPG, 2018-2024) et sa version consolidée (RPG complété, 2018-2022), la BD TOPO (2025), ainsi que la BD FORÊT v2.0 (2017). L'inventaire exhaustif de ce corpus de données détaillé au sein du tableau récapitulative (Figure 2 - lignes bleues).

La mobilisation de ces référentiels spatiaux s'inscrit dans un continuum méthodologique largement validé par la littérature scientifique. En effet, la caractérisation de l'Occupation des Sols (OCS) par des processus de fusion multi-sources s'appuie classiquement sur l'exploitation conjointe de ces bases de données : les travaux de Kalenga Tshingomba U. et al. (2020) capitalisent notamment sur l'utilisation en synergie des données OSO, RPG et BD FORÊT, tandis que l'approche développée par Thierry C. et al. (2020) intègrent en plus la BD TOPO afin de consolider le diagnostic spatial.

En s'appuyant sur la fiche transversale concernant les données de l'OCS par la Commissariat général au développement durable (CGDD) et Cerema (s.d.) ainsi que sur la méthodologie pour la cartographie de l'OCS par Thierry C. et al. (2020), trois grands axes permettent de déterminer les données pertinentes afin de répondre aux objectifs et hypothèses suivants :

- Disponibilité : open source, et présence à l'échelle de la France métropolitaine (1) ;
- Interopérabilité : format, attributs, nomenclature, intégrable sous l'outil libre (2) ;
- Performance des données : précision, taille, mode de production, fréquence d'actualisation (3).

L'exploitabilité opérationnelle de chaque source de données fait l'objet d'une qualification rigoureuse au sein du tableau d'évaluation (Figure 2). Cet audit repose sur un ensemble de métriques : résolution spatiale, richesse sémantique des attributs, granularité temporelle (millésimes), typologie de format, précision topologique et thématique, ainsi que sur l'intégrité des métadonnées de fiabilité et des méthodes de production. Bien que les autres sources (Figure 2 - lignes blanches) aient été consultées et recensées, elles n'ont pas été retenues. Certaines sources sont disponibles sous deux formats : vecteur et raster. En raison de la taille de données et de la cohérence avec les autres sources, le mode vecteur est privilégié pour cette étude. Ce choix s'inscrit également dans les travaux de Thierry C. et al. (2020), qui ont cartographié l'OCS par la combinaison de nombreuses données vectorielles car l'utilisation seule de photographies aériennes étant trop chronophage. De plus, ces sources ne reflètent pas parfaitement la réalité du terrain. Des marges d'erreur sont documentées pour chaque classe thématique ou pour chaque entité.

OSO fournit une cartographie annuelle, élaborée à partir d'une série temporelle (traité par IA) d'images Sentinel-2 depuis 2017. Il est avantageux pour son homogénéité sur l'ensemble du territoire français et offre une référence mobilisable en 24 classes thématiques. Cependant, il présente parfois des difficultés pour la distinction des classes entre elles ou pour la stabilité temporelle au niveau local.

Pour compléter certaines limites, COSIA (Couverture du Sol par IA) permet de prédire l'occupation du sol selon 15 classes. La production de ce référentiel est automatisée via des algorithmes d'intelligence artificielle, mobilisant spécifiquement des architectures d'apprentissage profond (*deep learning*). Ce

jeu de données bénéficie d'une mise à jour tous les trois ans, offrant une profondeur temporelle structurée autour de trois millésimes (2018, 2021 et 2024) sur la zone d'étude. Si cette source cartographique se distingue par une résolution spatiale fine (20 cm), elle comporte néanmoins un déficit de granularité sémantique pour les zones naturelles et agricoles : la nomenclature employée y peine à discriminer finement la complexité des usages fonctionnels (CGDD et Cerema, s.d.).

Le RPG permet l'identification des parcelles agricoles. Vingt-quatre (24) grands groupes de culture et 350 cultures principales y sont représentés, associés aux surfaces déclarées par les éleveurs dans le cadre de la Politique Agricole Commune (PAC) (IGN, 2025b). Le RPG complété, vient compléter les parcelles agricoles absentes du RPG en attribuant une catégorie de culture aux parcelles identifiées (Cantelaube P. et Lardot B., 2022) via des données externes traitées par IA. Elle est retenue comme le référentiel spécialisé pour les surfaces agricoles.

La BD FORÊT version 2 a été élaborée entre 2007 et 2017 par photo-interprétation d'images en infrarouge couleur de la BD ORTHO (IGN, 2025a). Elle est retenue comme le référentiel spécialisé pour la végétation et le milieu naturel.

La BD TOPO est une description des éléments du territoire et de ses infrastructures, de précision métrique. Elle est intéressante pour les thématiques naturelles et anthropiques. Une nouvelle édition est publiée chaque trimestre. Elle est retenue comme le référentiel spécialisé pour l'artificialisation et l'eau.

L'exploitation de ces sources complémentaires offre une base fiable pour la mise à jour automatisée de l'OCS sur le département de l'Aude, tout en tenant compte de la diversité des contextes paysagers.

Catégorie	Donnée	Producteur	Millésimes	Format	SCR	Résolution	Probabilité	Lien
Réglementaire	PLU	Picto-Occitanie	2024	Vecteur(WFS)	L93	-		https://www.picto-occitanie.fr/accueil
Limites (admin)	Admin Express	IGN	2025	Vecteur(.shp)	L93	-		https://geoservices.ign.fr/adminexpress
Limites (cadastre)	Parcellaire Express	IGN & DGFiP	2025	Vecteur(.shp)	L93	-		https://geoservices.ign.fr/parcellaire-express-pci
Relief	RGE ALTI	IGN	2024	Raster(.tif)	L93	5m/1m		https://geoservices.ign.fr/rgealti
OCS	OSO	Theia	2018-2023	Raster(.tif)/Vecteur	L93	10m	Entité	https://www.theia-land.fr
OCS	CoSIA	IGN	2018,2021,2024	Raster(.tif)/Vecteur	L93	20cm	Entité	https://geoservices.ign.fr/cosia
OCS	OSC GE	IGN & Cerema	2018-2021	Vecteur(.shp)	L93	Centimétrique		https://geoservices.ign.fr/ocsge
OCS	LiDAR HD (8+3)	IGN	2021–2026 selon zones	LAS / Raster(.tif)	L93	≥ 10 points/m² ≈ 10 cm (verticale) ≈ 50 cm (horizontale)		https://geoservices.ign.fr/lidarhd
OCS	CORINE Land Cove	Copernicus	1990, 2000, 2006, 2012 et 2018	Raster(.tif)/Vecteur	ETRS89	100m		https://land.copernicus.eu/en/products/corine-land-cover
Agricole	RPG	IGN & Agence de Services et de Paiement	2018-2024	Vecteur(.shp)	L93	Parcellaire agricole		https://geoservices.ign.fr/rpg
Agricole	RPG Complété	INRAE	2018-2022	Vecteur(.shp)	L93	Parcellaire agricole		https://entrepot.recherche.data.gouv.fr/dataverse/rpg_complete_2022
OCS / Bâtiments / Végétation / Hydrographie	BD TOPO	IGN	2025	Vecteur(.shp)	L93	Métrique		https://geoservices.ign.fr/bdtopo
Bâtiments	OSM	OSM	Temps Réel	PBF/Vecteur(sbp)	WGS84	-		https://www.openstreetmap.org/#map=6/46.46/2.23
Végétation	BD FORÊT V2	IGN	2021-2023	Vecteur(.shp)	L93	Métrique		https://geoservices.ign.fr/bdforet
Végétation	BD Haie	IGN	2022	Vecteur(.shp)	L93	25 cm		https://geoservices.ign.fr/bdhaie
Végétation	Inventaire des sites dans le cadre des Espaces Naturels Sensibles de l'Aude	Département de l'Aude	2025	Vecteur(.shp)	L93	-		https://opendata.aude.fr/explore/dataset/cd11_sites-des-espaces-naturels-sensibles/analyze/?location=8,43.05682,2.49884&basemap=jawg.streets&dataChart=eyJxIjoiVVgaWVdsIpbeyIjaGFydHMiOlt7InR5cGUiOiJibmFyIiwiY29uZmlnIjp7InNlcmllcyI6WyJjb3VudCJdLCJvcmllbnRhdGlvbiI6InZlcnRpY2FsIiwiZGF0YXNldCI6MH19XSwieEF4aXMiOlsiZiI6bnVsbH0sInlBeGlzIjp7InNob3dHaWRsaW5lcyI6dHJ1ZX19&sw=E6Zv6eyJkYXRhc2V0IjoiY2QxMV9zaXRlcy1kZXMtZXNwYWNlcy1uYXR1cmVscy1zZW5zaWJsZXMiLCJvcHRpb25zIjp7fX19XSwidGltZXNjYWxlIjoiIiwiZ6IzcGxheUxlZ2VuZCI6dHJ1ZSwiYWxpZ25Nb250aCI6dHJ1ZX0%3D	
Végétation	Natura 2000	INPN/DREAL	2024	Vecteur(.shp)	L93	Métrique		https://www.natura2000.fr/
Hydrogrphie	BD Topage	OFB / IGN / SANDRE	2022	Vecteur(.shp)	L93	Métrique		https://www.eaufrance.fr/actualites/le-millesime-2022-de-la-bd-topager-metropole-est-disponible

Figure 2 : Inventaires des données (en bleu : utilisées, en blanc : consultées)

3. Méthodologie

L'organigramme conceptuel (Annexe 3 : complément technique, section 0 : synthèse traitement) formalise l'architecture globale du pipeline de géotraitement, implémenté majoritairement en environnement Python. Ce flux opérationnel s'articule séquentiellement autour de quatre grandes étapes analytiques : (1) le prétraitement et l'harmonisation spatiale des données, (2) la modélisation des coefficients de fiabilité par source, (3) la fusion des référentiels, et (4) le contrôle qualité assorti d'une validation statistique. Concernant l'étape critique d'intégration multi-sources (phase 3 dans l'organigramme), le protocole déploie une architecture duale reposant sur deux approches de classification : un modèle déterministe d'agrégation de scores par classe thématique, et un système de classification hiérarchique par arbre de décision.

3.1 Le prétraitement

3.1.1. Harmonisation des nomenclatures

Le croisement de sources d'OCS ayant des sémantiques et nomenclatures hétérogènes, pose des contraintes méthodologiques importantes dans cette étude. Afin d'assurer une interprétation cohérente des flux, le protocole s'appuie sur la création d'une table de transformation des classes, servant de pivot pour comparer les sources de données autour de 6 classes thématiques : Artificialisé, Agriculture, Nature, Bois, Eau et Autre. Cette standardisation est une condition préalable pour comparer des données de sources différentes dans un référentiel commun. Le référentiel sémantique du Cerema constitue la nomenclature de base du modèle. Néanmoins, la variable mise à disposition (« cgrnum ») donne accès à 13 classes originelles reflétant une OCS fondée principalement sur une logique fiscale plutôt qu'une occupation physique des sols. Cette différence entre fiscalité et aspect physique complexifie la distinction entre certaines classes qui sont physiquement identiques sur le terrain.

Conformément aux recommandations et préconisations méthodologiques fournies par le Cerema, une adaptation regroupant les classes pour stabiliser la donnée a été effectuée. La classe Artificialisé est constituée par la fusion des groupes de culture : jardins (autres que les jardins d'agrément), terrains à bâtir, terrains d'agrément, chemins de fer et canaux, ainsi que sols et carrières. La classe Agriculture est constituée par les vignes, les vergers ainsi que les autres terres agricoles. Cette adaptation vise à assurer une cohérence thématique entre les sources.

Pour intégrer les données (OSO et COSIA), une matrice de passage systématique est appliquée afin de réduire la complexité de la nomenclature. Ce dictionnaire de correspondance permet de projeter les nomenclatures initiales vers les 6 classes thématiques pivots. Le détail complet de cette matrice de passage est disponible en Annexe 3 : complément technique (Section 1 : Matrice nomenclature).

3.1.2. Traitement par millésime

Chaque source est traitée indépendamment pour chaque millésime disponible. À ce stade, aucune synthèse temporelle n'est réalisée : chaque année produit un fichier autonome contenant la distribution surfacique par classe, la classe dominante et la probabilité associée.

- BD TOPO (2025) : Elle caractérise avec une précision élevée les dominances liées aux surfaces artificialisées et aux zones en eau. Les entités bâties, les constructions surfaciques ou encore les infrastructures publiques sont intersectées avec chaque parcelle d'étude pour caractériser une dominance « Artificialisé ». Les surfaces hydrographiques alimentent la classe « Eau ». La couche « Estran » est mobilisée pour la classe « Nature ».
- BD FORÊT (2017) : L'attribut « Typologie de Formation Végétale » (TFV) permet d'identifier les zones principalement arborées (classées en « Bois ») des landes et surfaces herbacées (classées en « Nature »).
- RPG Standard (2018-2024) et Complet (2018-2022) : Le Registre Parcellaire Graphique (RPG) est mobilisé pour la caractérisation des usages agricoles. Les données « Standard », issues des déclarations des agriculteurs à la PAC (Politique Agricole Commune) via le champ CODE_GROUP, permettent de discriminer les surfaces agricoles du milieu naturel. Ainsi, le groupe 17 correspondant aux « Estives et landes » est systématiquement rattaché à la classe « Nature ». Les données complémentaires (« RPG Complet ») apportent des informations supplémentaires en permettant de caractériser de larges parcelles non déclarées à la PAC. Elles fournissent une probabilité de classification pour chaque entité.
- OSO (2018-2023) et COSIA (2018, 2021, 2024) : Elles sont reclassées via la matrice de correspondance. OSO fournit une valeur de confiance pour chaque entité, représentant la fiabilité de sa classification. COSIA fournit pour chaque classe originelle une probabilité

issue de la matrice de confusion du modèle. Cette probabilité, propre à chaque nature de sol, est utilisée comme indicateur de confiance.

Pour chaque source et millésime disponible, une intersection géométrique est réalisée sur chaque parcelle. La surface de chaque entité intersectée est calculée puis agrégée par classe cible permettant de produire, pour chaque parcelle, des champs d'occupation exprimés en pourcentage.

La probabilité associée à la classe dominante est ensuite calculée par une moyenne pondérée par la surface, selon des modalités spécifiques à la nature de chaque source (données déclaratives, administratives ou issues de l'IA). Le détail de la formule mathématique et les règles d'attribution des probabilités/confiances par type de source sont reportés en Annexe 3 : complément technique (Section 2 : Complément traitement par millésime).

Il est à noter que les deux flux RPG sont traités simultanément pour fournir un fichier unique en sortie. Cette harmonisation permet de rendre comparables malgré leurs natures hétérogènes, tout en conservant des mesures de cohérence et de fiabilité propres à chaque expert thématique.

3.1.3. Synthèse temporelle et stabilisation des chronologies

Le passage d'une série annuelle à une consolidation des données par source repose sur des algorithmes individuels pour chaque source temporelle. Cette étape vise à limiter l'influence des erreurs de mesure, de classification ou de déclaration pour dégager une occupation dominante sur une période longue. Cette approche est particulièrement adaptée aux usages agricoles, où des rotations de cultures peuvent induire des changements de classe ponctuels.

L'exploitation de la dimension diachronique des sources (OSO, COSIA, RPG), repose sur un algorithme évaluant la succession des états d'occupation sur la fenêtre temporelle 2018-2024. L'affectation sémantique finale de l'entité géométrique est dictée par un vote de majorité fréquentielle, retenant la classe présentant la plus forte occurrence temporelle. Cette étape de lissage statistique s'avère impérative pour neutraliser les anomalies transitoires et fiabiliser la donnée brute en amont de l'intégration multi-sources. En cas d'éventuelles égalités entre fréquences, une règle d'arbitrage par récence est appliquée. Ainsi, si deux classes apparaissent un nombre de fois identique dans la chronologie de la source, le module permettrait de retenir la nature la plus récente, donc la plus actuelle. Ce protocole n'est pas appliqué aux données de la BD TOPO ou la BD FORÊT de l'IGN étant donné qu'un seul millésime est utilisé pour ces sources. Aucun arbitrage temporel ne leur est appliqué.

Ce formalisme algorithmique aboutit à l'extraction de deux métriques fondamentales de robustesse : d'une part, un indice de stabilité diachronique, destiné à quantifier la persistance thématique de la classe au fil des millésimes ; d'autre part, une probabilité moyenne pondérée, traduisant le niveau de confiance intrinsèque de la modalité sémantique retenue. La formalisation mathématique détaillée de ces estimateurs est explicitée au sein de l'Annexe 3 : complément technique (Section 3 : complément synthèse temporelle et stabilisation des chronologies). Ce lissage chronologique assure que les conflits d'attribution entre bases de données (à l'instar du RPG et d'OSO) ne relèvent pas d'une anomalie transitoire, mais bien d'une divergence structurelle entre leurs ontologies respectives. Une telle démarche fournit ainsi un levier objectif pour auditer la fiabilité inhérente à chaque producteur de données.

3.1.4. Audit de comparabilité et validation de l'interopérabilité

L'audit de comparabilité vise à évaluer la cohérence entre chaque source disponible en 2021, cette date étant la seule disponible où l'ensemble des sources, hors BD FORÊT, ont été rendus disponibles simultanément. L'audit fait office de contrôle méthodologique indépendant des traitements précédents, afin de valider la cohérence de nouvelles nomenclatures cibles.

L'interopérabilité sémantique des nomenclatures ayant été préalablement garantie par les routines de prétraitement annuel, l'audit de fiabilité se fonde sur la distribution proportionnelle des classes,

dérivée du formalisme analytique susmentionné. À l'échelle de l'unité parcellaire, des métriques d'emprise spatiale (ratios surfaciques) sont indexées à chaque classe cible pour l'ensemble des référentiels. Ces indicateurs de proportionnalité constituent le socle quantitatif indispensable à l'analyse de congruence inter-sources.

L'analyse comparative s'articule autour de la confrontation des pondérations surfaciques intra-classes, opérée à l'échelle de l'entité spatiale d'étude. L'adéquation inter-référentiels y est objectivée par une métrique de dispersion reposant sur l'écart-type (des proportions par classe et par parcelle). Des distributions, elle-même couplée à un indicateur composite de divergence globale et à l'extraction des consensus sémantiques. L'appareil mathématique régissant ces estimateurs d'audit est consigné *in extenso* en Annexe 3 : complément technique (Section 4 : Complément sur l'audit de comparabilité).

3.2 Le traitement

3.2.1. Calcul d'un coefficient source

Bien que l'ensemble des sources fournisse un indicateur de fiabilité interne (des probabilités locales par parcelle), ces valeurs ne prennent pas en considération la fiabilité intrinsèque de la source elle-même. L'hétérogénéité intrinsèque de leurs protocoles d'acquisition et de modélisation confère à ces jeux de données une absence d'interopérabilité native, proscrivant toute confrontation directe en l'état.

Afin de pallier cette hétérogénéité structurelle, un coefficient d'étalonnage spécifique à chaque référentiel est modélisé. Ce coefficient a pour objectif d'être un pondérateur global par source. Le principe est d'appliquer ce coefficient à la probabilité/fiabilité de chaque parcelle lors des croisements finaux. Ce mécanisme mathématique garantit la conservation des nullités, tout en pénalisant légèrement les scores de confiance maximaux au prorata de l'incertitude native du référentiel.

Le coefficient est établi via un score de 100 points selon la formule ci-dessous :

$$C_{src} = max\left(10,\ min\left(100,\ \left(Base - (\Delta_t \times M_{age})\right) + M_{meth}\right)\right)$$

Ce coefficient repose sur trois piliers :

- La précision initiale ($Base$) : Elle distingue les sources « expertes » (administratives et vectorielles comme le RPG ou la BD TOPO) des sources issues de télédétection (OSO, COSIA) qui, bien que précises, présentent des bruits ou des lissages géométriques plus importants.
- L'obsolescence temporelle (M_{age}) : Un malus annuel modélise la perte de fraîcheur de l'information. Ce malus est différencié selon la dynamique des classes : il est plus faible pour la BD FORÊT, dont les formations végétales évoluent plus lentement que les parcelles agricoles ou urbaines.
- La méthode de production (M_{meth}) : Ce critère ajuste la confiance selon que la donnée est issue de déclarations et complété par IA (RPG), de relevés neutres (BD TOPO) ou de traitements automatisés par Intelligence Artificielle (OSO, COSIA), ces derniers pouvant induire des incertitudes de classification.

La formalisation mathématique sous-jacente, couplée à l'exhaustivité de la matrice de pondération (incluant les scores intrinsèques de base, les dépréciations temporelles et les incréments méthodologiques), est explicitée *in extenso* au sein de l'Annexe 3 : complément technique (Section 5 : Complément coefficient source).

L'application de la formule de pondération aboutit à la définition des coefficients de confiance finaux utilisés pour le croisement multi-sources (Table 1). Le maintien d'un coefficient élevé permet de préserver la fiabilité relative de la source, tout en servant d'arbitrage pour les données dépourvues d'indice de confiance originel.

Source	Dernière année de la source	Age (2025)	$Base$	$\Delta_t \times M_{age}$	M_{meth}	Score	Coefficient
RPG FUSION (Standard + Complet)	2024	1	95	0.5	1.0	95.5	0.955
OSO	2023	2	85	1.0	-1.0	83.0	0.83
BD TOPO	2025	0	95	0.0	0.0	95.0	0.95
COSIA	2024	1	85	0.5	-1.0	83.5	0.835
BD FORET	2017	8	95	2.0	-0.25	92.75	0.927

Table 1: Résumé du calcul des coefficients sources

3.2.2. Croisement inter-sources

Afin de valider l'architecture définitive du processus de fusion spatiale, un protocole expérimental a été déployé pour confronter deux approches distinctes de classification régissant l'arbitrage sémantique final.

- **Méthode 1** : Une approche de score par classe, où la classe finale est déterminée par le score le plus élevé (Fusion décisionnelle par scoring**).**

Cette approche repose sur une fusion des indicateurs de chaque source via des scores calculés pour chaque classe. La classe obtenant le score le plus élevé est retenue.

$$Score_{cl} = \sum_{src} \left(P_{src} \times C_{src} \times \left(\frac{S_{src}}{1000} \right) \times B_{src,cl} \right)$$

Où P_{src} correspond à la confiance interne pour la parcelle, C_{src}le coefficient source, S_{src} la stabilité et $B_{src,cl}$ l'expertise thématique de la source (expert : 1.5, secondaire : 1.2, léger : 1.1) ces coefficients étant calibrés pour les sources expertes (ex : RPG pour l'agriculture) face aux sources généralistes.

Le score final d'une classe est la somme des scores calculés pour chaque source ayant identifié cette classe. Cette logique d'agrégation valorise intrinsèquement la redondance spatiale : toute convergence thématique inter-sources induit une amplification mathématique du score alloué à la modalité dominante.

Le modèle produit de la même manière des indicateurs de robustesse tels que la marge différentielle entre les deux scores les plus élevés, la fiabilité/confiance finale (moyenne $F_{src} \times C_{src}$ pour les sources ayant la dominance finale), ainsi que le nombre de sources utilisées et disponibles (Annexe 3 : complément technique, Section 6 : Exemple de nomenclature en sortie de la méthode 1).

- **Méthode 2** : Une approche par arbre de décision reposant sur une hiérarchie des sources.

La méthodologie employée pour cette deuxième approche repose sur une logique hiérarchique, nommée « arbre de décision » à plusieurs branches, dont les données d'occupation du sol issues des données satellitaires (OSO, COSIA) sont considérées comme la « couronne » de l'arbre, et les données exogènes (BD TOPO, BD FORÊT, RPG FUSION) considérées comme les « feuilles » de l'arbre, venant vérifier et arbitrer la classification finale de l'occupation du sol à l'échelle des parcelles.

Dans un premier cas, lorsque les classifications OSO et COSIA sont concordantes sur la parcelle (i), la classe commune est retenue, à condition qu'elle soit conforme aux données exogènes. Dans le cas contraire, une branche de l'arbre s'active et contrôle si une autre donnée de référence est proportionnellement dominante sur la parcelle. Si tel est le cas, cette classe est alors retenue. Sinon, la classe commune aux données OSO et COSIA demeure la classe définitive. Dans un second cas, lorsque les classifications OSO et COSIA divergent (ii), l'arbitrage s'effectue en priorisant les catégories thématiques (Eau, Agricole, Artificialisé, Nature, Bois) issues des données exogènes, en considérant leur prédominance spatiale au sein de la parcelle. Cette prédominance est préalablement calculée à partir d'une agrégation spatiale de ces catégories thématiques au sein de la parcelle (Annexe 3 : complément technique, Section 7). Lorsque plusieurs catégories sont disponibles, celle présentant la

plus forte proportion, en accord avec les classifications faites par OSO ou COSIA, est sélectionnée. Dans le cas d'un désaccord avec les données OSO/COSIA, elle reste néanmoins retenue, car issue des données exogènes. En absence de données de référence, un arbitrage est opéré entre OSO et COSIA selon leur niveau de fiabilité respectif. In fine, chaque attribution est assortie d'un indice de confiance, qui traduit la cohérence des sources utilisées et la solidité de la décision finale (Figure 3).

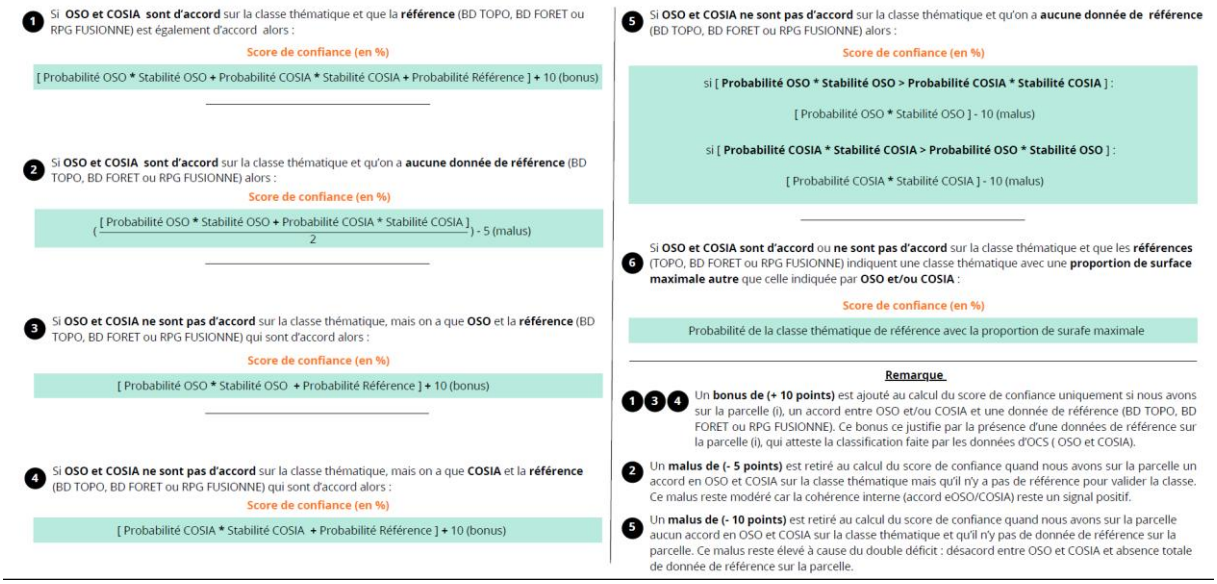


Figure 3 : Méthode de calcul de l'indice de confiance final de la méthode 2 (arbre de décision)

3.3 Validation méthodologique et analyse comparative

Une fois les deux méthodes établies, il convient de prouver leur fiabilité respective dans une phase d'évaluation et de validation des modèles. La comparaison des deux méthodes de croisement est axée sur plusieurs paramètres. Il s'agit d'une évaluation quantitative (i) et qualitative (ii) des résultats de classification. L'évaluation quantitative est basée sur la cartographie des résultats, comparant les résultats aux images satellitaires. Quant à l'évaluation qualitative des résultats, plusieurs métriques sont mobilisées. En effet, il est important de varier les mesures pour des questions de complémentarité de l'information et de robustesse de l'évaluation. Une matrice de confusion est utilisée pour décrire le taux de classification juste pour chaque classe. Le calcul du rappel (la moyenne), déterminant la capacité de chaque méthode à identifier la proportion de vrais positifs correctement classifiés, la précision (la moyenne), mesurant la fiabilité des prédictions positives, et le score F1 (accord harmonique de la précision et du rappel), permettant d'évaluer équitablement la performance de chaque méthode sur l'ensemble des classes, en particulier celles moins représentées dans le jeu de données.

Les métriques sont calculées sur la base des caractérisations morphologiques des parcelles sélectionnées, afin d'éviter un calcul généralisé pour chaque dalle. Ce choix permet une vérification deux en un (méthodes 1 et 2 avec la réalité terrain) et l'observation de la performance des modèles selon le type de classe thématique et le type de parcelle disponible. Cette caractérisation morphologique des parcelles cadastrales est réalisée selon la méthode développée par Puech T. et Mignolet C. (2022). A cette fin, un ensemble d'indicateurs morphologiques est calculé pour chaque parcelle, tels que : la surface (S), le périmètre (P), le nombre de segments qui composent le polygone (Nseg), le nombre minimum de segments nécessaires pour représenter la moitié du périmètre total (N50), l'élongation (ELONG) (Puech T. et Mignolet C., 2022) et le Parcel Shape Index (PSI) (Demetriou D. et al., 2013). Une analyse en composantes principales (ACP) sur l'ensemble du jeu de données (les parcelles cadastrales) est effectuée afin d'harmoniser et de normaliser les indicateurs exprimés dans des unités différentes. À l'issue de l'ACP, seuls les trois premiers axes factoriels sont sélectionnés car leur pouvoir explicatif est jugé suffisant pour structurer le jeu de données (inertie > 1 %). Les résultats

de l'ACP servent ensuite de support pour une classification ascendante hiérarchique (CAH) selon la méthode de Ward (Ward J. H., 1963). Cette démarche permet de partitionner les parcelles en quatre classes morphologiques distinctes, identifiées sur la base du dendrogramme de classification (sauts d'inertie). L'annexe 3 complément technique (Section 8 : Caractérisation morphologique des parcelles cadastrales) synthétise l'ensemble de cette chaîne de caractérisation.

L'ensemble des métriques utilisées pour l'évaluation quantitative (la précision, le rappel, le score F1) est calculé pour chaque type de parcelle, à partir d'un jeu de données réalisé par nos soins, appelé « vérité terrain », constitué à partir de la photo-interprétation de l'imagerie satellitaire « Google Satellite – année 2025 », selon le principe suivant : choisir des parcelles sur lesquelles nous sommes capables de déduire une classe d'occupation du sol selon le principe de la majorité de la surface occupée. Si une parcelle donnée présente différentes classes thématiques, mais qu'à partir de la photo-interprétation la classe « Nature » par exemple apparait comme la dominance, alors la parcelle est classée comme telle (Figure 4).



Figure 4 : Exemple de photo-interprétation des parcelles pour le jeu de données de la vérité terrain

Le nombre de parcelles retenues pour constituer la vérité terrain varie d'un type de parcelle à l'autre, en raison des parts occupées par les classes thématiques, qui varient également d'un type de parcelle à l'autre. Cependant, un total de 157 parcelles a servi à constituer la vérité terrain.

3.4 Diagnostic de la méthode retenue

Au terme du protocole expérimental et après validation du modèle présentant les performances optimales, un diagnostic approfondi est intégré aux résultats afin d'en permettre une analyse détaillée.

Il évalue la qualité même de l'arbitrage en corrélant le nombre de sources en accord avec la dominance finale et la stabilité de l'indice de confiance, permettant ainsi de contrôler si l'ajout de sources augmente la fiabilité globale. Une analyse par netteté décisionnelle est appliquée en contrôlant le score le plus élevé entre la première classe déterminée et son dauphin, afin d'identifier les classes thématiques où l'arbitrage est le plus sensible. Des indicateurs sont agrégés pour produire trois niveaux de confiance (moyen < 50%, acceptable (50 - 75%) et robuste >75%). Ce protocole permet ainsi de qualifier précisément la robustesse et la qualité de la méthode retenue, permettant d'orienter les besoins futurs nécessaires pour une classification plus juste de certaines classes.

3.4 Evaluation de l'influence des sources de données sur la performance du modèle

Pour finaliser le protocole d'étude, un audit de robustesse a été mise en œuvre afin de quantifier l'importance de chaque source de données dans la décision finale. La méthode repose sur une exclusion itérative (leave-one-out) qui consiste à retirer, pour chaque itération, une source dans la décision. Ainsi, pour chaque scénario d'exclusion, l'intégralité du processus méthodologique explicité auparavant pour la méthode retenue est appliquée, puis confrontée à la référence intégrant l'ensemble des sources.

$$\Delta F_s = C^{-}_{total} - C^{-}_{excl(s)}$$

Où ΔF_s est l'impact de la source s sur la confiance globale du modèle, $\bar{C}_{total}$ la confiance moyenne obtenue lorsque toutes les sources sont utilisées et $\bar{C}_{excl(s)}$ la confiance moyenne obtenue après exclusion de la source s.

Ce protocole extrait différents indicateurs permettant d'identifier le rôle de chaque flux. L'analyse des transferts nets et la matrice de mutation quantifient l'instabilité physique du modèle en comptabilisant les changements de classe. Si l'absence d'une source provoque une mutation massive vers une classe secondaire, son rôle de « verrou thématique » est alors démontré. Une surveillance de la variation des marges de victoire et de l'érosion de l'indice de confiance est menée, afin de distinguer les sources permettant le choix de la dominance finale et celles utiles simplement pour stabiliser l'indicateur de fiabilité final.

3. Résultats

Dans un souci de clarté, seuls les résultats présentant un intérêt direct pour l'évaluation des méthodes sont rapportés.

3.1 Audit de comparabilité et validation de l'interopérabilité

La Figure 5 - A montre le profil distributionnel du taux d'accord entre les cinq sources de données (OSO, COSIA, BD TOPO, BD FORÊT, RPG). Cet indicateur quantifie le degré de convergence des estimations surfaciques pour les cinq classes d'occupation du sol (Eau, Agriculture, Artificialisé, Bois, Nature) à l'échelle de l'unité foncière. L'observation de l'histogramme met en exergue un mode statistique important compris dans l'intervalle [87,5 - 90 %], indiquant que pour la grande majorité des parcelles, les sources convergent vers une estimation quasi unanime des proportions thématiques. Toutefois, l'examen de la divergence inter-classes (Figure 5 - B) révèle une forte disparité selon les thématiques observées. La classe « Artificialisé » est celle qui présente le niveau de litige le plus marqué, affichant un indice de divergence supérieur à 25 %. Les classes « Nature », « Agriculture » et « Bois » présentent des niveaux de divergence intermédiaires, avec des indices s'établissant respectivement à 15%, 12% et 15%. En définitive, la classe « Eau » se distingue par une adéquation optimale entre les référentiels, la divergence inter-sources y demeure strictement résiduelle (2%).

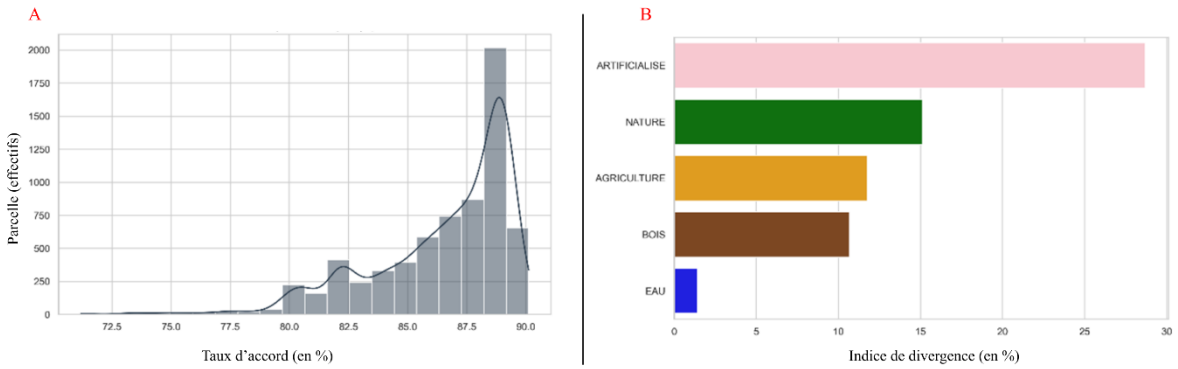
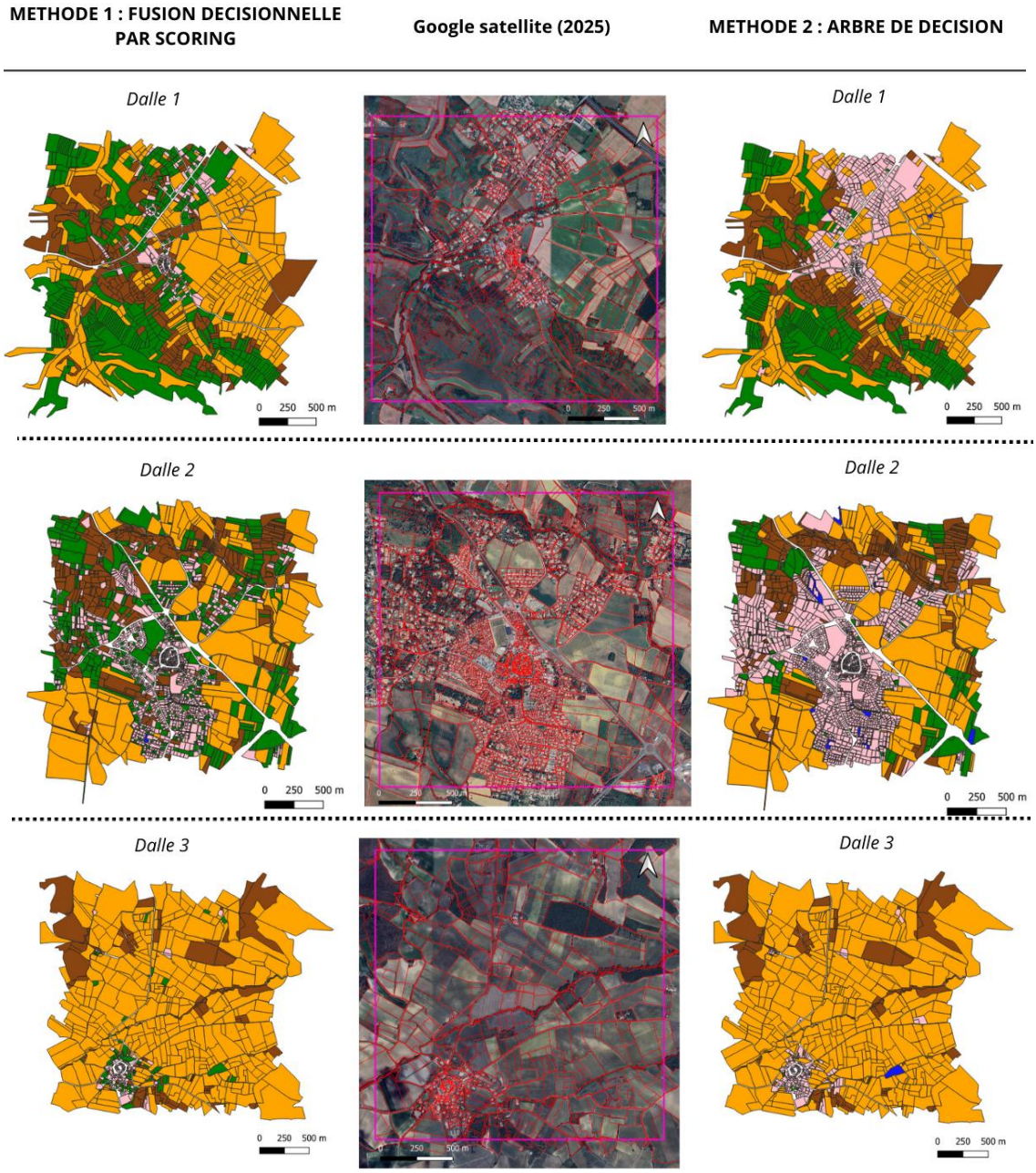


Figure 5 : Résultats de l'audit de comparabilité et validation de l'interopérabilité

3.2 Analyse comparative des deux approches de mise à jour de l'OCS des parcelles cadastrales

3.2.1. Evaluation qualitative

La Figure 6 décrit visuellement les résultats issus des deux techniques de classification : l'approche par la méthode 1 (fusion décisionnelle par scoring) montre une occupation et l'approche par la méthode 2 (arbre de décision). La méthode 1 décrit une OCS plus fragmentée, notamment visible sur les dalles 1, 2, 3 et 4 où les parcelles classées en « Nature » et « Bois » sont très représentés dans le tissu urbain. Cela parait, à première vue, différente de la configuration observée sur les images « Google satellite ». Cependant, lorsque l'on observe l'imagerie satellitaire à une échelle plus grande (Figure 7), on peut constater que ces parcelles classées en « Nature » ou « Bois » au sein du tissu urbain correspond à la végétation ou arbre isolé généralement intégrée à l'habitat, dont l'occupation est quasi majoritaire à l'échelle de la parcelle. A contrario, l'approche par la méthode 2, montre une OCS plus lissée, avec d'avantage de parcelles classées en « Artificialisé » par rapport à la méthode 1. De plus, l'examen détaillé de la dalle 2 met en lumière quelques unités parcellaires classé en « Eau », bien que la confrontation avec l'imagerie satellitaire n'y atteste la présence d'aucune surface hydrographique.



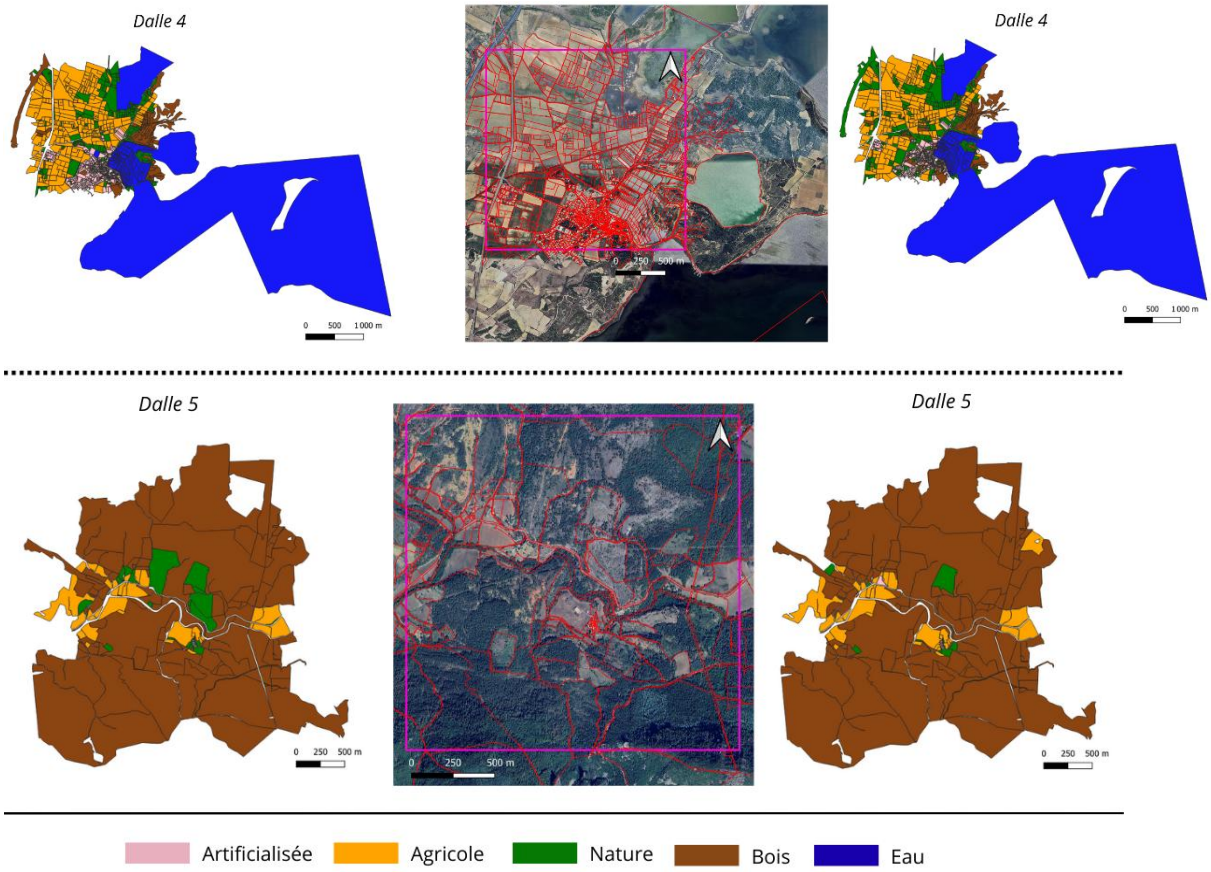


Figure 6 : Cartographie des deux méthodes de classification en comparaison de l'image satellitaire

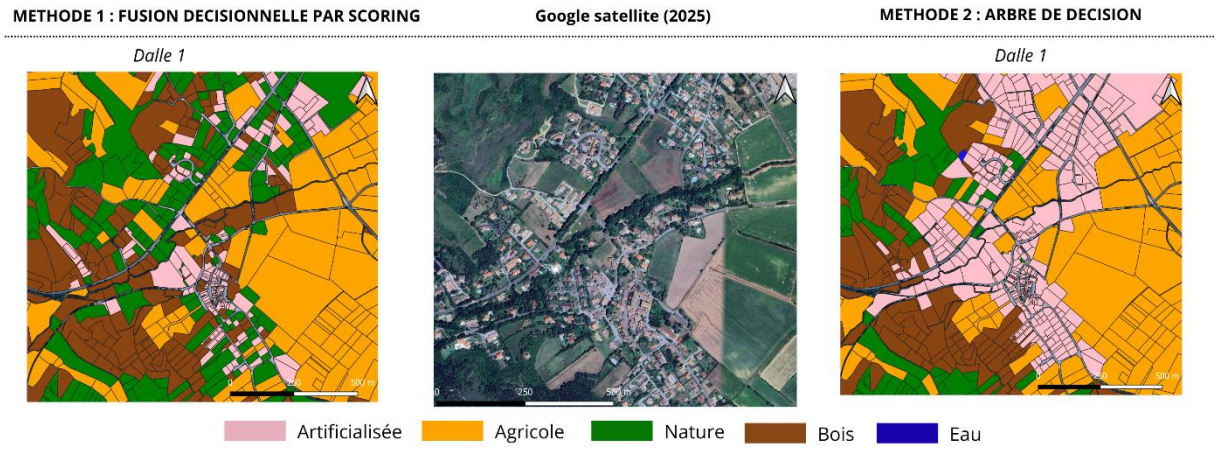


Figure 7 : Observation à grande échelle

L'analyse des matrices de "ressemblance/divergence'' (Figure 8), réalisées sur l'ensemble des cinq dalles, montre statistiquement que les classes « Agricole » et « Eau » sont celles pour laquelle les deux méthodes sont le plus en adéquation, avec moins de divergence. En revanche, les divergences majeures concernent le passage de la classe « Nature » (Méthode 1) vers la classe « Artificialisé » (Méthode 2) ainsi que celui de la classe « Bois » (Méthode 1) vers « Artificialisé » (Méthode 2). Ce glissement est présent sur l'ensemble des dalles, mais il est particulièrement marqué sur les dalles 1 et 2, où l'on compte respectivement 141 et 367 parcelles classées en « Nature » par la méthode 1, mais en « Artificialisé » par la méthode 2.

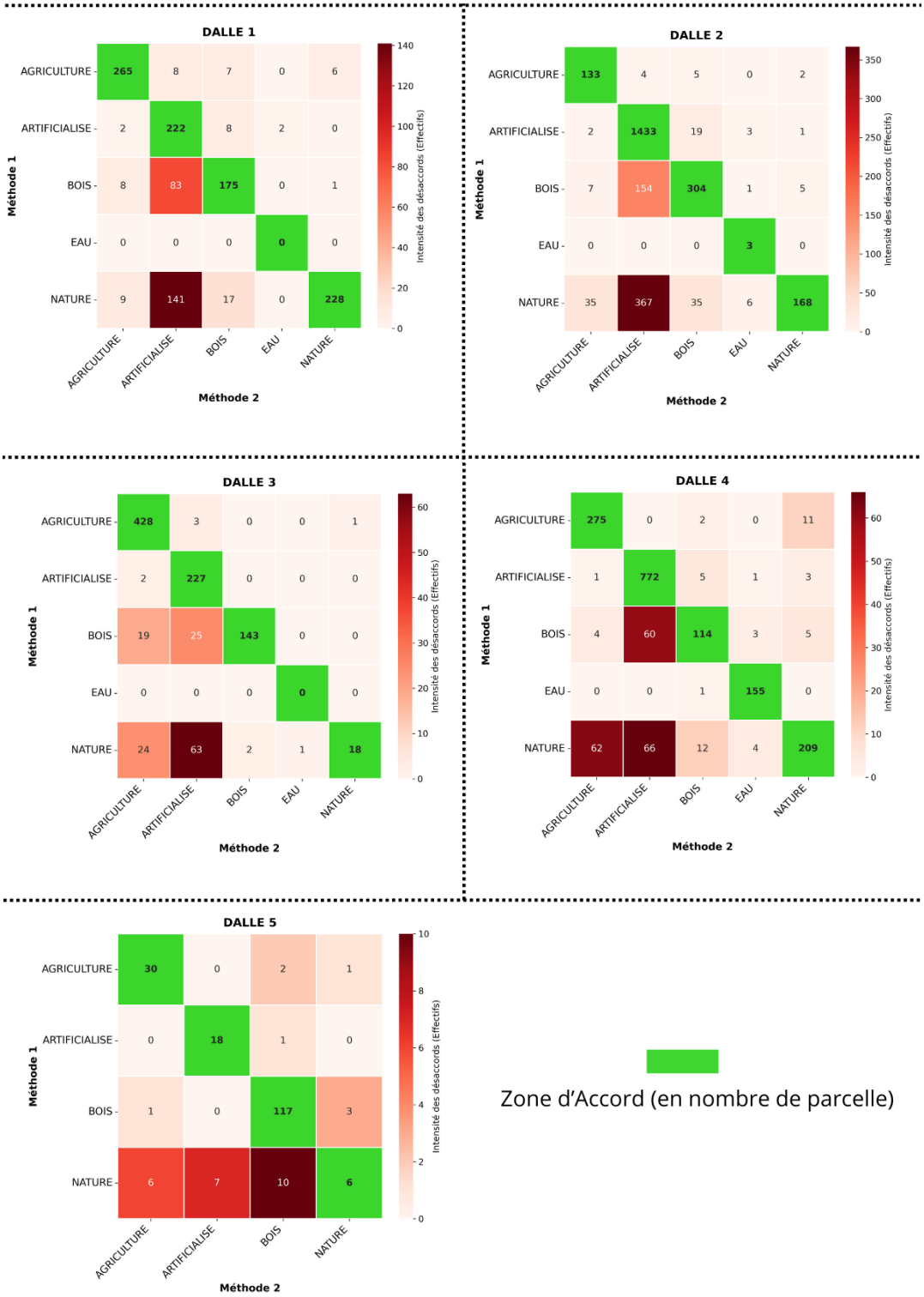


Figure 8 : Matrice de convergence/divergence de la classification issue des deux méthodes

3.2.2. Evaluation quantitative

Dans un premier temps, la caractérisation morphométrique des parcelles a permis de segmenter l'ensemble des parcelles tests (6 792) en quatre typologies structurelles distinctes (Figure 9) : les grandes parcelles de forme complexe (263), les grandes parcelles de forme simple (1 620), les petites parcelles de forme rectangulaire ou carrée (4 381) et les petites parcelles de forme étirée ou allongée

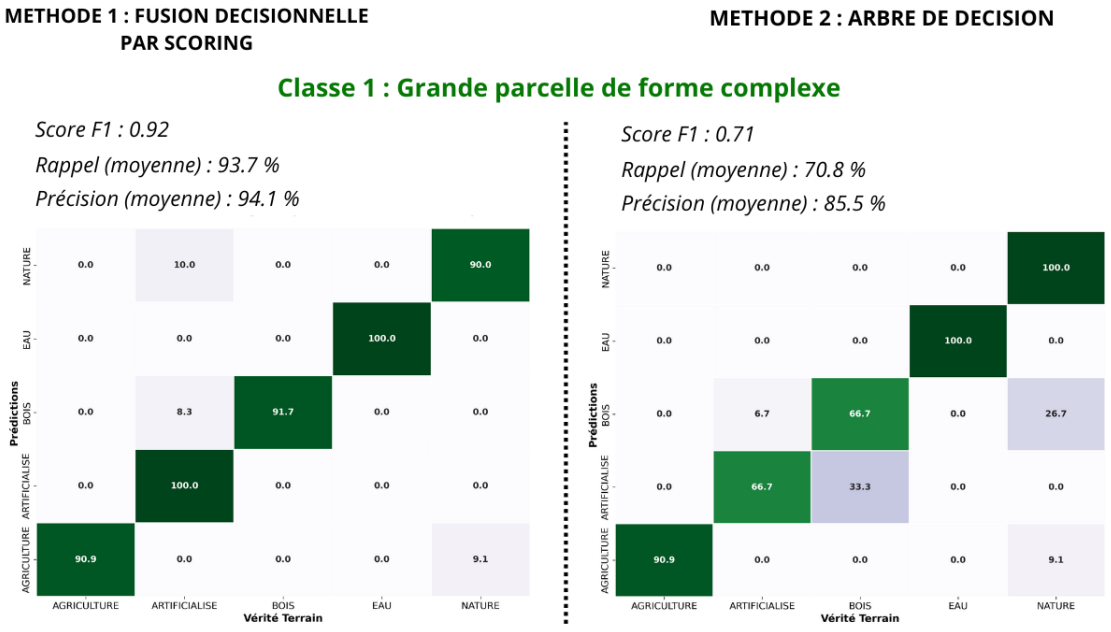
(528). Ces morphologies sont caractérisées à partir de la distribution statistique des indicateurs calculés (par exemple l'aire) et la complexité de leurs contours.

L'analyse des matrices de confusion établies pour chaque type de parcelle met en exergue des différences marqué de performance entre les deux méthodes. La méthode 1 démontre une stabilité systémique, maintenant un F-score global supérieur à 0,91 quel que soit la morphologie des parcelles, tandis que la méthode 2 présente des F-score inférieures à 0,80 dans certains cas.

Pour les grandes parcelles, qu'elles soient de forme complexe (classe 1) ou simple (classe 2), la méthode par scoring présente des F-score compris entre 0,92 et 0,95, avec des rappels et des précisions supérieurs à 93 %. La lecture des matrices de confusion montre une identification de haute précision pour les classes « Artificialisé », « Bois », « Eau » et « Nature ». A l'inverse, la méthode par arbre de décision affiche une baisse significative des performances sur ces mêmes emprises, en particulier pour la classe 2 (précision moyenne = 84,1 %).

Pour les petites parcelles de forme rectangulaire ou carrée, la méthode 1 présente d'excellente performance, avec une précision et un rappel de 95,6 % et 92 % respectivement. Les classes « Artificialisé » et « Eau » affichent une précision supérieure à 90 %. Seule la classe « Agricole » présente une confusion élevée avec la classe « Artificialisé ». Sous la méthode 2, cette même classe présente une précision supérieure à 88 %, à l'exception de la classe « Artificialisé » ou on observe une forte confusion avec la classe « Nature ».

Enfin, sur les petites parcelles de forme allongée la situation s'inverse. La méthode par arbre de décision obtient les meilleures performances (F-score = 0,93), légèrement devant la méthode par scoring (F-score = 0,90). Les classes « Artificialisé » et agricole sont bien identifiées avec la méthode 1, tandis que les classes « Eau » et « Nature » le sont moins. Avec la méthode 2, ce sont les classes « Artificialisé » et « Bois » qui présentent des confusions.



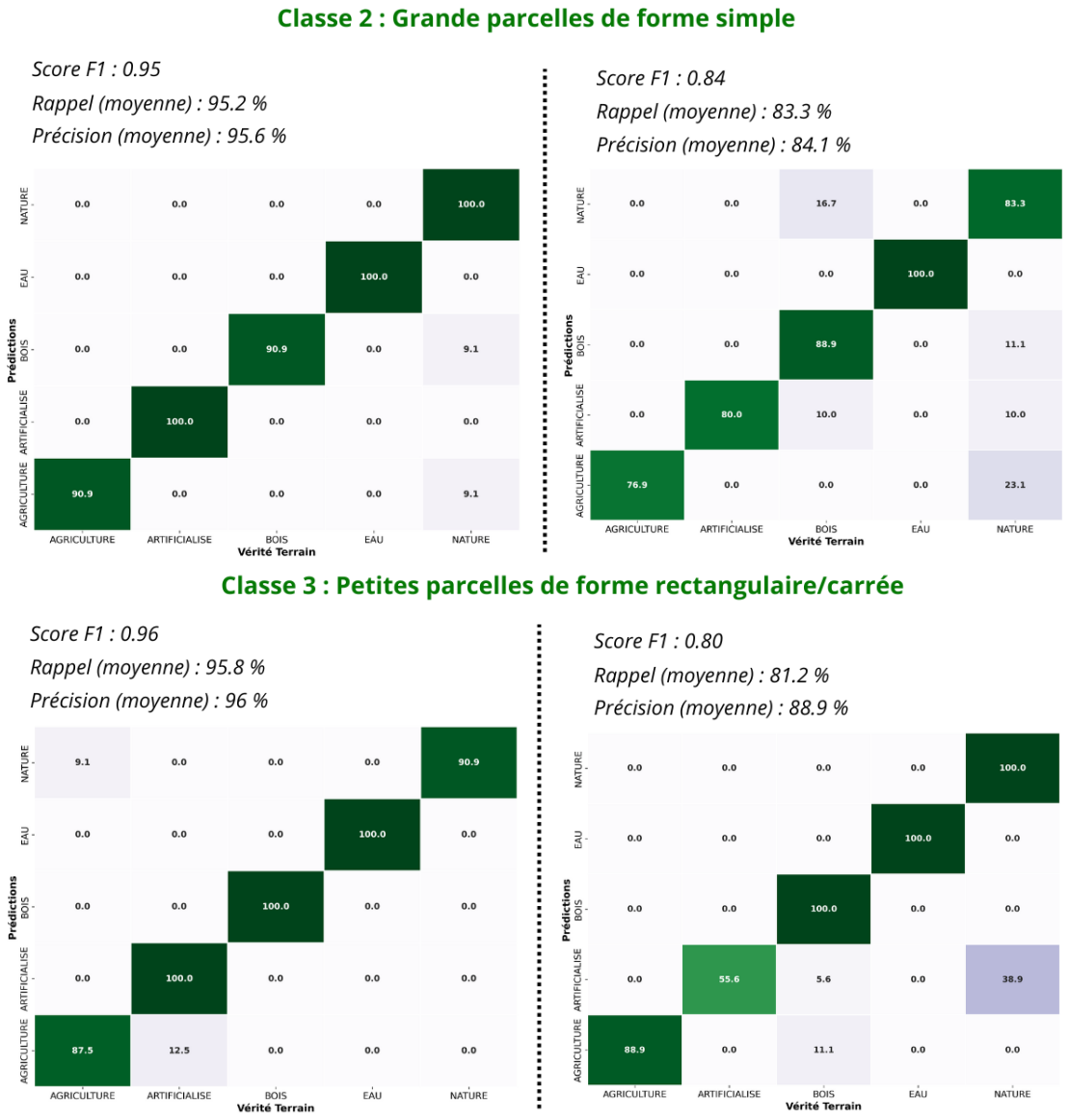


Figure 9 : Matrice de confusion

3.3. Diagnostic de la méthode retenue : méthode 1

Les Figure 10 et 11 restituent la synthèse statistique des métriques sous-jacentes de la Méthode 1 (fusion décisionnelle par scoring).

En premier lieu, la distribution de l'indice de confiance final à l'échelle des parcelles, montre un niveau de de fiabilité globalement assez élevé (Figure 10 - A). Environ 20 % des parcelles atteignent un seuil de confiance qualifié de robuste (> 75 %), tandis que 80 % s'inscrivent dans une marge de confiance acceptables (50 – 75 %). La part des parcelles présentant un indice de confiance moyen (< 50 %) demeure strictement résiduelles (0,4 %). Cette performance globale masque néanmoins une hétérogénéité selon les classes thématiques (Figure 10 - B), la dispersion intra-classe étant matérialisée par l'écart-type. La classe « Eau » se distingue par un indice de confiance élevé (plus de 80 %) et une très faible variabilité. Les classes « Agriculture », « Bois » et « Artificialisé » occupent la deuxième position, avec des indices de confiance globalement satisfaisants (entre 60 et 80 %). Seule la classe « Nature » présente des indices de confiance plus faibles (entre 40 et 60 %).

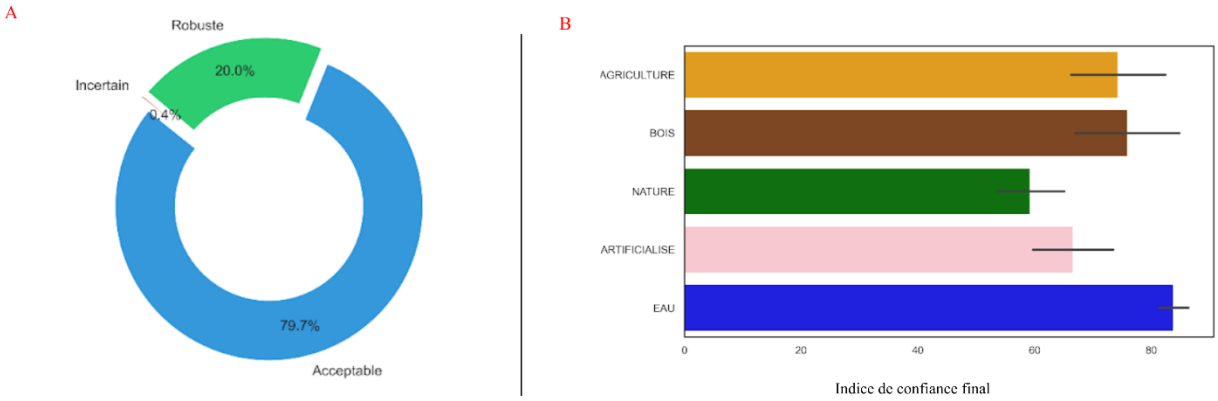


Figure 10 : Synèse statistique de la méthode 1 – Indice de confiance

Par ailleurs, le profil de mobilisation des sources (Figure 11) indique que l'attribution de la classe thématique final dans une parcelle repose sur un consensus trivalent (trois sources sur cinq), représentant 37,5 % des parcelles tests. Les parcelles classées par une seul source représentent que 28,5 % (Figure 11 - A). Cependant, plus le nombre de source impliqué dans la classification de la parcelle est élevé, plus la confiance finale augmente (Figure 11 - B). Lorsque trois ou quatre sources sont mobilisées, la médiane de la confiance est relativement élevée, autour de 74 %, et les valeurs sont peu dispersées (environ 5 %). À l'inverse, lorsqu'une seule source est mobilisée, la médiane reste à 57 % et les valeurs sont plus dispersées (environ 15 %).

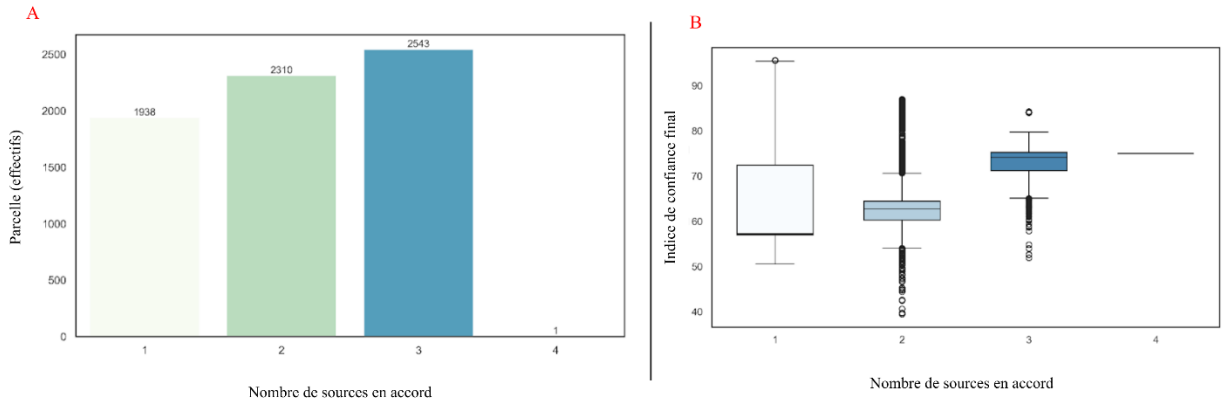


Figure 11 : Synthèse statistique de la méthode 1 - Profil de mobilisation des sources

3.4. Evaluation de l'influence des sources, réalisée avec la méthode 1

La Figure 12 décrit les bilans de transferts interclasses à l'exclusion itérative d'une source (approche leave-one-out). Une valeur positive indique que la classe "gagne'' des effectifs de parcelle en l'absence de la source testé, tandis qu'une valeur négative indique qu'elle perd des effectifs. L'analyse des amplitudes révèle la forte sensibilité du modèle pour la variable COSIA (Figure 12 – A) : son retrait entraîne une perte d'effectif massive pour les classes « Nature » et « Bois », mais un gain pour la classe « Artificialisé ». De manière analogue, l'exclusion de la BD FORÊT induit un déficit marqué pour la classe « Bois », assorti d'une légère surreprésentation au profit de la classe « Nature ».

Concomitamment, la Figure 12- B synthétise les résultats de l'audit de sensibilité, évaluant l'impact de chaque source sur la variation de l'indice de confiance final. Une valeur négative indique une dégradation des indices de confiances, tandis qu'une valeur positive signale une consolidation de la confiance. On remarque le retrait d'OSO entraine un gain de confiance quasi généralisé pour l'ensemble des classes thématiques. À l'inverse, l'exclusion des données exogène tels que le RPG ou la BD FORÊT, entraine une chute brutale des indices de confiances.

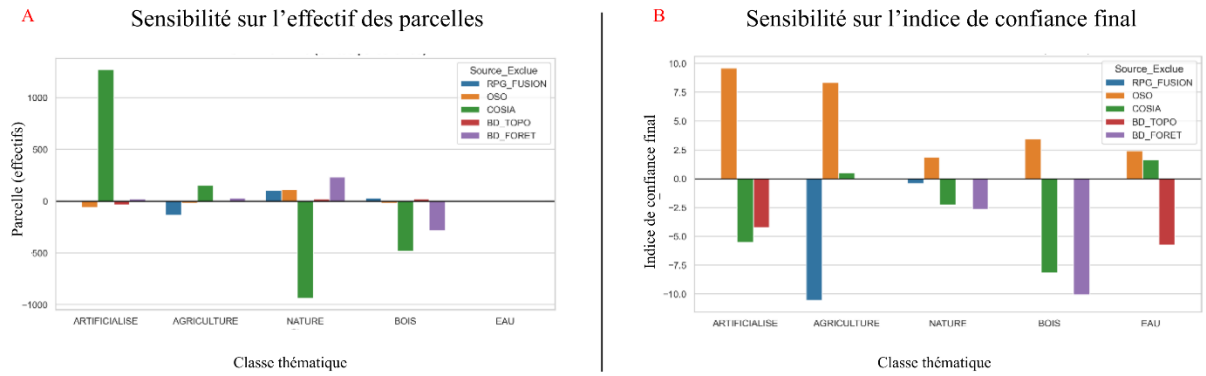


Figure 12 : Audit de sensibilité des sources

4. Discussion

4.1 Analyse des résultats obtenus

Le résultat de l'analyse de la cohérence inter-sources affirme la comparabilité des sources et leur validité, malgré la présence de certaines différences entre elles. Deux facteurs majeurs expliquant certaines divergences observées ont été identifiés : morphologique et sémantique. Lorsque les formes deviennent plus complexes, cela peut créer des désaccords. En ce qui concerne la typologie, la classe Eau est très stable et les données sont presque toujours d'accord à ce sujet. En revanche, la classe Artificialisé est celle qui présente le plus de différences. Cette vulnérabilité ne relève pas uniquement d'erreurs techniques, mais peut s'expliquer par une hétérogénéité sémantique majeure, notamment sur les zones anthropiques : inclusion variable des jardins, du bâti léger ou de la voirie. De plus, la simplification vers une classe dominante fournit un indicateur de fiabilité. La classe Nature marque une instabilité, traduisant la difficulté à caractériser les mosaïques paysagères complexes et les milieux de transition.

Les deux approches du croisement multi-sources mettent en évidence aussi bien des ressemblances que des dissemblances pour la mise à jour de l'OCS des parcelles cadastrales. L'analyse comparative relève visuellement et statistiquement que ces différences entre les deux approches s'expriment : (1) dans l'organisation spatiale de l'OCS, (2) dans la morphologie des parcelles et (3) dans leur sens thématique, donc l'occupation du sol ou son usage.

La méthode 1 (fusion décisionnelle par scoring) produit une OCS plus fragmentée, surtout dans les zones urbaines. Beaucoup de parcelles classées Nature ou Bois se trouvent au milieu du bâti, ce qui peut paraître contre-intuitif au premier abord. Cependant, sur l'imagerie satellitaire à grande échelle, ces zones correspondent bel et bien à des espaces verts : jardins, pelouses, haies ou arbres isolés, qui occupent en réalité une grande partie de nombreuses parcelles résidentielles. Ainsi, cette méthode semble mieux coller à la réalité du terrain à l'échelle de la parcelle, en cohérence avec la définition stricte de l'OCS. À l'inverse, la méthode 2 (arbre de décision) produit une OCS plus lisse et homogène, où la classe Artificialisé est davantage représentée, notamment dans les zones urbanisées. Cette approche a tendance à absorber les petits espaces verts internes aux parcelles bâties dans une grande catégorie Artificialisée. Le résultat se rapproche ainsi davantage d'une lecture fonctionnelle ou d'usage du sol que d'une description détaillée de sa couverture réelle. En effet, la matrice de ressemblance et de divergence vient confirmer statistiquement ces différences. La forte correspondance observée pour la classe Agricole s'explique assez facilement par le fait que ces parcelles sont généralement de grande taille et couvertes par une seule OCS, ce qui réduit considérablement les risques de confusion lors de la classification. En revanche, les écarts récurrents entre les classes Nature (méthode 1) et Artificialisé (méthode 2) révèlent à quel point les règles intégrées dans l'arbre de décision jouent un rôle clé dans la décision finale des classes.

L’évaluation qualitative par la caractérisation morphologique des parcelles cadastrales apporte un éclairage important sur les conditions de performance de chaque méthode. La méthode 1 est présente une classification très fiable, peu importe la taille ou la complexité géométrique des parcelles. Cette performance peut s’expliquer par sa capacité à bien croiser différentes sources, notamment grâce au “scoring” des sources (nombre de sources compétentes, nombre de sources en accord, degré de fiabilité et stabilité des OCS OSO et COSIA). À l'inverse, la méthode 2 est plus sensible à la forme des parcelles. Ses résultats faiblissent nettement sur les grandes parcelles, surtout celles de forme simple, ainsi que sur les petites parcelles de forme carrée ou rectangulaire. En revanche, elle reste efficace sur les petites parcelles allongées, car ces dernières correspondent souvent à des éléments assimilés à une seule occupation du sol, comme des routes, des bandes de culture ou des pistes. En résumé, la méthode 2 devient plus fidèle à la réalité du terrain, lorsque la lecture est orientée vers l’usage du sol. Elle met l'accent sur la dominance surfacique des classes issues des données exogènes. C’est la raison pour laquelle elle est très sensible à leur absence par rapport à la réalité du terrain. Les règles de décision ont des difficultés à saisir la diversité du sol dans une parcelle, surtout lorsqu’elle présente une grande superficie. En revanche, la méthode 1 est considérée comme une approche plus équilibrée et robuste en termes d'applicabilité et de gestion de l’incertitude, permettant une excellente mise à jour de l’OCS des parcelles cadastrales. Ce modèle conserve une lecture graduelle de la classification finale en intégrant les informations complémentaires en dernier. Cela est particulièrement utile dans les cas de disponibilité inégale des données ou de conflits entre sources au sein d’une parcelle. Il limite également l’influence excessive d’une source unique : l’incertitude est intégrée et visible plutôt que masquée.

En effet, l'analyse synthétique de la méthode retenue (méthode par scoring) confirme les principes précédemment constatés. La classe Eau est, dans la plupart des cas, bien identifiée avec une faible ambiguïté, montrant une confiance élevée et très stable. À l’inverse, la classe Nature présente les niveaux de confiance les plus faibles. Par ailleurs, une forte divergence pour la classe Artificialisé a été constatée. Dans l'ensemble, l'indice de confiance est assez élevé, ce qui met en avant la pertinence de cette méthode. La corrélation positive entre le nombre important de sources utilisées et la confiance finale met en lumière l'intérêt et la robustesse de l'analyse multi-sources.

L'analyse finale de la sensibilité sur l'approche 1 met en lumière l'importance de chaque source et leur rôle dans le modèle. COSIA joue un rôle structurant dans la détermination finale. D'une part, elle contribue à l'intégrité des classes Nature et Bois. D'autre part, elle freine la sur-artificialisation. En outre, l'augmentation de l'indice de confiance suite au retrait d'OSO constitue un résultat contre-intuitif majeur qui doit être interprété avec prudence. OSO et COSIA étant souvent les deux seules sources ubiquistes, elles opèrent en contre-pouvoir. OSO est donc indispensable, non pour sa précision intrinsèque, mais pour sa fonction systémique de modérateur de certitude. À l'inverse, le RPG, la BD FORÊT et la BD TOPO agissent comme des amplificateurs de certitude pour leurs thématiques (Agriculture, Bois et Eau), lorsque l’accord est déjà établi. Il apparaît également les limites liées à l’absence de véritables sources expertes pour les classes Nature et Artificialisé. Par ailleurs, la domination de COSIA dans le volume de reclassifications s'explique par la conjonction de sa couverture exhaustive et de sa qualité interne favorable. Cela peut également traduire le fait que le modèle final dépend fortement d'un seul producteur de données pour certaines classes. En effet, OSO et COSIA sont disponibles dans toutes les parcelles. Ces deux sources étaient en accord dans la moitié des cas (53 %). Toutefois, en cas de désaccord, COSIA contribue fortement à la détermination finale (44 %). En résumé, toutes les sources sont indispensables et se complètent entre elles.

4.2 Limites et apports globaux

Bien que la méthodologie construite permette toutes ces analyses approfondies, plusieurs limites doivent être prises en compte. Dans les analyses, quatre grands facteurs influençant le choix méthodologique, les résultats et l'interprétation ont été relevés : (1) sémantique et typologie, (2)

morphologie parcellaire, (3) organisation spatiale de l'OCS et (4) information recherchée (occupation ou usage).

Premièrement, les limites concernant la nature et la qualité des données mobilisées doivent être mises en avant. Une harmonisation approfondie de la nomenclature est nécessaire au début du processus afin de la mettre en cohérence avec celle du cadastre. Cette complexité de la nomenclature et de la définition des classes est, selon nous, un élément central susceptible d'influencer la détermination finale dans le cadre du croisement multi-sources. De plus, les sources présentent des résolutions spatiales et des précisions géométriques inégales. Pour rappel, la vocation première du cadastre n'est ni thématique ni physique (segmentation sémantique), mais administrative, ce qui implique une forte probabilité de diversité d'OCS au sein d'une même parcelle. De nombreuses entités spatiales présentent un paysage en mosaïque, ce qui peut provoquer des erreurs de représentation et de classification, notamment dans les espaces hétérogènes ou fortement artificialisés (Verburg P. et al., 2011). Pour y répondre au mieux, le choix des données et de la méthodologie doit être prudent. Par exemple, l'utilisation exclusive d'OSO peut poser des problèmes. Sa résolution intermédiaire (10 m) peut s'avérer limitante pour des analyses à l'échelle parcellaire (la superficie moyenne d'une parcelle cadastrale ≈ 5 930 m² dans le département de l'Aude).

Deuxièmement, certaines limites concernant les choix méthodologiques ont été identifiées. La méthodologie repose sur des règles hiérarchiques, des seuils et des paramètres (priorités entre sources, importance, etc.) que nous avons définis, ce qui peut influencer directement les résultats. Bien que ces choix soient justifiés en nous appuyant sur la littérature scientifique et sur les caractéristiques des données, ils restent en partie subjectifs. Une analyse de sensibilité pourrait être menée en modifiant les seuils afin d'étudier l'importance de leur impact sur les résultats. De plus, l'absence de données de référence indépendantes reflétant pleinement la réalité limite la validation des résultats. Les attributs OCS du cadastre ne sont pas accessibles au grand public. L'évaluation réalisée repose principalement sur des comparaisons internes entre sources et méthodologies utilisées.

Par ailleurs, l'automatisation du processus permet une reproductibilité et un gain de temps considérables. En revanche, elle implique également une simplification de la réalité et du processus. Afin de renforcer la stabilité et la fiabilité, il serait possible d'ajouter d'autres données ou d'utiliser des données plus simples ou plus étendues mais plus précises. À titre d'exemple, l'intégration de données linéaires des réseaux de transport ou de données raster comme le RGE ALTI serait intéressante à tester, bien qu'elles n'aient pas été retenues dans cet article. Cependant, cela peut entraîner des temps de calcul élevés et des traitements lourds, comme nous l'avons déjà observé avec le traitement de COSIA. Il est donc indispensable de prendre en considération les compromis entre finesse des résultats, robustesse méthodologique et coûts de calcul (Thierry C. et al., 2020).

En somme, le produit final doit être interprété comme une représentation synthétique et informationnelle issue du croisement raisonné de sources hétérogènes, et non comme une description exhaustive reflétant parfaitement la réalité du sol. Néanmoins, notre méthodologie propose des pistes susceptibles de pallier ces limites. Les deux approches de croisement attribuent une classe dominante reposant davantage sur l'occupation (modèle 1) ou sur l'usage (modèle 2). Il est donc impératif de choisir l'approche la plus pertinente en fonction de l'information recherchée. De plus, notre méthodologie permet de consulter la deuxième ou la troisième classe possible ou concurrente considérée comme dominante dans une parcelle, ainsi qu'une marge d'incertitude. Pour ces raisons, cet article ouvre une première piste dans la mise à jour automatique du cadastre et propose plusieurs perspectives d'amélioration concernant l'occupation et l'utilisation du sol.

5. Conclusion et perspectives

En définitive, cette recherche valide l'hypothèse centrale de notre étude et atteint les objectifs méthodologiques fixés : l'actualisation automatisée de l'attribut « Occupation du Sol » (OCS) de la matrice cadastrale gagne considérablement en fiabilité grâce à la synergie d'une approche multi-sources et diachronique. En confrontant deux paradigmes de classification, l'étude objective la supériorité du modèle par fusion décisionnelle (Méthode 1), qui garantit une robustesse systémique face aux hétérogénéités morphométriques des unités foncières. Ce protocole préserve la réalité physique du terrain (*Land Cover*) et érige l'incertitude en une métrique de qualification explicite. A contrario, l'heuristique par arbre de décision (Méthode 2) s'avère pertinente pour l'analyse des usages anthropiques (*Land Use*), bien qu'elle demeure tributaire de la prépondérance surfacique des intrants exogènes. Si l'automatisation garantit la reproductibilité opérationnelle, elle simplifie irréductiblement la réalité terrain, imposant un compromis strict entre finesse de résolution, robustesse et coût computationnel. Posant les jalons d'un dispositif territorial évolutif, ce protocole s'affinera à terme via l'assimilation d'infrastructures vectorielles linéaires (réseaux de transport), de données altimétriques (RGE ALTI), ainsi que d'autres données spécialisées, optimisant la caractérisation des milieux complexes sans pour autant saturer l'architecture algorithmique.

6. Bibliographie

Batista F e S., Lavalle C. et Koomen E., (2013). A procedure to obtain a refined European land use/cover map, *Journal of Land Use Science*, 8(3), pp. 255-283. DOI : 10.1080/1747423X.2012.667450

Bonnefond M. et Desrousseaux M., (2021). *Le cadastre : un outil foncier tout autant qu'un outil fiscal.* Consulté le 4 février 2026. URL : https://www.foncier-developpement.fr/wp-content/uploads/2021_Fiche-foncier-Cadastre_fr.pdf

Cerema, (s.d.), *Liste des attributs par table*, Dans : Table des Fichiers Fonciers. Consulté le 4 février 2026. URL : https://piece-jointe-carto.developpement-durable.gouv.fr/NAT004/DTerNP/html3/_ff_descriptif_tables_image.html

Cerema Nord-Picardie, (2018), *Qualification de l'usage des zones US 235 de l'OCSGE par les Fichiers fonciers Étude - d'approfondissement et déploiement sur un périmètre élargi*, 40 p., URL : https://datafoncier.cerema.fr/sites/datafoncier/files/fichiers/2019/02/Enrichissement_OCSGE_FF_m%C3%A9thode_approfondie_0.pdf

CGDD (Commissariat général au développement durable) et Cerema (Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement), (s.d.). *Fiche transversale / L'occupation du sol : concept, usages et bases de données,* in : Applisat Communauté du satellitaire. Consulté le 20 janvier 2026. URL : https://www.applisat.fr/fiches-produits/fiche-transversale-occupation-du-sol

Chauvin M-C., (2020). *Délais de mise à jour du plan cadastra*l, in : Le Sénat. Consulté le 4 février 2026. URL : https://www.senat.fr/questions/base/2022/qSEQ220701807.html

D'Andrimont R., Yordanov M., Martinez-Sanchez L., Eiselt B., Palmieri A., Dominici P., Gallego P., Francisco J., Reuter H., Joebges C., Lemoine G. et Velde M., (2020). *Harmonised LUCAS in-situ data and photos on land cover and use from 5 tri-annual surveys in the European Union*, 38 p. DOI : 10.48550/arXiv.2005.05272.

Delattre N., (2022). *Intelligence artificielle dans le projet foncier innovant*, in : Le Sénat. Consulté le 20 janvier 2026. URL : https://www.senat.fr/questions/base/2022/qSEQ221204652.html

Demetriou D., See L. et Stillwell J., (2013). A Parcel Shape Index for Use in Land Consolidation Planning, *Transactions in GIS*, 17(6), pp. 861-882. DOI : 10.1111/j.1467-9671.2012.01371.x

Douglas D. H. et Peucker T. K., (1973). ALGORITHMS FOR THE REDUCTION OF THE NUMBER OF POINTS REQUIRED TO REPRESENT A DIGITIZED LINE OR ITS CARICATURE, *Cartographica*, 10(2), pp. 112-122. DOI : 10.3138/FM57-6770-U75U-7727

IGN (Institut National de l'Information Géographique et Forestière), (2025a). *BD FORÊT - Documentation technique.* Consulté le 4 février 2026. URL : https://geoservices.ign.fr/bdforet

IGN (Institut National de l'Information Géographique et Forestière), (2025b). *RPG - Documentation technique*. Consulté le 4 février 2026. URL : https://geoservices.ign.fr/rpg

Kalenga Tshingomba U., Jouven M., Sautot L., Shaqura I. et Teisseire M., (2020). *Cartographie des surfaces pastorales à l'aide des données Sentinel 2 L3A et des données ouvertes : Promesses et réalités*, Revue Internationale de Géomatique, 30, p. 245-277. DOI : 10.3166/rig.2021.00112

Kasser M., 2008. *Impact des évolutions technologiques actuelles sur les levers cadastraux*, in : De l'estime au cadastre en Europe. Les systèmes cadastraux aux XIXe et XXe siècles. DOI : 10.4000/books.igpde.10998

Le Roux B. et Albouy S. (de la Département de l'Aude), (2024). *STRATÉGIE DÉPARTEMENTALE POUR LA BIODIVERSITÉ DE L'AUDE 2025 - 2035*, 183 p.

Liu L., Olteanu-Raimond A-M., Jolivet L., Bris A. et See L., (2021). A data fusion-based framework to integrate multi-source VGI in an authoritative land use database, *International Journal of Digital Earth*, 14(4), DOI : 10.1080/17538947.2020.1842524

Naegelen C., (2021). *Missions de mise à jour des plans cadastraux par les géomètres du cadastre*, in : Le Sénat. Consulté le 4 février 2026. URL : https://questions.assemblee-nationale.fr/q15/15-39420QE.htm

Pérez-Hoyos A., García-Haro F-J. et San-Miguel-Ayanz J., (2012). A methodology to generate a synergetic land-cover map by fusion of different land-cover products, *International Journal of Applied Earth Observation and Geoinformation*, Volume 19, 2012, pp. 72-87, DOI : 10.1016/j.jag.2012.04.011

Piombini A., Leduc T. et Woloszyn P., (2013). *Environnement bâti et mobilité piétonne - Analyse morphométrique des espaces de visibilité et choix d'itinéraires pédestres, Colloque International Francophone Piéton COPIE 2013 - La ville sous nos pieds: connaissances et pratiques favorables aux mobilités piétonnes*, Nov 2013, Montreal, Canada. DOI : 10.13140/2.1.1801.0566

Puech T. et Mignolet C., (2022). Caractériser la morphologie des parcelles agricoles à l'échelle de la France, *Cybergeo: European Journal of Geography*. DOI : 10.4000/cybergeo.39924

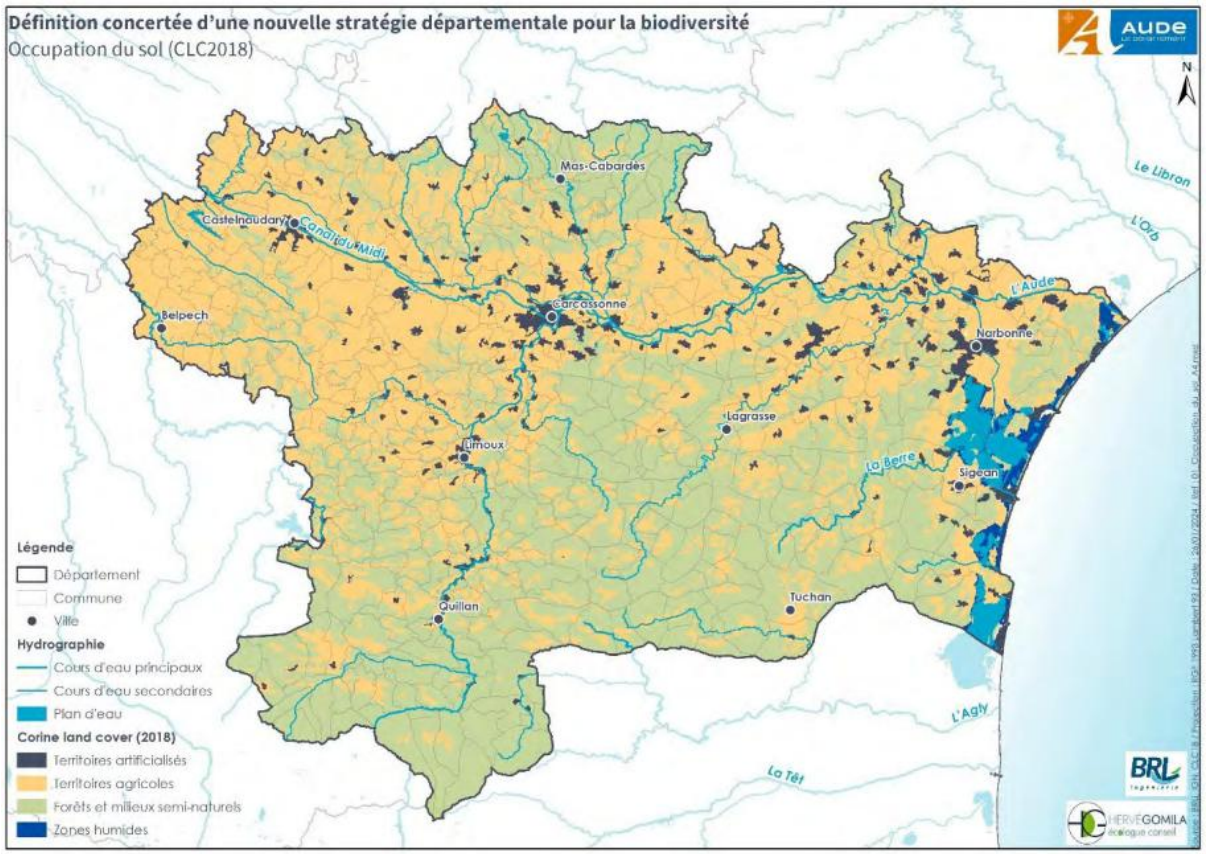
Thierry C., Lesieur-Maquin N., Fournier C., Delzons O., Gourdain P. et Hérard K., (2020). Comment cartographier l'occupation du sol en vue de modéliser les réseaux écologiques ? Méthodologie générale et cas d'étude en Ile-de-France, *Sciences Eaux & Territoires*, Articles hors-série 2020, pp. 2-8, DOI : ffmnhn-04171816f

Verburg P., Hermans K. et Nol L., (2011). Challenges in Using Land Use and Land Cover Data for Global Change Studies, *Global Change Biology*, 17, 2011, 2. 17, DOI : 10.1111/j.1365-2486.2010.02307.x

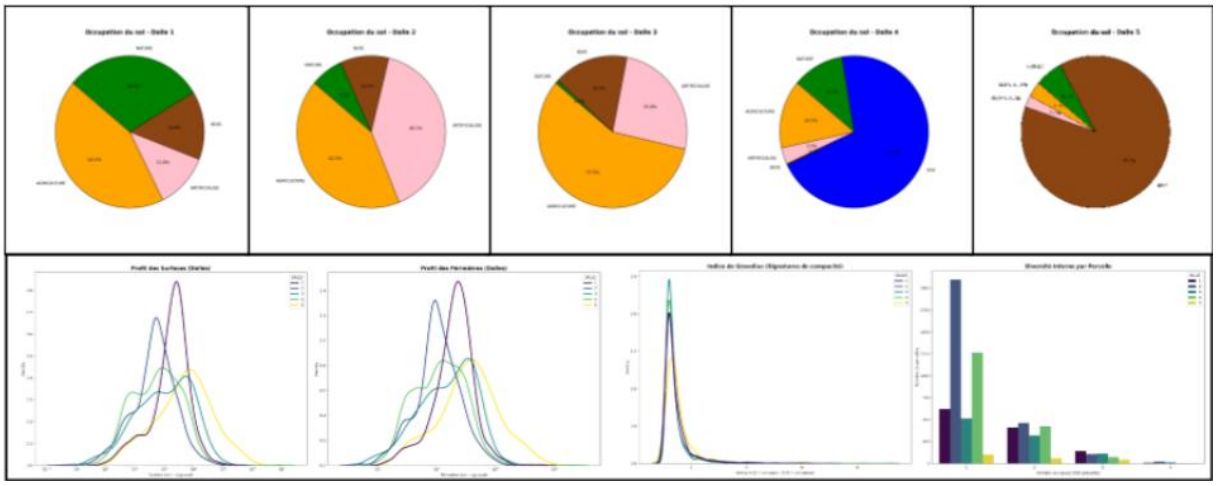
Ward J. H., (1963). Hierarchical Grouping to Optimize an Objective Function, *Journal of the American Statistical Association*, 58(301), pp. 236-244. DOI : 10.1080/01621459.1963.10500845

7. Annexe

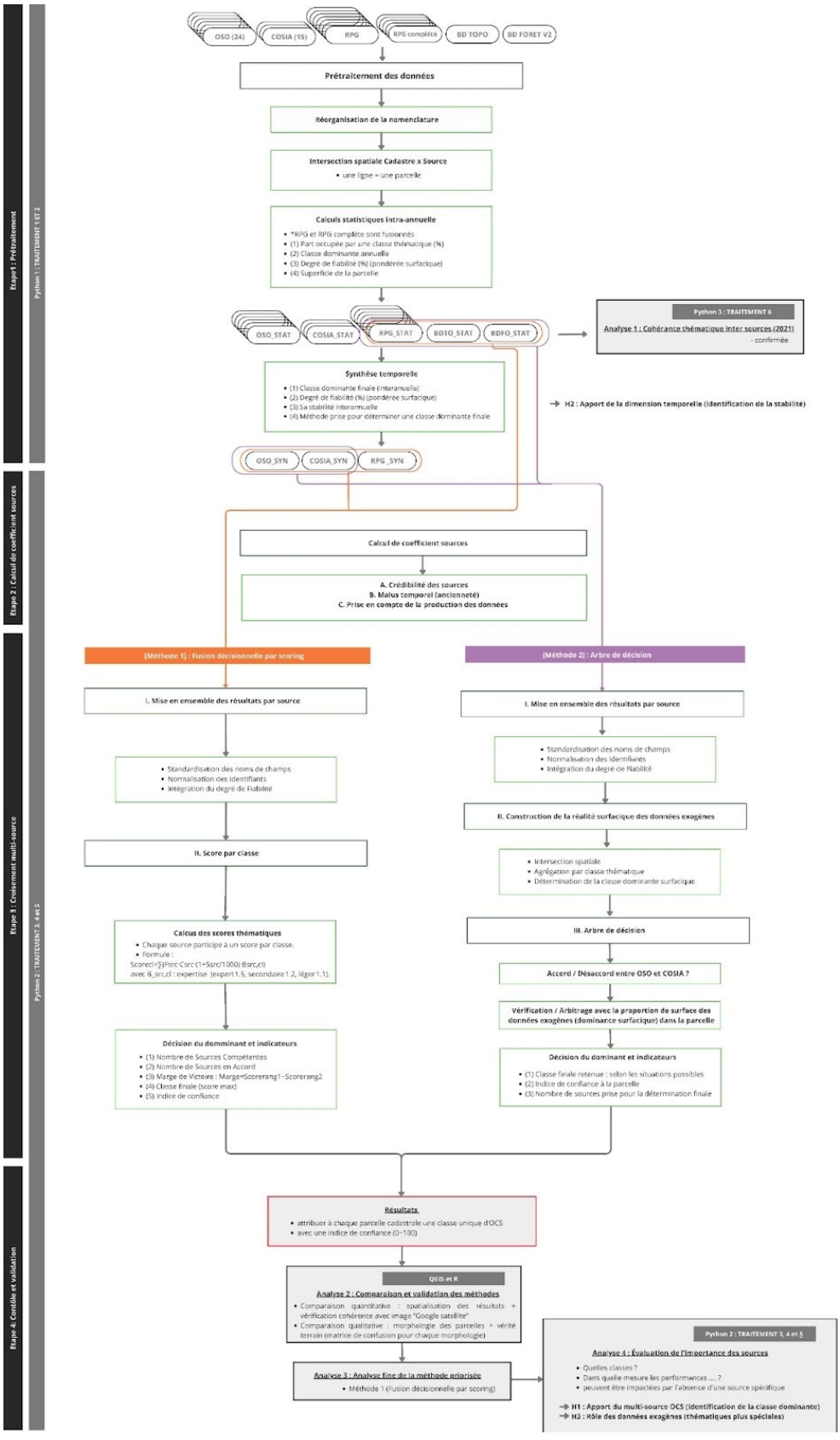
Annexe 1 : Graphiques statistiques concernant les choix des dalles (calculées sur les données OSO 2023)



Annexe 2 : Informations complémentaires sur les échantillons d'étude



Section 0 : synthèse traitement



Section 1 : Matrice nomenclature

INFORMATION DU CADSTRE	Nomenclature cadastre CEREMA (cgrnumdtxt : Groupe de culture dominant parmi les sufs (décodé))	CODE CADASTRE (cgrnumdtxt)	Nomenclature enrichie OSO	CODE OSO	Nomenclature COSIA	CODE COSIA
dcntagri : "Surface de suf de type agricole (en m²)" Nul si pas de type agricole (cgrnum=01, 02, 03 ou 04)	TERRES	1	Urbains denses	1	Batiment	1
	PRES, PRAIRIES NATURELLES, HERBAGES ET PATURAGES	2	Urbains diffus	2	Zone imperméable	3
	VERGERS, CULTURES FRUITIERES D ARBRES, ARBUSTES	3	Zones industrielles et commerciales	3	Zone perméable	2
	VIGNES	4	Surfaces routes	4	Piscine	4
dcntbois : "Surface de suf de type bois (en m²)" Nul si pas de type bois (cgrnum=05)	BOIS, AULNAIES, SAUSSAIES, OSERAIES	5	Oléagineux d'hiver	5	Serre	15
dcntnat : "Surface de suf de type naturel (en m²)" Nul si pas de type naturel (cgrnum=06 ou 07)	LANDES, PATIS, BRUYERES, MARAIS, TERRES VAINES ET VAGUES	6	Céréales à pailles	6	Sol nu	5
	CARRIERES	7	Protéagineux de printemps	7	Surface eau	6
dcnteau : "Surface de suf de type eau (en m²)" Nul si pas de type eau (cgrnum=08)	LACS, ETANGS, MARES, ABREUVOIRS, FONTAINES, CANAUX NON NAVIGABLES ET DEPENDANCES, SALINS, SALINES, MARAIS SALANTS	8	Soja	8	Neige	7
dcntarti : "Surface de suf de type artificialisé (en m²)" Nul si pas de type artifcialisé (cgrnum=07, 09, 10, 11, 12 ou 13)	JARDINS (AUTRES QUE LES JARDINS D AGREMENT)	9	Tournesol	9	Conifère	8
	TERRAINS A BATIR	10	Maïs	10	Feuillu	9
	TERRAINS D AGREMENT	11	Riz	11	Broussaille	10
	CHEMINS DE FER ET CANAUX	12	Tubercules/racines	12	Pelouse	12
	SOLS	13	Prairies	13	Culture	13
dcntsol : "Surface de suf de type sol (en m²)" Nul si pas de type sol (cgrnum=09, 10, 11, 12 ou 13) "GT national occsol : prend comme référence dcntarti qui inclus également dcnt07 (carrières)" dout le choix de l'exclure			Vergers	14	Terre labourée	14
			Vignes	15	Vigne	11
Autre (non classifiable)			Forêts de feuillus	16		
			Forêts de conifères	17		
			Pelouses	18		
			Landes	19		
			Surfaces minérales	20		
			Plages et dunes	21		
			Glaciers ou neiges	22		
			Eau	23		
			Autre	24		

Section 2 : Complément traitement par millésime

La probabilité associée à la classe dominante est calculée comme une moyenne pondérée par la surface :

$$P_{dom} = \frac{\sum_{i=1}^{n} (S_i \cdot p_i)}{\sum_{i=1}^{n} S_i}$$

Où S_i désigne la surface intersectée de chaque entité appartenant à la classe dominante, et P_i la probabilité associée fournie par la source.

La probabilité associée à la classe dominante détectée est calculée selon la nature de la source :

- Pour OSO, elle correspond à la moyenne du champ confiance pondérée par la surface intersectée.
- Pour COSIA, elle dérive des probabilités de la matrice (pour les classes), également pondérées par la surface une fois appliqué à l'entité.
- Pour le RPG Complet, la méthode de calcul est similaire à celle d'OSO, utilisant les probabilités fournies par entité.
- Pour les sources administratives (BD TOPO, BD FORÊT, RPG Standard), la confiance est fixée à 100 % pour une classe identifiée et 0 % pour la classe « Autre » donc non connue par la source.

Section 3 : Complément synthèse temporelle et stabilisation des chronologies

L'indicateur de stabilité chronologique représente le rapport entre le nombre de millésimes ayant la classe dominante finale et le nombre total de millésimes. Un indice de 100 % représente une permanence chronologique, tandis qu'un score faible signale une zone de transition ou une incertitude de classification.

$$Stab = \frac{N_{dom}}{N_{tot}}$$

Où N_{dom} est le nombre de millésimes où la classe dominante finale apparaît et N_{tot} le nombre total de millésimes disponibles pour la source.

La probabilité/fiabilité moyenne pondérée calculée uniquement sur les millésimes où la classe dominante est élue par le vote de majorité. Il correspond à la robustesse de la classe élue.

$$P_{moy} = \frac{1}{n}\sum_{i=1}^{n} p_i$$

Où P_i correspond à la probabilité / confiance du millésime i et n le nombre de millésimes où la classe dominante est élue par vote de majorité.

Section 4 : Complément sur l'audit de comparabilité

La divergence inter-sources est calculée via l'écart-type des proportions de chaque classe thématique. Une valeur faible indique un accord élevé entre les sources.

$$DIV_c = \sqrt{\frac{1}{m}\sum_{j=1}^{m}\left(p^{-}_{c}\right)^2}$$

Où c représente la classe thématique (Artificialisé, Agriculture, Nature, Bois, Eau), j la source (OSO, RPG, BD TOPO, COSIA, BD FORÊT), m le nombre de sources disponibles pour la parcelle, $P_{c,j}$ la proportion de la classe c fournie par la source j et p^{-}_{c} la moyenne des proportions de la classe c sur toutes les sources.

Un indice global de divergence est ensuite obtenu par une moyenne des divergences de chaque classe, permettant d'évaluer la cohérence globale entre les sources pour une parcelle.

$$INDICE_DIVERGENCE = \frac{1}{k}\sum_{c=1}^{k} DIV_c$$

Où k est le nombre de classes thématiques réelles (ici k = 5) étant donné que la classe Autre n'est pas incluse dans l'analyse et DIV_c est la divergence de la classe c.

Pour faciliter l'interprétation, un indicateur complémentaire est défini :

Taux d'accord précis (%) = 100 - indice global de divergence (*INDICE_DIVERGENCE*)

Section 5 : Complément coefficient source

Le coefficient est établi via un score de 100 points selon la formule ci-dessous :

$$C_{src} = max\left(10,\ min\left(100,\ \left(Base - (\Delta_t \times M_{age})\right) + M_{meth}\right)\right)$$

Où $Base$ est le score initial attribué à la source, Δ_t est la différence en années entre le dernier millésime disponible de la source et l'année de référence de l'étude (2025), M_{age} est le malus temporel appliqué annuellement pour modéliser la perte de fraîcheur de l'information et M_{meth} est un ajustement méthodologique reflétant la confiance accordée au mode de production (supervision humaine, IA, automatisation). Les valeurs sont personnelles.

La $Base$ permet d'évaluer la fiabilité des sources par rapport à la réalité du terrain :

- 95 points pour les sources vecteurs de précision élevée telles que le RPG, la BD TOPO ou la BD FORÊT (référence en précision).
- 85 points pour COSIA (bruit).
- 85 points pour OSO (lissage)

Le malus d'ancienneté $\Delta_t \times M_{age}$ ajoute une dimension d'obsolescence à la source, réduisant sa fiabilité. Un malus annuel est appliqué à partir de la date du dernier millésime :

0,50 point par an pour l'ensemble des sources disponibles, sauf pour la BD FORÊT fixée à 0,25 point par an.

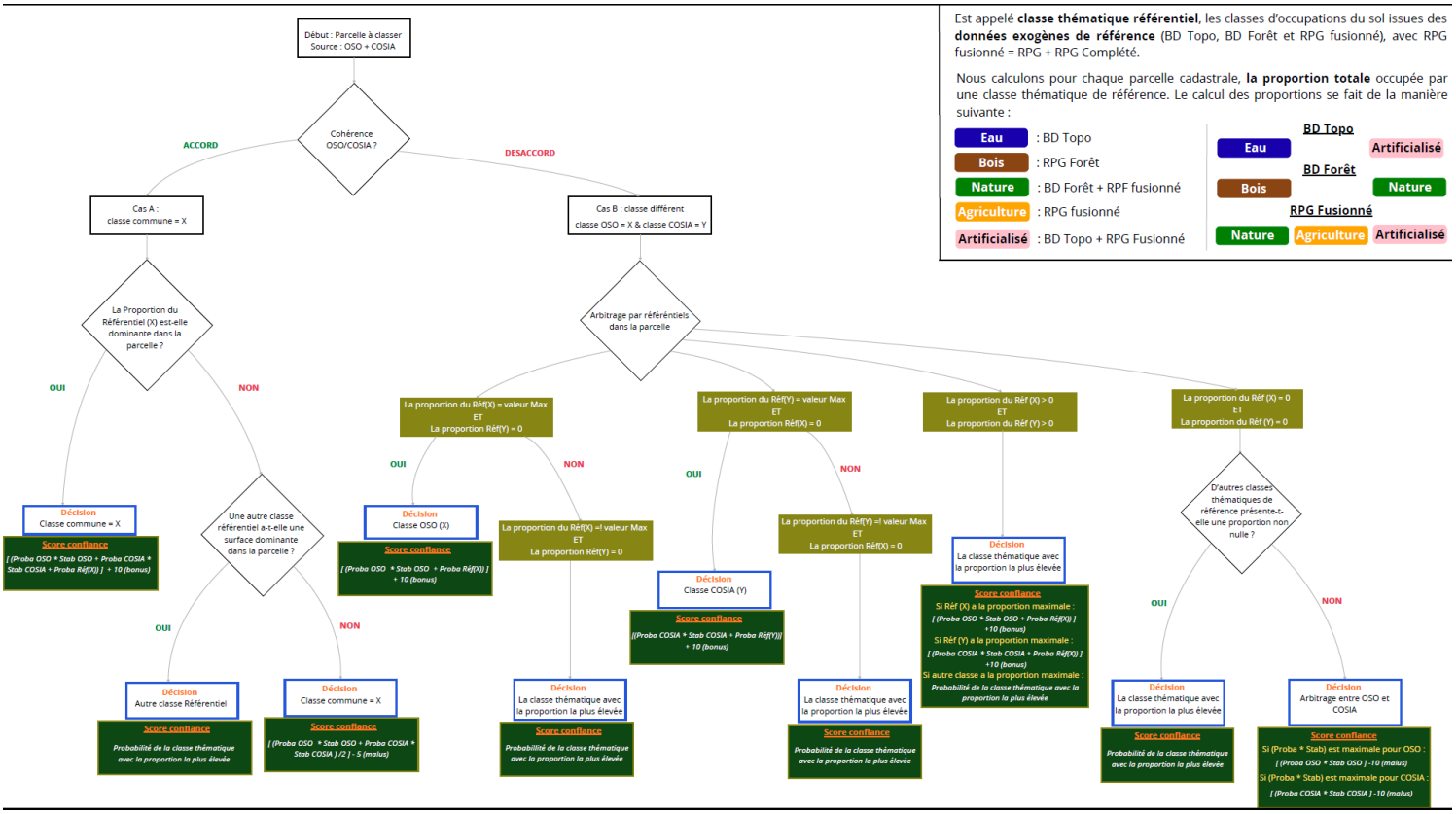
Enfin le malus/bonus lié à la méthode d'acquisition M_{meth} ajoutant une fiabilité selon le processus de classification/acquisition :

- +1 pour le RPG FUSION (Administrative + ajout de traitements par IA)
- 0 pour la BD TOPO (Une méthode de référence, humaine)
- -0,25 pour la BD FORÊT, en partie automatisée via des traitements par IA mais aussi humaine.
- -1 pour OSO et COSIA, ayant une prédominance de traitements liés à l'intelligence artificielle pour leur classification.

Section 6 : Exemple de nomenclature en sortie de la méthode 1

I. Informations préparatoires de mise en ouevre de la méthode 1												II. Résultats finaux											
Info		RPG				OSO	COSIA	BDTOPO	BDFORET	NB_SOURCES		MARGE_VICT	CL_FIN	SCORE_%						INDICE	NOM_SOURCE
PL	DALLE	CL_DOM	FIAB_DOM	STAB_DOM	COEF_SOURCE	...	...	...	...	COMPET	ACCORD			ARTI	AGRI	NAT	BOIS	EAU			
110040A0065	3	AGRI	100	100	0.955	...	...	...	...	3	3	296.27	AGRI	0	296.2729	0	0	0		73.86	RPG, OSO, COSIA
:	:	:	:	...	:	:	:	:	:	:	:	:	:	:	:	:	:	:		:	:

Section 7 : Arbre de décision (Méthode 2)



Section 8 : Caractérisation morphologique des parcelles cadastrales

