



**Production partagée d'une base de données
« Occupation des Sols » à grande échelle en
Région Grand Est (OCS GE2)**

CONTRÔLE QUALITÉ DPT 88

Version 3

David LOY – Jean-Pierre GÉRARD

TTI Production



Earth Observation Consulting Services

Decembre 2020

Table des matières

Introduction.....	4
Types de contrôles.....	5
Contrôles sur la forme.....	7
LE NOMMAGE	7
FICHIERS SHAPE FILE d'ArcGis (Contrôle automatisé)	7
ENCODAGE	7
PROJECTION	7
DIMENSION	7
SUPERPOSITION COUCHES	7
STRUCTURE DE LA TABLE ATTRIBUTAIRE	7
Types et noms de champs.....	7
Relation entre polygones et attributs	9
Contrôles topologiques	10
TOPOLOGIE DE NIVEAU 1	10
Absence d'intersection et de superposition	10
Absence de vides entre les polygones	11
Absence d'arc pendant ou de polygones ouverts	11
Contrôles géométriques.....	12
CONTRAINTES DE PRODUCTION	12
Respect des surfaces minimales.....	12
Contrôle des polygones voisins identiques	14
Contrôles sémantiques	14
REPLISSAGE DE LA TABLE ATTRIBUTAIRE	14
Contrôle des ID.....	14
Contrôle des intitulés dans les tables attributaires	14
Contrôle des surfaces de polygones.....	15
L'IMPERMÉABILISATION AU NIVEAU 5	15
Taux d'erreurs sur le Niveau 5 de la nomenclature	15
Contrôles thématiques	17
ÉCHANTILLONNAGE	17
Methodologie.....	17
Methodologie de contrôle	20
Classement des erreurs	21
RÉSULTATS.....	21
Taux d'erreurs sur le Niveau 4 de la nomenclature	21
Conclusion	23

Matrice de confusion.....	23
Analyse de la matrice de confusion.....	26
Conclusion	28

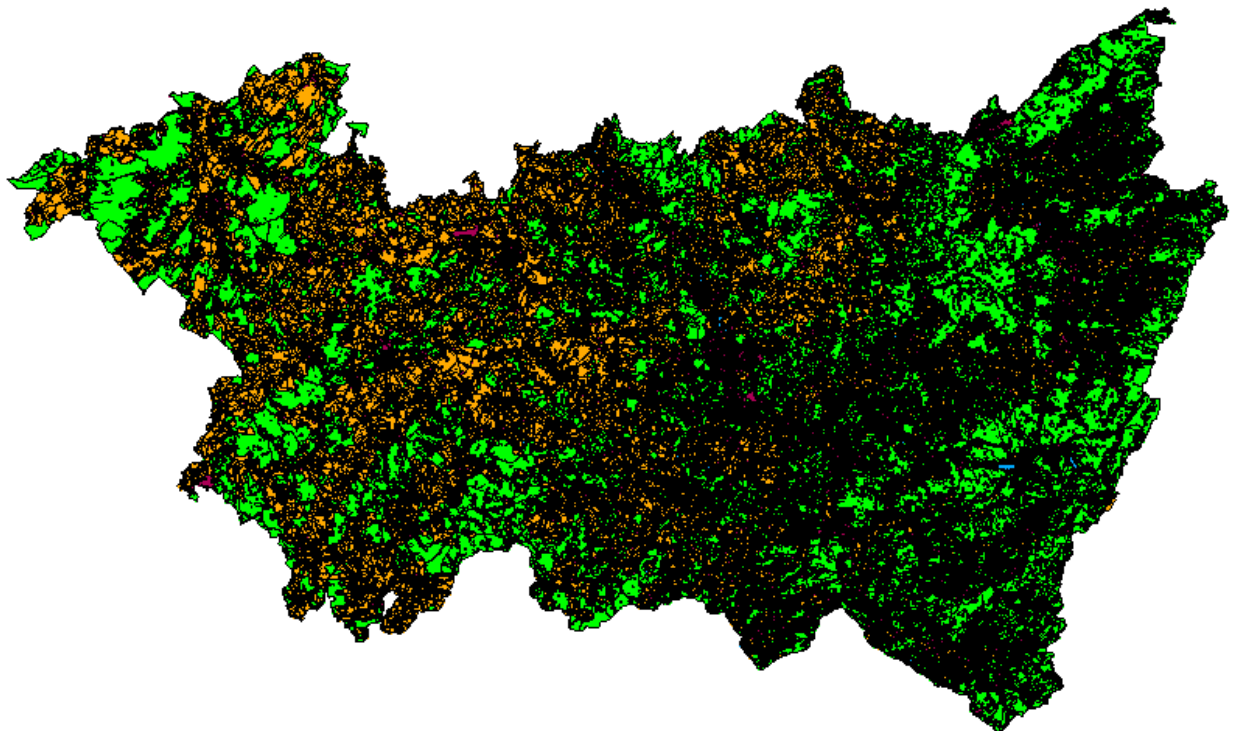
Introduction

Le phasage de production du département 88 est le suivant :

1. Production du socle (squelette + bâtis) sur le département.
2. Production de l'occupation du sol 2019 et 2010 (*)
3. Contrôle qualité CQE – Contrôle Externe du lot 3 (TTI Production) + CQC Contrôle Externe COPIL & COTECH
4. Rejet de la première version suite des % de fiabilité insuffisants
5. Contrôle qualité CQE – Contrôle Externe

(*) Pour simplifier la compréhension, les millésimes seront nommés 2019 et 2010. Néanmoins, il conviendra de noter que selon les départements, les dates diffèrent sensiblement.

Ce contrôle concerne l'intégralité du département 88.



Attention : ce rapport correspond partiellement au contrôle qualité de la seconde livraison après corrections du producteur réalisées sur ce département. Or, une troisième livraison a été faite après d'ultimes corrections. Un troisième contrôle complet n'étant pas prévu par le CCTP, seule la topologie et les classes révisées ont été recontrôlées. De ce fait, le rapport ne présente pas les bons chiffres pour le nombre de polygones des couches, mais les informations de topologie restent valables. Pour la thématique, les tableaux de statistique ont été révisés dans le rapport et dans l'Excel, mais pas dans la matrice qui est donc celle du second contrôle.

Fichiers originaux de l'OS :

OCS_GE2_D88_2019.shp

Département 88 (5953.22 km²), Millésime 2019 > 322552 polygones

OCS_GE2_D88_2010.shp

Département 88 (5953.22 km²), Millésime 2010 > 319812 polygones

Fichiers après suppression du niveau 5 de l'OS :

OCS_GE2_D88_2019_NIV4.shp

Département 88 (5953.22 km²), Millésime 2019 > 147170 polygones

OCS_GE2_D88_2010_NIV4.shp

Département 88 (5953.22 km²), Millésime 2010 > 152185 polygones

Fichiers multi-dates et d'évolution au niveau 5 de l'OS :

OCS_GE2_D88_Multi_Date_2010_2019.shp

Département 88 (5953.22 km²), Millésimes 2019 & 2010 > 357278 polygones

OCS_GE2_D88_Multi_Date_Evolution_2010_2019.shp

Département 88 (516.51 km²), Millésimes 2019 & 2010 > 53019 polygones

NOTE : le niveau 4 n'est pas un livrable prévu au CCTP. Il a été fourni par le producteur à la demande du CQE pour optimiser les contrôles.

Types de contrôles

Le contrôle qualité porte sur de nombreux points. Certains contrôles sont réalisés sur les couches vectorielles entières et d'autres sur un échantillonnage. De même, une partie des contrôles est automatisée (ou semi-automatisée lorsqu'un contrôle visuel vient vérifier les erreurs) ou intégralement visuelle.

Contrôles sur la forme (Fichiers entiers)

- LE NOMMAGE (**visuel**)
- FICHIERS SHAPE FILE d'ArcGis (**automatisés**)
- ENCODAGE (**automatisés**)
- PROJECTION (**automatisés**)
- DIMENSION (**automatisés**)
- SUPERPOSITION COUCHES (**visuel**)
- STRUCTURE DE LA TABLE ATTRIBUTAIRE
 - Types et noms de champs (**automatisés**)
 - Relation entre polygones et attributs (**automatisés**)

Contrôles topologiques (Fichiers entiers)

- TOPOLOGIE DE NIVEAU
 - Absence d'auto-intersection et de micro polygones (**automatisés**)
 - Absence de vides entre les polygones (**automatisés**)
 - Absence d'arcs ou de points dupliqués (**semi-automatisés**)
 - Absence d'arc pendant ou de polygones ouverts (**automatisés**)
- TOPOLOGIE DE NIVEAU 2
 - Absence d'auto-intersection ou de polygones ouverts (**automatisés**)

Contrôles géométriques (Fichiers entiers)

- CONTRAINTES DE PRODUCTION
 - Respect des surfaces minimales (**automatisés**)

Contrôles sémantiques (Fichiers entiers)

- REMPLISSAGE DE LA TABLE ATTRIBUTAIRE
 - Contrôle des ID (**automatisés**)
 - Contrôle des intitulés dans les tables attributaires (**automatisés**)
 - Contrôle des surfaces de polygones (**automatisés**)

Contrôles complémentaires (Fichiers entiers)

- MÉTADONNÉES (**visuel**)
- STATISTIQUES (**automatisés**)

Contrôles thématiques (Échantillon de points) - (visuel)

- ÉCHANTILLONNAGE
 - Méthodologie
 - Classement des erreurs
- RÉSULTATS SUR L'IMPERMÉABILISATION AU NIVEAU 5
 - Taux d'erreurs
 - Type d'erreurs
- RÉSULTATS SUR L'OS DE NIVEAU 4
 - Taux d'erreurs
 - Type d'erreurs
 - Matrice de confusion

La production s'est faite sur la base de la nomenclature révisée lors de la phase de terrain et présentée lors de la mission de formation. Cette nomenclature est considérée comme validée et définitive. Seuls de petits ajustements de définition sont autorisés en cas de difficultés observées.

Contrôles sur la forme

IMORTANT : le contrôle sur la forme n'est réalisé que sur le niveau 5 et les évolutions. En effet, le niveau 4 et le fichier multidates ne sont pas des livrables.

LE NOMMAGE

On constate que le nommage de fichiers est conforme à l'attente avec précision du département, du millésime, du secteur et du type de produit. À l'exception du shape-file suivant qui présente un espace en trop :

OCS_GE2_D88_Multi_Date Evolution_2010_2019.shp

Ce défaut a été corrigé

FICHIERS SHAPE FILE d'ArcGis (Contrôle automatisé)

Les fichiers ne présentent aucune anomalie. Un fichier de style pour le SIG ArcMap a été livré par le producteur.

ENCODAGE

L'encodage des fichiers est l'UTF-8, il convient donc d'importer ou d'utiliser les shape-files en respectant ce dernier.

PROJECTION

La projection des fichiers est leRGF93_Lambert_93 comme défini par le CCTP.

DIMENSION

Les fichiers sont bien en 2 D.

SUPERPOSITION COUCHES

La superposition des couches avec les BD-Ortho de référence est parfaite. On n'observe ni décalage ni distorsion.

STRUCTURE DE LA TABLE ATTRIBUTAIRE

Types et noms de champs

Un script a été réalisé afin de contrôler les noms de champs, le type (texte ou nombre) et la dimension (nombre de caractères ou nombre de chiffres après la virgule).

OCS_GE2_D88_2010.dbf

Field Name	Type	Dec	Msize	Dsize	Moffs	Dofts	Dwd
gid	Double	0	8	8	0	0	10
cod_n1	Double	0	8	8	8	8	10
lib_n1	String	0	46	46	16	16	45
cod_n2	Double	0	8	8	64	62	10
lib_n2	String	0	61	61	72	70	60
cod_n3	Double	0	8	8	136	131	10
lib_n3	String	0	53	53	144	139	52
cod_n4	Double	0	8	8	200	192	10
lib_n4	String	0	53	53	208	200	52
cod_n5	Double	0	8	8	264	253	10
lib_n5	String	0	23	23	272	261	22
surf_m	Double	15	8	8	296	284	10
surf_ha	Double	15	8	8	304	292	10
perimetre	Double	15	8	8	312	300	10
millesime	String	0	10	10	320	308	9
source	String	0	71	71	330	318	70
comment	String	0	167	167	401	389	166
doute	Double	0	8	8	568	556	10

OCS_GE2_D88_2019.dbf

Field Name	Type	Dec	Msize	Dsize	Moffs	Dofts	Dwd
gid	Double	0	8	8	0	0	10
cod_n1	Double	0	8	8	8	8	10
lib_n1	String	0	46	46	16	16	45
cod_n2	Double	0	8	8	64	62	10
lib_n2	String	0	61	61	72	70	60
cod_n3	Double	0	8	8	136	131	10
lib_n3	String	0	53	53	144	139	52
cod_n4	Double	0	8	8	200	192	10
lib_n4	String	0	53	53	208	200	52
cod_n5	Double	0	8	8	264	253	10
lib_n5	String	0	23	23	272	261	22
surf_m	Double	15	8	8	296	284	10
surf_ha	Double	15	8	8	304	292	10
perimetre	Double	15	8	8	312	300	10
millesime	String	0	10	10	320	308	9
source	String	0	71	71	330	318	70
comment	String	0	167	167	401	389	166
doute	Double	0	8	8	568	556	10

OCS_GE2_D88_Multi_Date_Evolution_2010_2019.dbf

Field Name	Type	Dec	Msize	Dsize	Moffs	Dofts	Dwd
gid	Double	0	8	10	0	1	10
cod_19_n1	Double	0	8	10	8	11	10
ocs19n1lib	String	0	46	45	16	21	45
cod_19_n2	Double	0	8	10	64	66	10
ocs19n2lib	String	0	61	60	72	76	60
cod_19_n3	Double	0	8	10	136	136	10
ocs19n3lib	String	0	53	52	144	146	52
cod_19_n4	Double	0	8	10	200	198	10
ocs19n4lib	String	0	53	52	208	208	52
cod_19_n5	Double	0	8	10	264	260	10
ocs19n5lib	String	0	23	22	272	270	22
cod_10_n1	Double	0	8	10	296	292	10
ocs10n1lib	String	0	46	45	304	302	45
cod_10_n2	Double	0	8	10	352	347	10
ocs10n2lib	String	0	61	60	360	357	60
cod_10_n3	Double	0	8	10	424	417	10
ocs10n3lib	String	0	53	52	432	427	52
cod_10_n4	Double	0	8	10	488	479	10
ocs10n4lib	String	0	53	52	496	489	52
cod_10_n5	Double	0	8	10	552	541	10
ocs10n5lib	String	0	23	22	560	551	22

surf_m	Double	15	8	24	584	573	10
surf_ha	Double	15	8	24	592	597	10
perimetre	Double	15	8	24	600	621	10
source2019	String	0	71	70	608	645	70
source2010	String	0	71	70	679	715	70
comment	String	0	161	160	750	785	160
doute	Double	0	8	10	912	945	10

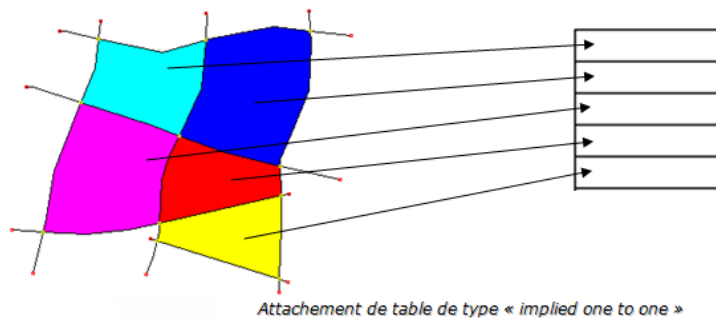
OCS_GE2_D88_Multi_Date_2010_2019.dbf

Field Name	Type	Dec	Msize	Dsize	Moffs	Doffs	Dwd
gid	Double	0	8	8	0	0	10
cod_19_n1	Double	0	8	8	8	8	10
ocs19n1lib	String	0	46	46	16	16	45
cod_19_n2	Double	0	8	8	64	62	10
ocs19n2lib	String	0	61	61	72	70	60
cod_19_n3	Double	0	8	8	136	131	10
ocs19n3lib	String	0	53	53	144	139	52
cod_19_n4	Double	0	8	8	200	192	10
ocs19n4lib	String	0	53	53	208	200	52
cod_19_n5	Double	0	8	8	264	253	10
ocs19n5lib	String	0	23	22	272	270	22
cod_10_n1	Double	0	8	10	296	292	10
ocs10n1lib	String	0	46	45	304	302	45
cod_10_n2	Double	0	8	10	352	347	10
ocs10n2lib	String	0	61	60	360	357	60
cod_10_n3	Double	0	8	10	424	417	10
ocs10n3lib	String	0	53	52	432	427	52
cod_10_n4	Double	0	8	10	488	479	10
ocs10n4lib	String	0	53	52	496	489	52
cod_10_n5	Double	0	8	10	552	541	10
ocs10n5lib	String	0	23	22	560	551	22
surf_m	Double	15	8	24	584	573	10
surf_ha	Double	15	8	24	592	597	10
perimetre	Double	15	8	24	600	621	10
source2019	String	0	71	70	608	645	70
source2010	String	0	71	70	679	715	70
comment	String	0	161	160	750	785	160
doute	Double	0	8	10	912	945	10

Relation entre polygones et attributs

Il s'agit de vérifier qu'à chaque polygone correspond un enregistrement dans la table attributaire et un seul. Ainsi :

- Il ne peut y avoir un nombre différent d'objets et d'enregistrements dans la table attributaire.
- Il ne peut pas exister des éléments sans attachement à un enregistrement dans la table attributaire.
- Il ne peut pas exister d'enregistrement dans la table attributaire sans attachement à un polygone.
- Un polygone ne peut être attaché qu'à seul enregistrement dans la table attributaire.
- Un enregistrement dans la table ne peut être attaché qu'à un seul polygone.



Attachement de table de type « implied one to one »

Après reconstruction topologique, la couche montre la présence de 54 polygones sans attributs. Ces polygones correspondent à des îles dans d'autres polygones d'évolution. Le shape file ne considère pas les îles (trous) comme des polygones à part entière, c'est donc la reconstruction qui génère ces 54 polygones et il est normal qu'ils ne soient liés à aucun attribut.

Donc, aucune anomalie n'a été constatée sur les différents fichiers livrés pour le secteur 3.

Contrôles topologiques

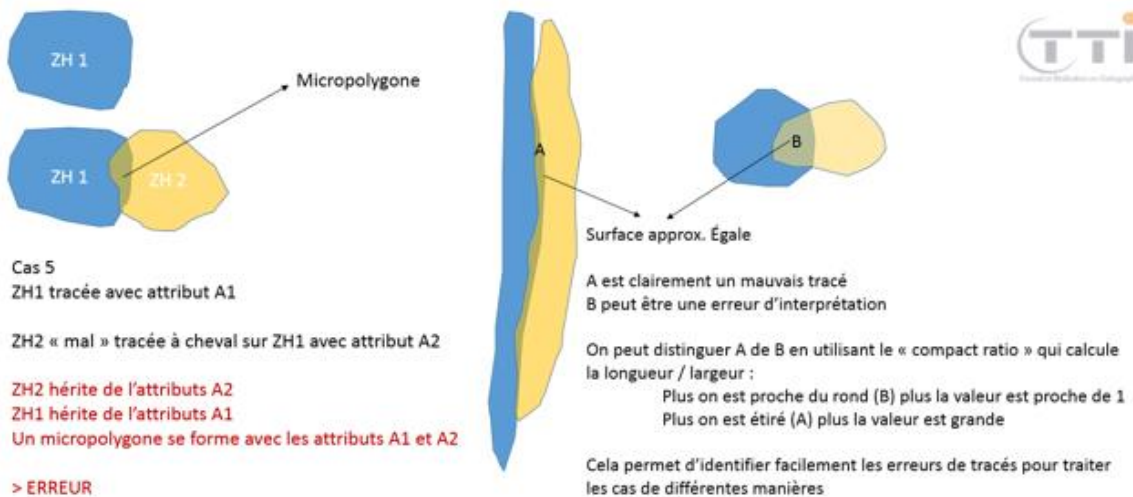
TOPOLOGIE DE NIVEAU 1

Il s'agit des contrôles élémentaires de la topologie. La structure topologique surfacique est valide lorsque :

- Aucun arc n'est dupliqué.
- Toutes les surfaces sont fermées.
- Aucun arc n'est « pendant » : un arc « pendant » est un arc dont l'une au moins de ses extrémités n'est connectée à aucun autre arc.
- Aucun polygone ne se chevauche avec un autre.
- Aucun vide ne peut être observé entre deux polygones voisins
- Deux arcs ne peuvent se croiser sans être interconnectés.
- Aucun point sur une même ligne n'est dupliqué

Absence d'intersection et de superposition

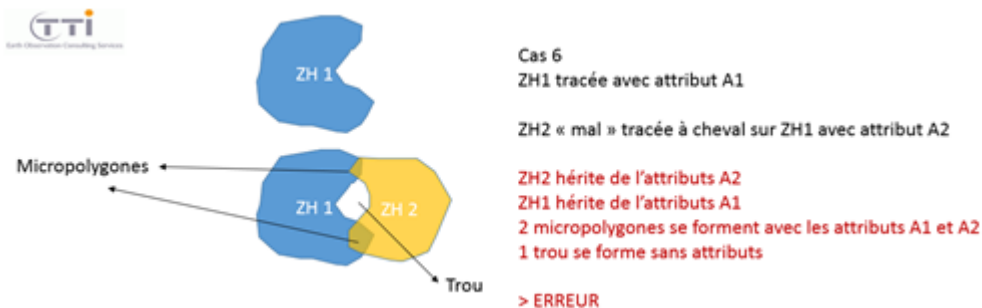
Les erreurs topologiques ne sont pas liées qu'à la qualité du calage ou du tracé entre les polygones. Il s'agit également de superposition de polygones qui se cachent les uns les autres, mais qui produisent des intersections lorsque l'on reconstruit la topologie. Ces superpositions forment de nouveaux polygones à la reconstruction de la topologie qui sont reliés à plusieurs éléments dans la table attributaire (autant que de recouvrements) et il est alors très facile de les identifier par requête.



Aucun micro-polygone n'a été repéré sur les couches livrées.

Absence de vides entre les polygones

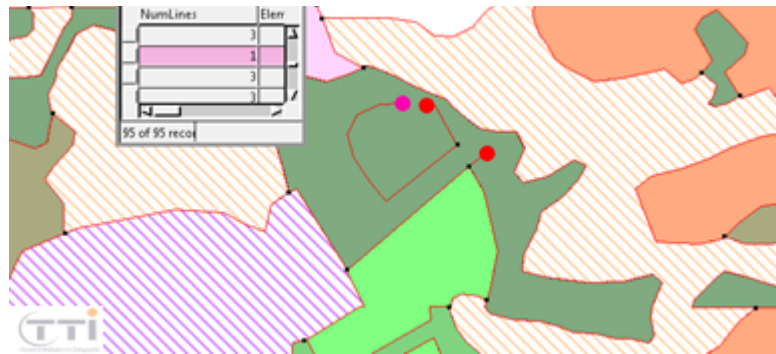
Comme pour les superpositions, des espaces entre les polygones normalement jointifs peuvent générer des micro-polygones à la reconstruction topologique, Ces vides forment de nouveaux polygones qui ne sont reliés à aucun élément dans la table attributaire et il est alors très facile de les identifier par requête.



Aucun espace vide entre polygones n'a été repéré sur les fichiers livrés.

Absence d'arc pendant ou de polygones ouverts

Les polygones non fermés produisent ce que l'on appelle des arcs pendants. Il s'agit de lignes dont l'un des nœuds (initial ou final) n'est relié qu'à elles-mêmes alors que dans une topologie propre, ils devraient être reliés à au moins trois lignes. Ainsi, en SIG le repérage des arcs pendants est très simple par une requête sur le nombre de lignes partant d'un nœud.



Exemple d'arc pendant et de polygone ouvert.

Aucun arc pendant ou polygone ouvert n'a été repéré sur les couches livrées.

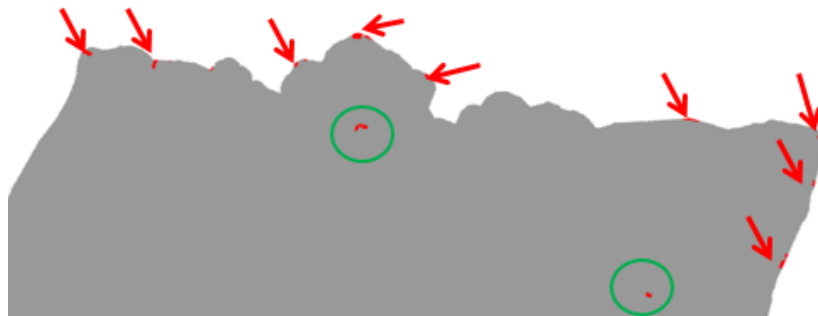
Contrôles géométriques

CONTRAINTES DE PRODUCTION

Respect des surfaces minimales

L'**unité minimale de collecte** (UMC) correspond à la plus petite unité spatiale cartographiée. Normalement, aucune parcelle ne doit être de surface inférieure à celle-ci et dans le même temps, un objet supérieur à cette surface doit être délimité. Dans la pratique, ce n'est pas toujours vrai. Le photo-interprète est parfois obligé de « tricher » afin de prendre en compte une entité importante (notamment dans les postes urbains) plus petite que l'UMC ou la continuité des réseaux.

Le respect des UMC prend en compte deux paramètres : la surface et la nomenclature. Ce repérage des polygones inférieurs à la surface de l'UMC en fonction de son code de nomenclature est fait sur SIG via une requête automatisée.



De plus, le découpage lié aux bordures de la zone de travail ne suit pas la réalité de l'occupation du sol. C'est une limite administrative. De fait, la délimitation de l'occupation du sol doit s'arrêter sur cette limite et implicitement, cela peut engendrer de petits polygones de bordure inférieurs à l'UMC. Ces polygones ne sont évidemment pas à prendre en compte.

Observation des tests automatisés

UMC OCS_GE2_D88_ZONE2_2010_NIV4,

Pas significatif, car continuité réseau ferré :

Continuité réseaux > Polygone 77699, surface 250.9887 < 500 - UMC niv4 invalide pour le code 1411

Continuité réseaux > Polygone 16135, surface 376.1935 < 500 - UMC niv4 invalide pour le code 1411

Continuité réseaux > Polygone 30940, surface 231.4810 < 500 - UMC niv4 invalide pour le code 1411

Continuité réseaux > Polygone 75051, surface 424.6383 < 500 - UMC niv4 invalide pour le code 1411

Pas significatif, car continuité réseau hydrographique :

Continuité réseaux > Polygone 115293, surface 368.5503 < 500 - UMC niv4 invalide pour le code 5110

Continuité réseaux > Polygone 5177, surface 447.0848 < 500 - UMC niv4 invalide pour le code 5110

Continuité réseaux > Polygone 46870, surface 329.7292 < 500 - UMC niv4 invalide pour le code 5110

Continuité réseaux > Polygone 88738, surface 395.3578 < 500 - UMC niv4 invalide pour le code 5110

Pas d'erreur d'UMC repérées

UMC OCS_GE2_D88_ZONE2_2019_NIV4

Pas significatif, car continuité réseau ferré :

Continuité réseaux > Polygone 75527, surface 250.9887 < 500 - UMC niv4 invalide pour le code 1411

Continuité réseaux > Polygone 72941, surface 424.6383 < 500 - UMC niv4 invalide pour le code 1411

Continuité réseaux > Polygone 15763, surface 376.1935 < 500 - UMC niv4 invalide pour le code 1411

Continuité réseaux > Polygone 30200, surface 231.4810 < 500 - UMC niv4 invalide pour le code 1411

Pas significatif, car continuité réseau hydrographique :

Continuité réseaux > Polygone 111841, surface 368.5503 < 500 - UMC niv4 invalide pour le code 5110

Continuité réseaux > Polygone 5005, surface 447.0848 < 500 - UMC niv4 invalide pour le code 5110

Continuité réseaux > Polygone 86103, surface 395.3578 < 500 - UMC niv4 invalide pour le code 5110

Continuité réseaux > Polygone 45634, surface 329.7292 < 500 - UMC niv4 invalide pour le code 5110

Pas d'erreur d'UMC repérées

UMC OCS_GE2_D88_ZONE2_2010

Pas significatif, car continuité réseau ferré :

Continuité réseaux > Polygone 213928, surface 231.4810 < 250 - UMC niv5 invalide pour le code 14113

Pas d'erreur d'UMC repérées

UMC OCS_GE2_D88_ZONE2_2019

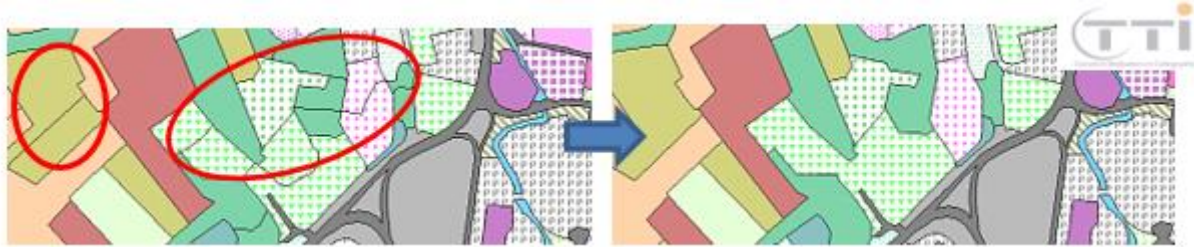
Pas significatif, car continuité réseau ferré :

Continuité réseaux > Polygone 221961, surface 231.4810 < 250 - UMC niv5 invalide pour le code 14113

Pas d'erreur d'UMC repérée

Contrôle des polygones voisins identiques

Ce contrôle automatique est effectué sur SIG. La routine permet de prendre en compte plusieurs tables et plusieurs champs pour détecter les lignes dont les attributs du polygone sélectionné sont identiques de chaque côté de celle-ci.



La couche de niveau 5 présente plusieurs polygones voisins de code identique. Cela s'explique par la présence des champs « comment » et « indice » (doute). Si deux polygones voisins ont la même OS, mais que sur l'un d'eux un doute a été émis, la limite est préservée. L'OS est alors identique à droite et à gauche de la limite commune.

La fusion des polygones de niveau 5 pour donner la couche de niveau 4 a utilisé les champs cod_n4, mais aussi les champs « comment » et « indice » (doute). De ce fait, il persiste également au niveau 4 quelques polygones voisins ayant le même code.

Aucune erreur de codes voisins identiques hors présence de doutes ou commentaires n'a été relevée, même sur les versions simplifiées au niveau 4.

Contrôles sémantiques

REPLISSAGE DE LA TABLE ATTRIBUTAIRE

Contrôle des ID

Le numéro maximal doit correspondre au nombre de polygones de la couche. Par ailleurs, il ne doit pas exister deux ID avec le même nombre. Cette validation est automatisée.

Les ID sont conformes aux exigences.

Contrôle des intitulés dans les tables attributaires

Il s'agit de contrôler que le remplissage de la table attributaire et plus particulièrement celui des intitulés de classe est conforme à la nomenclature. Néanmoins, il s'agit également de contrôler qu'il n'y a pas de codes (niveaux 1 à 5) présents qui n'existent pas dans la nomenclature.

Pour ce contrôle, un processus automatisé compare les valeurs ou les intitulés des tables attributaires avec ceux de la nomenclature et signale toute anomalie. Les anomalies signalées sont :

- Incohérence des codes de niveaux 1 à 5 au sein d'un même polygone
- Présence d'un code qui n'existe pas dans l'un des champs
- Absence de code dans l'un des champs
- Absence d'intitulé dans l'un des champs
- Présence d'un intitulé qui n'existe pas dans l'un des champs
- Présence d'un intitulé non conforme (exemple au pluriel alors qu'il est au singulier dans la nomenclature).
- Incohérence entre les codes et les intitulés correspondants

Pas d'incohérences observées entre les codes et les intitulés.

Contrôle des surfaces de polygones

Les champs contenant les surfaces en m² et ha ont été recalculés et comparés à ceux de la couche. Généralement, si une erreur est détectée, cela traduit un oubli de mise à jour du champ. Il sera alors demandé au prestataire de corriger cela. Le calcul étant lié au logiciel, il est normal de trouver de petites variations.

Nous avons donc ignoré les différences de +/- 5m² et les écarts jusqu'à 25 m² pour les très grands polygones (> 8 km²) et/ ou avec de nombreuses îles (> 40) ou très complexes du réseau routier.

Pas d'erreurs significatives observées dans les calculs de surfaces des autres couches.

L'IMPERMÉABILISATION AU NIVEAU 5

Le producteur procède à des requalifications de certains éléments du socle et à une resegmentation du niveau 4 sur les postes urbains par photo-interprétation.

Pour le contrôle, il s'agit donc de contrôler par PIAO un échantillon de points. Le contrôle, bien que plus simple, s'apparente alors au contrôle thématique.

Il est à noter que les zones imperméabilisées et bâties sont l'héritage de l'intégration du bâti dans le socle. Il ne s'agit donc pas de contrôler de façon exhaustive cette classe « imperméable bâti », mais plutôt d'essayer de relever des erreurs de la BD-Topo : bâti indiqué, mais n'existant plus ou pas encore construit. Ce contrôle est encore plus important pour le millésime 2010, car il est produit à partir du socle 2019 et il faut donc enlever les bâtis inexistantes.

Échantillonnage :

Un échantillon aléatoire a été fait sur le département. Il n'est pas nécessaire de contrôler un nombre de points important pour estimer la qualité, car cette information a déjà été largement contrôlée sur les trois subdivisions du département en cours de production et que les résultats ont toujours été en dessous du taux d'erreur maximal attendu. 550 points ont été contrôlés sur les millésimes 2019 et 2010.

Taux d'erreurs sur le Niveau 5 de la nomenclature

Il est important de noter que ce contrôle ne s'attache qu'à la (im) perméabilité de l'occupation du sol. La nature du niveau 4 n'est pas vérifiée durant ce contrôle, néanmoins, quelques points ont été relevés à titre indicatif.

Secteur2	Échantillon	Refusé	Redécouper	Toléré	Doute
2019	400	13	7	7	8
2010	150	9	8	2	5
TOTAL	550	22	15	9	13
		4.0 %	2.72 %	1.63 %	2.36 %

Refusé : il y a erreur sur le type de (im) perméabilité

Redécouper : on peut isoler des zones ayant un autre type de (im) perméabilité

Toléré : le taux de imperméabilité/perméabilité est aux alentours de 50/50

Doute : le CQE n'arrive pas à déterminer le type

L'analyse des erreurs montre surtout qu'il s'agit essentiellement d'erreurs sur la perméabilité des classes difficiles : zones en mutation, exploitations agricoles, etc. du fait de la nature du sol parfois difficile à estimer. Il faut d'ailleurs noter que le CQE a rencontré un nombre non négligeable de doutes.

NOTE : aucun contrôle sur StreetView n'a été fait.

Conclusion :

Le taux d'erreur (refusé + à redécouper) sur le volet imperméabilisation est de **6.72 %** ce qui est au-delà de la fiabilité attendue. Le résultat est donc excellent.

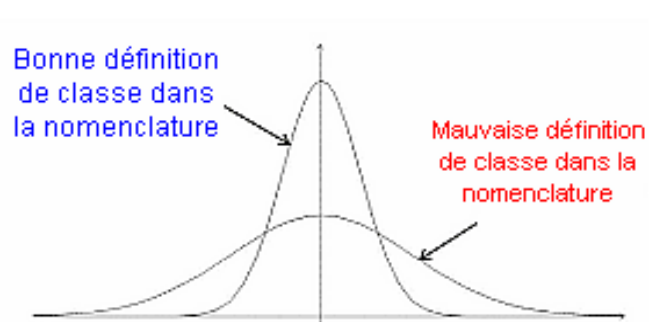
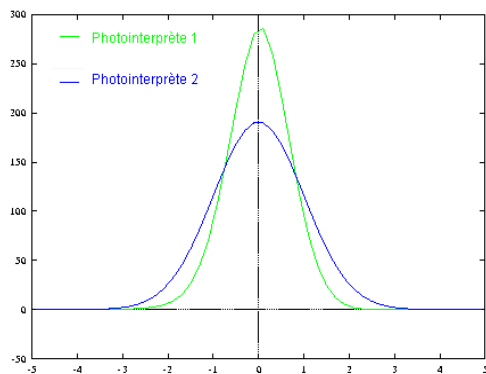
Contrôles thématiques

ÉCHANTILLONNAGE

Le contrôle de la qualité thématique vise à identifier le taux d'erreurs, mais aussi à tenter d'en trouver l'origine.

Origine des erreurs :

- Les clartés des définitions de classes dans la nomenclature
- L'expérience du photo-interprète ou sa compréhension de la nomenclature
- L'erreur de saisie du code
- L'erreur de segmentation
- L'erreur propagée d'une mauvaise donnée exogène
- L'erreur induite de la qualité des ortho-photos
- L'erreur d'omission par manque de donnée exogène
- Le contrôleur peut se tromper en validant ou refusant un polygone à tort



Exemple de la répercussion du taux de réussite en fonction à gauche, des photo-interprètes, à droite de la définition de la nomenclature

À cela s'ajoutent deux paramètres :

- Certaines classes sont moins représentées que d'autres
- Le tirage doit être représentatif de toute la zone et l'échantillonnage doit donc être correctement réparti sur toute la zone.

L'échantillonnage des polygones est donc fait suivant une méthode probabiliste, il s'agit d'un tirage aléatoire en grappes et stratifié. C'est-à-dire un tirage aléatoire orienté qui prend en compte à la fois l'occurrence des classes et la répartition.

Méthodologie

Il s'agit donc de tirer aléatoirement des polygones en respectant ces **4 règles** :

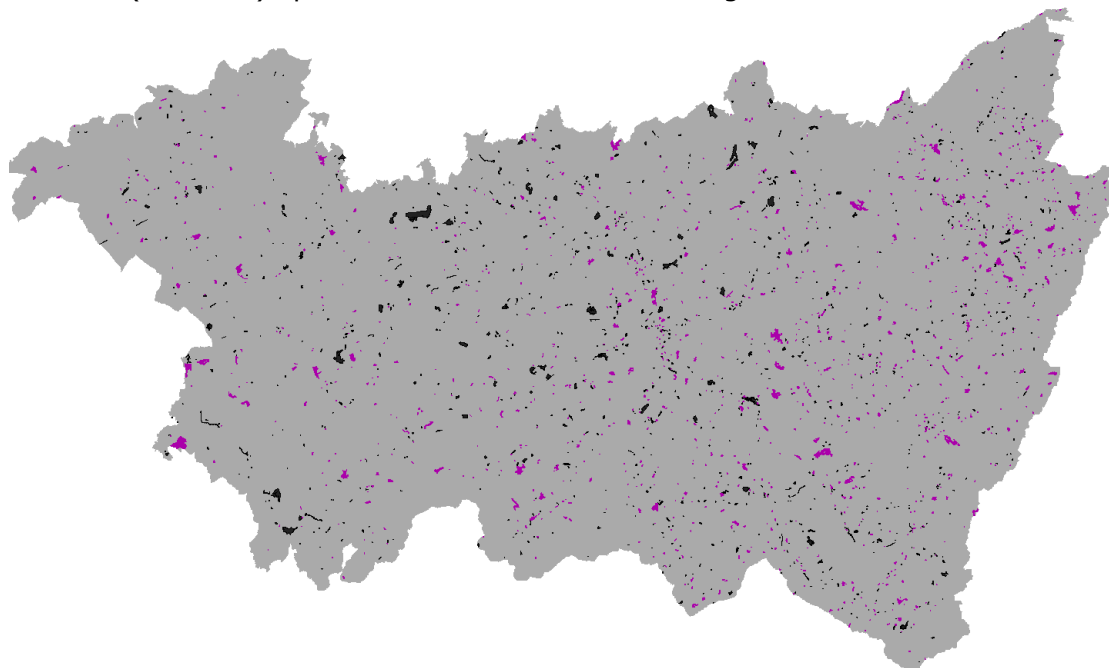
- Inventorier avec au minimum 1000 polygones.
- La répartition par classe doit correspondre à l'occurrence réelle de la classe dans la couche/nombre total de polygones.
- Pour les postes présentant moins de 3 polygones, l'intégralité des polygones de la classe est vérifiée.
- Répartition équitable dans le secteur analysé.

Le socle est contrôlé à part, et donc l'échantillonnage exclut les réseaux routiers et ferrés.

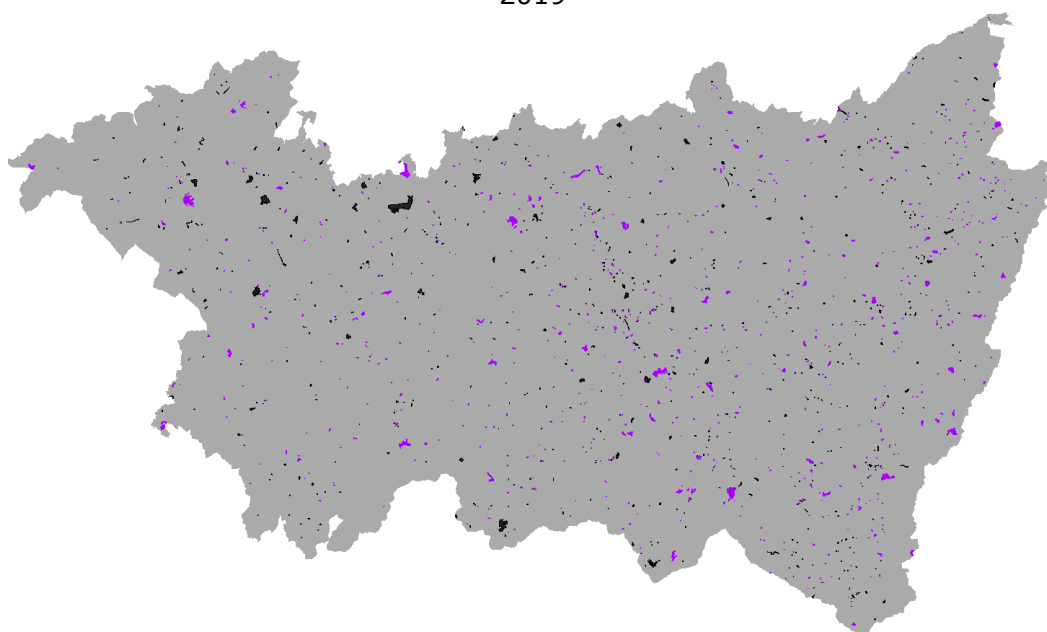
Total 2019	2500 polygones
Total 2010	1500 polygones

Une fois l'échantillonnage effectué, un script extrait automatiquement les polygones 2019 concernés. Les polygones 2010 sont ensuite traités de la même manière. L'occupation du sol 2010 étant majoritairement héritée de 2019 ; l'échantillon est plus faible.

Ci-dessous, sont présentés en orange les polygones contrôlés, sans limites de taille. En noir l'échantillonnage du premier contrôle auquel s'ajoute l'échantillonnage du second contrôle (en violet) après livraison de la version corrigée.



2019



2010

OCCURRENCE		
typo_niv4	2019	2010
Bâti continu dense	570	571
Bâti continu aéré	2067	2052
Bâti collectif	753	742
Bâti mixte	13	12
Bâti individuel dense	7857	7763
Bâti individuel lâche	19374	19571
Bâti isolé en zone agricole ou naturelle	5588	5544
Espaces libres en milieu urbain	1005	1233
Emprises scolaires et universitaires	452	443
Emprises hospitalières	184	172
Équipements sportifs et de loisirs ; campings	1199	1171
Cimetières	413	415
Autres équipements collectifs	1349	1324
Équipements eau, énergies, T.I.C. et déchets	641	548
Emprises d'activités à dominante industrielle	626	639
Emprises d'activités à dominante commerciale	585	547
Emprises d'activité à dominante mixte ou tertiaire	1644	1618
Anciennes emprises d'activité	73	26
Emprises militaires	70	72
Exploitations agricoles	3034	2948
Zones d'extraction	95	82
Emprise réseau ferré	300	298
Emprise réseau routier	97	104
Espaces associés aux réseaux routiers et ferrés	3752	3667
Emprises aéroportuaires	5	5
Emprises portuaires	4	4
Espaces verts urbains	1052	1062
Espaces en transition	1699	2010
Places	34	34
Cultures annuelles et pluri-annuelles	6028	6389
Cultures spécifiques	401	350
Vignes	252	270
Vergers traditionnels	3887	3951
Vergers intensifs	115	113
Pépinières	28	28
Prairies, friches et délaissés agricoles	20937	21049
Bosquets et haies	9259	9733
Forêts de feuillus	10905	10981
Forêts de conifères	7038	7510
Forêts mixtes	8793	9220
Coupes à blanc et jeunes plantations	7680	6634
Peupleraies et sapinières	277	245
Pelouses et pâturages de montagne	466	489
Formations pré-forestières	5491	8027
Surfaces enherbées semi-naturelles	3463	5043
Plages et sables	51	31
Roches nues	147	148
Zones de sinistre (incendie, tempête)	139	64
Ripisylves et rivulaires	3000	3012
Autres milieux humides	364	393
Cours d'eau et canaux	448	450
Plans d'eau	3252	3222
Bassins artificiels	245	188

Méthodologie de contrôle

Le CQE procède au contrôle en passant en revue à l'échelle de travail du producteur les polygones sélectionnés. Il affiche également les différentes images (couleur naturelle ou IRC) et les données exogènes disponibles. Dans un premier temps, les scripts automatisés remplissent les champs concernant les codes, les intitulés, les UMC. Puis le photo-interprète du CQE complète les observations.

COD	UMC	LMC	DOUTE	GEO	Texte_GEO	THE	Texte_THE
<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	7	LISTE PB : Mauvaise limite avc voisinage; - REFUSE		
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	7	LISTE PB : Mauvaise limite avc voisinage; - REFUSE		
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	7	LISTE PB : Mauvaise limite avc voisinage; - REFUSE		
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	7	LISTE PB : Mauvaise limite avc voisinage; - REFUSE		
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	7	LISTE PB : Mauvaise limite avc voisinage; - REFUSE		
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	7	LISTE PB : Mauvaise limite avc voisinage; - REFUSE		

COD : si coché, erreur sur un code ou un intitulé. Ce champ est rempli par script.

UMC : si coché, erreur sur l'UMC. Ce champ est rempli par script, mais un contrôle visuel le complète pour les cas particuliers.

LMC : si coché, erreur sur la LMC. Ce contrôle se fait visuellement lors de l'affichage du polygone.

GEO : code(s) d'erreur(s) géométrique(s). Il peut y en avoir plusieurs pour un même polygone.

Texte_GEO : C'est l'intitulé correspondant au code(s) GEO et dans le même temps la qualification « REFUSE » si non acceptable.

THE : code(s) d'erreur(s) thématique(s). Il peut y en avoir plusieurs pour un même polygone.

Texte_THE : C'est l'intitulé correspondant au code(s) THE et dans le même temps la qualification « REFUSE » si non acceptable.

Note : le refus thématique ou géométrique est cumulé. Si l'un des deux est « REFUSE », le polygone est refusé.

CODE	CODE2	RMQ	EXO	cod_n4	lib_n4
ok	...	limite N arbitraire	Foret IFN; Photo IR	3130	Forêts mixtes
ok	...	limite à refaire à gauche		2310	Surfaces enherbées, friches e
ok	...	prairie à côté au NORD > limite à modifier		2310	Surfaces enherbées, friches e
ok	ok	prairie à droite		2110	Cultures annuelles et pluri-a
ok	...	le polygone d'à côté devrait l'être aussi		1610	Espaces en transition
ok	...	il manque les résineux à gauche		3120	Forêts de conifères
ok	...	la limite ne devrait pas se trouver là mais englober toute la parcelle		2110	Cultures annuelles et pluri-a
Peupleraies et sapinières	Formations pré-forestières	la partie N est en préforestier comme à côté. La limite est arbitraire		3150	Peupleraies et sapinières
nk	...	involonne à retracer		2320	Rosquets et haies

CODE 1 : code proposé pour le remplacement ou la re-segmentation

CODE 2 : code proposé pour le remplacement ou la re-segmentation ou autre alternative.

RMQ : commentaire libre du CQE pour indiquer la nature de l'erreur ou indiquer qu'il n'a pas trouvé de source de donnée exogène pour justifier un classement

EXO : donnée exogène utilisée pour contrôler le polygone et éventuellement le recours à google street.

Classement des erreurs

Le CQE contrôle par photo-interprétation la pertinence et identifie les problèmes :

- de délimitations (tracé imprécis, polygone pouvant être subdivisé ou au contraire agrégé)
- de confusions thématiques
- d'erreurs de saisie

Liste des erreurs GÉOMÉTRIQUES :

1 Couloir inférieur à LMC	> REFUSE
2 Précision du tracé	> REFUSE
3 Subdivision possible	> REFUSE
4 Cas limite de subdivision	
5 La limite est celle de la Réf forêt, non ce qui visible	> REFUSE
6 Rétrécissement localisé	
7 Mauvaise limite avec voisinage	> REFUSE
8 Cas limite de précision du tracé	
9 Cas limite de délimitation	

Liste des erreurs THÉMATIQUES :

1 Erreur d'identification	> REFUSE
2 Cas limite acceptable ;	
3 Cas éventuellement à discuter	
4 Doute du CQE sur nature ou classement	
5 Autre possibilité de classement	
6 Cas logique = repenser le découpage	
7 Commentaire libre	
8 Regroupement avec polygone voisin	> REFUSE

RÉSULTATS

Pour mémoire, le CCTP demande un taux de fiabilité de 80 % par classe et 90 % au total.

Taux d'erreurs sur le Niveau 4 de la nomenclature

typo_niv4	2019			2010			TOTAL		
	ech.	refusés	pct	ech.	refusés	pct	ech.	refusés	pct
Bâti continu dense	12	1	91.67	10	1	90.00	22	2	90.91
Bâti continu aéré	40	6	85.00	24	3	87.50	64	9	85.94
Bâti collectif	16	1	93.75	10	0	100.00	26	1	96.15
Bâti mixte	10	1.5	85.00	10	1	90.00	20	2.5	87.50
Bâti individuel dense	149	32	78.52	65	12.5	80.77	214	44.5	79.21
Bâti individuel lâche	142	21	85.21	68	12.5	81.62	210	33.5	84.05
Bâti isolé en zone agricole ou naturelle	103	5	95.15	55	3.5	93.64	158	8.5	94.62
Espaces libres en milieu urbain	22	2	90.91	16	3	81.25	38	5	86.84
Emprises scolaires et universitaires	15	2	86.67	15	2	86.67	30	4	86.67
Emprises hospitalières	17	1	94.12	13	1	92.31	30	2	93.33

	2019			2010			TOTAL		
Équipements sportifs et de loisirs, campings	25	4	84.00	17	1.5	91.18	42	5.5	86.90
Cimetières	15	0	100.00	16	2	87.50	31	2	93.55
Autres équipements collectifs	28	1	96.43	16	1	93.75	44	2	95.45
Équipements eau, énergies, T.I.C. et déchets	14	2	85.71	16	2	87.50	30	4	86.67
Emprises d'activités à dominante industrielle	14	0	100.00	18	3	83.33	32	3	90.63
Emprises d'activités à dominante commerciale	11	1	90.91	14	2.5	82.14	25	3.5	86.00
Emprises d'activité à dominante mixte ou tertiaire	36	7	80.56	24	4.5	81.25	60	11.5	80.83
Anciennes emprises d'activité	13	1.5	88.46	10	2	80.00	23	3.5	84.78
Emprises militaires	13	0	100.00	10	1	90.00	23	1	95.65
Exploitations agricoles	60	2	96.67	36	6.5	81.94	96	8.5	91.15
Zones d'extraction	18	5	72.22	14	1.5	89.29	32	6.5	79.69
Emprise réseau ferré									
Emprise réseau routier				10	1.5	85.00	10	1.5	85.00
Espaces associés aux réseaux routiers et ferrés	70	5	92.86	42	5.5	86.90	112	10.5	90.63
Emprises aéroportuaires	6	0	100.00	6	0	100.00	12	0	100.00
Emprises portuaires	4	0	100.00	4	0	100.00	8	0	100.00
Espaces verts urbains	33	9.5	71.21	19	5.5	71.05	52	15	71.15
Espaces en transition	37	3	91.89	24	3	87.50	61	6	90.16
Places	13	1	92.31	12	0	100.00	25	1	96.00
Cultures annuelles et pluri-annuelles	127	3	97.64	61	5.5	90.98	188	8.5	95.48
Cultures spécifiques	24	4.5	81.25	21	2.5	88.10	45	7	84.44
Vignes	22	1.5	93.18	21	1	95.24	43	2.5	94.19
Vergers traditionnels	71	7.5	89.44	44	7.5	82.95	115	15	86.96
Vergers intensifs	10	1	90.00	24	2	91.67	34	3	91.18
Pépinières	10	0	100.00	10	0	100.00	20	0	100.00
Prairies, friches et délaissés agricoles	144	23	84.03	79	17	78.48	223	40	82.06
Bosquets et haies	144	20.5	85.76	74	14	81.08	218	34.5	84.17
Forêts de feuillus	150	18.5	87.67	78	15.3	80.38	228	33.8	85.18
Forêts de conifères	130	10.5	91.92	72	6	91.67	202	16.5	91.83
Forêts mixtes	150	26.5	82.33	79	15	81.01	229	41.5	81.88
Coupes à blanc et jeunes plantations	147	15	89.80	60	10	83.33	207	25	87.92
Peupleraies et sapinières	13	2	84.62	19	2	89.47	32	4	87.50
Pelouses et pâturages de montagne	11	2	81.82	17	3.5	79.41	28	5.5	80.36
Formations pré-forestières	108	11.5	89.35	63	9.5	84.92	171	21	87.72
Surfaces enherbées semi-naturelles	74	11.5	84.46	56	6	89.29	130	17.5	86.54
Plages et sables	10	1	90.00	10	2	80.00	20	3	85.00
Roches nues	11	0	100.00	13	1	92.31	24	1	95.83
Zones de sinistre (incendie, tempête)	10	0	100.00	10	2	80.00	20	2	90.00
Ripisylves et rivulaires	63	6	90.48	38	7	81.58	101	13	87.13
Autres milieux humides	12	1.5	87.50	17	3.5	79.41	29	5	82.76
Cours d'eau et canaux	10	0	100.00	11	0.5	95.45	21	0.5	97.62
Plans d'eau	70	4.5	93.57	37	6	83.78	107	10.5	90.19
Bassins artificiels	16	2.5	84.38	19	1	94.74	35	3.5	90.00
GLOBAL	2473	288.5	88.33	1527	222.3	85.44	4000	510.8	87.23

Analyse de la matrice de confusion

Sur l'urbain, les confusions entre le Bâti individuel dense et lâche dans les deux sens sont assez présentes. On rappellera que la limite entre les deux densités peut être difficile à voir de manière homogène, néanmoins les erreurs relevées sont assez évidentes. Il s'agit toujours d'une mauvaise appréhension de la densification qui ne correspond pas forcément au ratio de surface anthropisée / surface végétalisée, mais également du contexte environnant : classement d'une habitation isolée en dense sans polygones urbains voisins. Il s'agit de tenir compte du contexte (distance aux tissus continus, environnement proche, à savoir si la parcelle est largement entourée d'espaces ouverts) et de la possibilité d'implanter de nouvelles habitations.

Millesimes 2018 & 2010 Département 88			Confusion de >				
			Cumul erreurs				
			Bâti continu dense	Bâti mixte	Bâti individuel dense	Bâti individuel lâche	Bâti isolé en zone agricole ou naturelle
			1111	1122	1123	1124	1130
Bâti continu dense	1111	2					
Bâti continu aéré	1112	10	3		3	1	
Bâti individuel dense	1123	62	1	15		39	3
Bâti individuel lâche	1124	35			22		9
Bâti isolé en zone agricole ou naturelle	1130	7				2	

Il est important de relever que la distance respective entre les habitations pour bien identifier le bâti isolé et cela reste parfois ignoré.

Le second extrait de matrice ci-après met en évidence des confusions entre les espaces verts urbains et les espaces libres en milieu urbain ou, dans une moindre mesure, les espaces en transition et bosquets. La distinction entre l'espace d'agrément, et l'espace potentiellement constructible est essentiellement basée sur l'accessibilité et l'aménagement. Sans doute la notion de poumon vert vient-elle perturber cette analyse, car il faut garder à l'esprit qu'elle ne s'applique que dans l'hyper-urbain et plus particulièrement pour des surfaces boisées et/ou difficilement exploitables.

Millesimes 2018 & 2010 Département 88			Confusion de >					
			Cumul erreurs					
			Espaces libres en milieu urbain	Emprises d'activités à dominante industrielle	Exploitations agricoles	Espaces en transition	Prairies, friches et délaissés agricoles	Bosquets et haies
			1140	1311	1330	1610	2310	2320
Espaces libres en milieu urbain	1140					2	2	
Emprises d'activité à dominante mixte ou tertiaire	1313		6	6		1		
Zones d'extraction	1340		1			4		2
Espaces associés aux réseaux routiers et ferrés	1413	1						2
Espaces verts urbains	1510	5				2	1	2

On observe également des confusions entre les espaces libres en milieu urbain et les emprises d'activités à dominante industrielle ou les exploitations agricoles. Il s'agit pour le premier d'une mauvaise interprétation de la nomenclature, car les espaces libres dans les ZAC ne sont pas à isoler.

Sur l'agricole, on note surtout des confusions entre les classes prairie, friches et délaissés agricoles et les surfaces en herbes naturelles. Outre la rugosité de la surface qui doit être mieux prise en compte, on revient sur l'analyse au niveau 1, l'un étant du naturel et l'autre de l'agricole. Il faut donc tenir compte du contexte environnant (forêts, reliefs, accès) pour mieux classer les polygones.

Le classement de prairies en espaces libres en milieu urbain est toujours présent. Il s'agit encore une fois de revenir au niveau 1 de la nomenclature, car on analyse mieux le contexte : une parcelle en herbe noyée dans un tissu urbain est-elle vraiment une prairie à but agricole ? On voit généralement assez facilement si le cas est logique ou non.

Sur le forestier et naturel, on note toujours des confusions entre les bosquets et les formations arborées. Il s'agit parfois de l'absence de prise en compte de la surface maximale des bosquets ou de la position isolée de l'îlot forestier dans l'agricole par exemple.

On retrouve toujours des confusions intra-forestières (feuillus / conifères / mixtes) avec un nombre d'erreurs plus élevé sur le mixte. Le nombre de confusions avec le pré-forestier est toujours important. Néanmoins, il est difficile à traiter, car il s'agit d'estimer la hauteur de la strate arborée. Il est à noter que l'image 2010 sur ce département est de qualité hétérogène et que certaines parties présentent peu d'ombres ce qui peut induire en erreur.

Un certain nombre d'erreurs sur les coupes forestières a été observé sur 2010. Dans les faits, n'ayant pas de référentiel avant 2010, on ne peut se baser que sur des indices fiables : fûts au sol, trace d'engins, etc. La mise en coupe d'un espace dégagé sur une simple suspicion n'est pas acceptable, car elle introduit une surestimation des coupes.

Millesimes 2018 & 2010 Département 88		Cumul erreurs	Confusion de >										
			Bosquets et haies	Forêts de feuillus	Forêts de conifères	Forêts mixtes	Coupes à blanc et jeunes plantations	Peupleraies et sapinières	Pelouses et pâturages de montagne	Formations pré-forestières	Surfaces enherbées semi-naturelles	Ripisylves et rivulaires	Autres milieux humides
			2320	3110	3120	3130	3140	3150	3210	3220	3230	4110	4120
Prairies, friches et délaissés agricoles	2310	41							1	3	25		
Bosquets et haies	2320	37		6		3		1		17		5	
Forêts de feuillus	3110	33	6		5	2	4			12		2	
Forêts de conifères	3120	19	1	1		7	2	2		3	1	1	
Forêts mixtes	3130	46	4	17	14		1			8	1	1	
Coupes à blanc et jeunes plantations	3140	29			3	1				16	8	1	
Peupleraies et sapinières	3150	5								2	1	1	
Pelouses et pâturages de montagne	3210	7					1				5		
Formations pré-forestières	3220	20		1	2	2	2				11		
Surfaces enherbées semi-naturelles	3230	28			2				8	10		1	3

Enfin, la confusion formations pré forestière soit avec les forêts, soit avec les surfaces enherbées semi-naturelles vient également de la difficulté à apprécier la hauteur des arbres. C'est un aspect difficile, qui peut être facilité lors de la mise à jour 2010, puisqu'en 9 ans, le préforestier devrait en toute logique se retrouver classé en forêt. Évidemment, certaines formations préforestière n'évoluent pas, mais elles restent minoritaires.

Sur les surfaces en eau, on note juste une petite erreur récurrente de confusion entre bassins et plans d'eau. Le classement entre ces deux classes peut être suggestif notamment lorsque l'on est dans une zone de loisir (agrément ou pêche).

Millesimes 2018 & 2010 Département 88			Confusion de > avec				
			Cumul erreurs	Ripisylves et rivulaires 4110	Autres milieux humides 4120	Cours d'eau et canaux 5110	Plans d'eau 5120
Ripisylves et rivulaires	4110	8		2			
Autres milieux humides	4120	5	2			3	
Cours d'eau et canaux	5110	0					
Plans d'eau	5120	10		2			7
Bassins artificiels	5130	3			1		

Conclusion

Pour mémoire, les spécifications du CCTP sont :

- 80% de fiabilité par classe
- 90% de fiabilité au global

	2019	2010	TOTAL
Espaces verts urbains	71.21	71.05	71.15
GLOBAL	88.33	85.44	87.23

Néanmoins, les données vectorielles :

- OCS_GE2_D88_2019.shp
- OCS_GE2_D88_2010.shp
- OCS_GE2_D88_Multi_Date_2010_2019.shp
- OCS_GE2_D88_Multi_Date_Evolution_2010_2019.shp

sont **validées**