



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

RÈGLES DE PRODUCTION

Version Finale

David LOY – Jean QUAILE

Septembre 2021

Table des matières

Préambule	3
Règles de digitalisation	4
Degré de tolérance logicielle.....	4
Lissage automatique des lignes	4
Digitalisation	6
Règles d'UMC et LMC	9
Règles de l'agrégation	13
Règles de segmentation	14
Règles concernant l'imbrication	16
Règles de jonctions.....	17
Intégration d'une grille dans le squelette	19
Utilisation des données exogènes	20
Gestion des commentaires.....	20
Gestion des tables attributaires	20
Futures mises à jour	22

Préambule

Dans ce marché, une occupation du sol sur la région Grand Est va être produite par photo-interprétation pour 2 millésimes différents : 2018/19 et 2007/10. Le millésime 2010 sera réalisé par mise à jour du millésime 2018 et donc produit dans un second temps.

L'utilisation de cette donnée peut intervenir dans différentes thématiques : étude environnementale (trames vertes et bleues par exemple), urbanistique (densification, mutations...) ou agricole pour observer les évolutions surfaciques ou typologiques.

La production est divisée en 2 lots : Le lot 1 concerne les départements 54, 55, 57, 67, 68 et 88. Le lot 2 concerne les départements 08, 10, 51 et 52.



Le territoire régional couvre 10 départements : 57 433 km² + zones tampon

La finalité est de produire sur l'ensemble du territoire de la région Grand Est :

- Une couche SIG d'occupation du sol et du degré d'imperméabilisation 2018/2019.
- Une couche SIG d'occupation du sol et du degré d'imperméabilisation 2007/2010.
- Une couche SIG des mutations entre 2007/2010 et 2018/2019.

Afin d'assurer l'homogénéité sur un tel projet, certaines règles de production doivent être mises en place afin de garantir une qualité optimale et minimiser les artefacts ou incohérences.

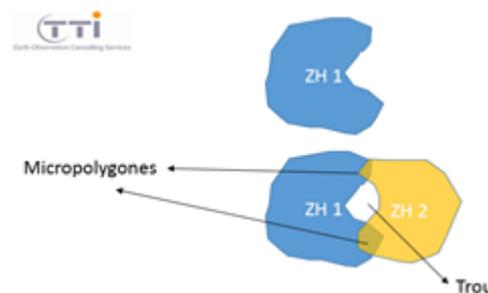
Ce rapport détaille donc les différentes règles de production à utiliser pour ce projet. Ces règles devront également être respectées lors des prochaines mises à jour de ce travail.

Règles de digitalisation

Degré de tolérance logicielle

Les différents logiciels utilisés pour la photo-interprétation ont des seuils variables de tolérance dans la construction topologique. De fait, les formats dans lesquels les polygones peuvent se superposer ne partagent pas de lignes identiques entre eux, mais en font la copie (exemple le Shape-file). Cela induit que suite à des manipulations, une ligne peut être légèrement modifiée sur un polygone et non sur son voisin.

Lors d'une reconstruction topologique, il s'agit de recouper les limites des polygones entre elles afin de voir si cela génère des intersections (cas des superpositions) ou à l'inverse des vides.



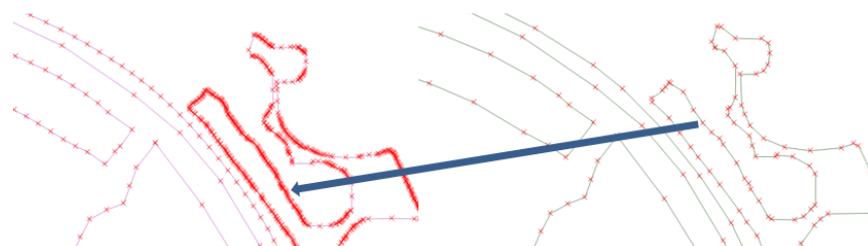
Ainsi, durant les calculs d'intersection, les logiciels utilisent un nombre de décimales plus ou moins grand. Cela induit que certains logiciels trouvent des intersections là où d'autres ne trouvent rien.

Le producteur de l'OCS 2018 – 2010 utilise ArcMap pour sa production. Il a été défini d'utiliser un facteur de précision de **0.0001**.

Pour les contrôles qualité ou les prochaines mises à jour, il conviendra de réutiliser le même paramètre afin de ne pas générer de micro-polygones non pertinents.

Lissage automatique des lignes

Certaines méthodologies de production génèrent des densités de points bien plus élevées que ne le fait un photo-interprète en traçant à la main. Il s'agit notamment de la création SIG de « buffer ». Ces zones tampon sont couramment utilisés pour produire le squelette (routes, voies ferrées, hydrographie), mais également d'autres classes comme les ripisylves par exemple.

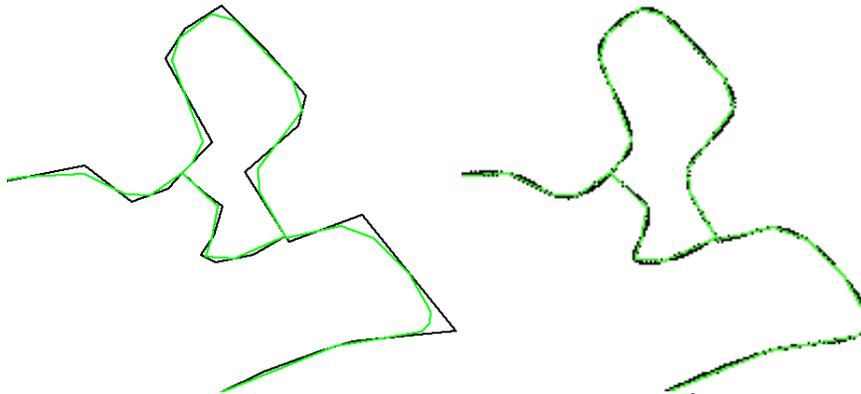


Exemple de ré-échantillonnage du nombre de points à droite.

OCS GE2

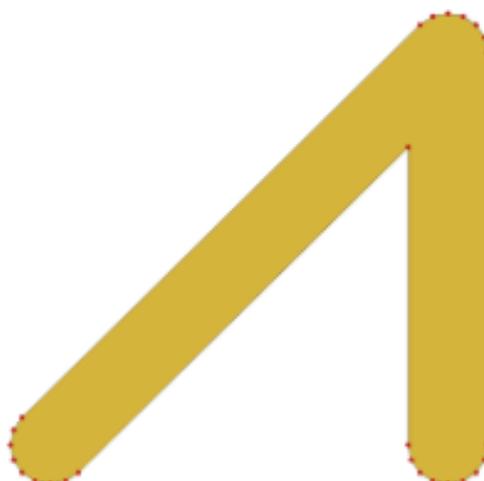
Or, la densité de points produite n'est pas nécessaire pour un rendu optimal de la base de données. En revanche, elle alourdit considérablement et inutilement le poids de la couche vectorielle. De ce fait, il est recommandé de sous-échantillonner ces lignes trop denses pour réduire le nombre de points avant intégration dans la couche vectorielle.

Néanmoins, il convient de régler au mieux ce processus qui a le désavantage de créer des limites anguleuses s'il n'est pas maîtrisé. Dans l'exemple ci-dessous, on voit la différence entre deux versions.



À gauche, les paramètres de simplification sont trop forts. À droite, ils sont mieux maîtrisés, car bien qu'allégé, le résultat en noir se superpose presque à la donnée d'origine.

Un nouveau processus de simplification par le producteur a été mis en place. Ce processus permet d'obtenir 10 points par arrondis de zone tampon contre 17 dans la construction initiale du socle. Cela permettra de gagner un poids non négligeable sur les fichiers Shape-File.



Exemple de la densité de point à 10 points max par arrondi

OCS GE2

Digitalisation

Les règles de précision de la digitalisation utilisent généralement les paramètres suivants :

- Classe de précision totale [p] : 1 mm à l'échelle d'interprétation soit au maximum 2.5 m pour le 1/2500 et 2 m pour le 1/2000
- Le coefficient de sécurité C est égal à 2
- Le coefficient k fonction du nombre de coordonnées X, Y, est de 2.42

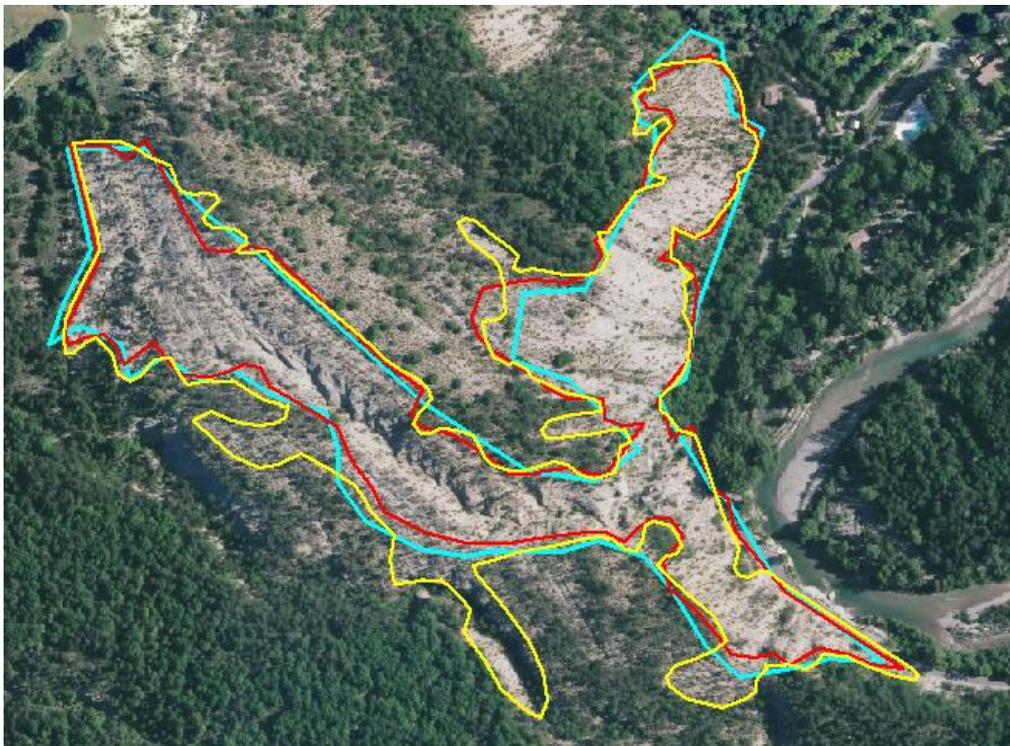
EM (moyenne arithmétique des écarts) : $[p] \times (1 + 1/(2 C^2)) = 2.8 \text{ m}$ au maximum

EJO (Écart ne pouvant jamais être observé) : $1.5 k [p] \times (1 + 1/(2 C^2)) = 10.5 \text{ m}$

Ces paramètres sont évidemment théoriques et il faut ajouter à cela une distance significative. Un écart ponctuel ou sur une courte longueur n'a pas à être sanctionné sur un travail d'occupation du sol si important. Comme la LMC, ils permettent de cadrer la photo-interprétation afin qu'elle soit homogène et avec le moins possible de tracés aberrants. Pour l'EJO (Écart ne pouvant jamais être observé), une longueur de 0.6 cm à l'échelle d'interprétation paraît raisonnable. Pour l'EM (moyenne arithmétique des écarts), une tolérance d'une longueur de 1 cm à 3 cm à l'échelle d'interprétation peut être appliquée selon les cas de figure.

La précision du tracé tient à plusieurs paramètres :

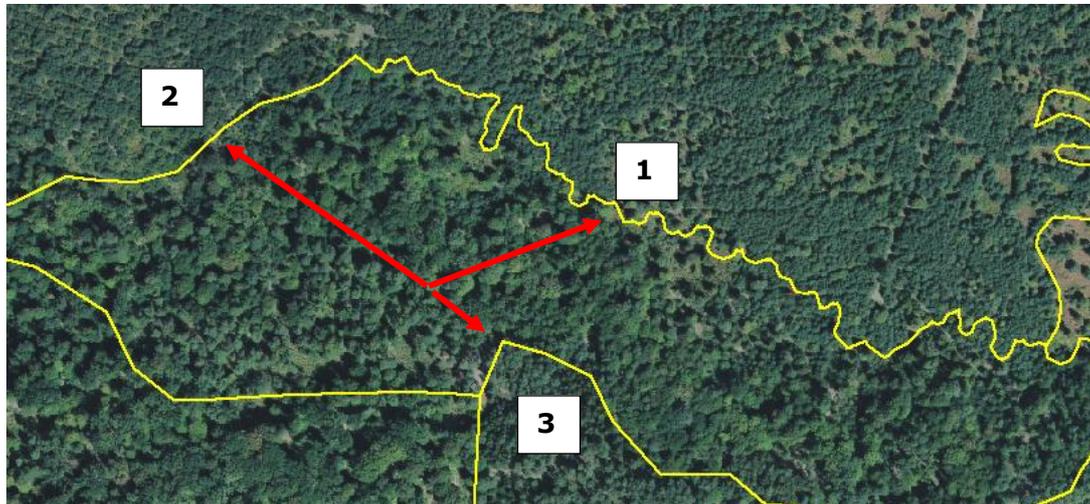
- L'échelle de travail
- Le soin du photo-interprète
- L'intégration de lignes existantes provenant du squelette, du socle ou d'une autre base de données (ex. : BD Forêt).



Exemple de trois tracés plus ou moins détaillés pour une même entité, faits par trois photo-interprètes

OCS GE2

Ces différences peuvent être préjudiciables à l'homogénéité d'une base de données vectorielle. Si les photo-interprètes ne travaillent pas à la même échelle, si certains ont tendance à tracer rapidement (ou inversement trop précisément), ou si des lignes sont intégrées via SIG dans la couche, cela va biaiser le découpage de l'occupation du sol sur l'ensemble de la zone.



Sur cet exemple, un même polygone montre 3 cas de figure :

- 1) Un surdécoupage inutile qui ressemble fortement à l'intégration d'une base forêt
- 2) Une limite pouvant être qualifiée de « normale »
- 3) Une limite anguleuse, trop imprécise, tracée trop vite ou à une échelle inadaptée.

Un contrôle interne ET externe dans les premières phases de photo-interprétation doit être fait pour repérer d'éventuelles différences au sein de l'équipe de photo-interprètes ou d'un assemblage de données exogènes trop ou trop peu précises. On peut notamment citer deux contrôles.

Contrôle des écarts à la réalité

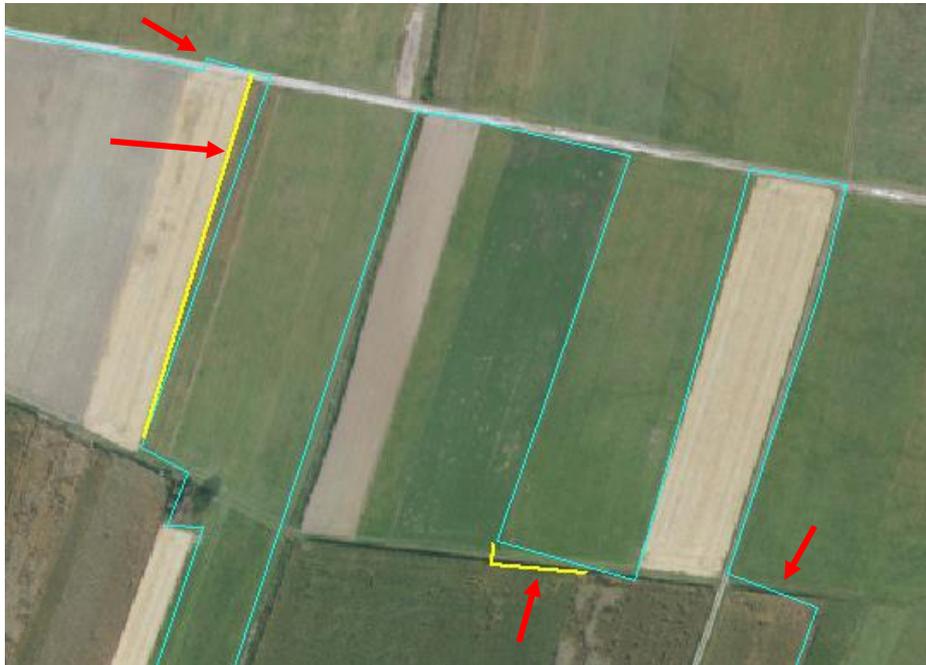
C'est le contrôle des EM (moyenne arithmétique des écarts) et EJO (Écart ne pouvant jamais être observé). Cela ne peut se faire que visuellement par rapport à l'image sur des secteurs aléatoires ou relatifs à l'équipe de production.



À gauche, le tracé s'écarte du cours d'eau : on est dans la marge de l'EM, à droite il s'écarte de la berge, on est typiquement dans l'EJO.

OCS GE2

Les deux cas de figure ci-dessus sont caractéristiques d'une non-mise à jour de tracé, soit issus d'une base (plan d'eau de la BD-Topo par exemple) ou d'un millésime sur l'autre. Mais d'autres cas sont tout simplement liés à un tracé trop rapide.



Exemple d'un tracé trop rapide : les traits de correspondent pas à la réalité (en jaune les limites qui auraient dû être faites)

Contrôle de la densité d'information

Il s'agit de contrôler si la segmentation est suffisamment fine par rapport à l'échelle d'interprétation. Pour rappel, l'UMC impose de diviser ou d'isoler un polygone ou deux occupations du sol sont observables si elles entrent dans cette UMC. Le même contrôle visuel par rapport à l'image sur des secteurs aléatoires ou relatifs à l'équipe de production doit être fait.



À gauche, la segmentation est incomplète car il manque clairement des entités. À droite, on a également des extensions oubliées probablement lors de la mise à jour.

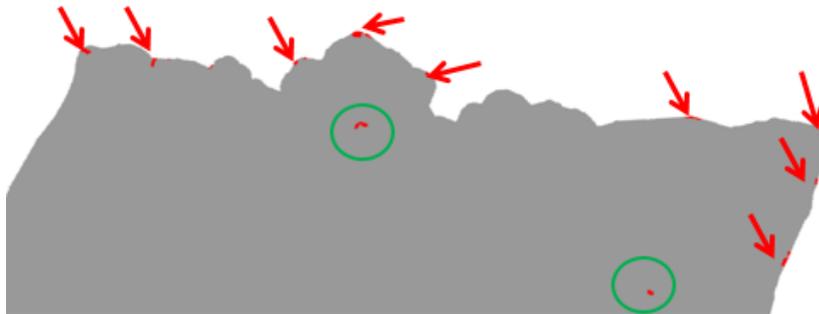
Règles d'UMC et LMC

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 adaptée.

Exceptions :

1) Le découpage lié aux bordures d'une zone de travail (département ou son buffer extérieur) 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 donc pas à prendre en compte.



2) La continuité des réseaux est privilégiée dans cette étude. Cela concerne notamment les classes suivantes :

- 1411 Emprise réseau ferré
- 1412 Emprise réseau routier
- 5110 Cours d'eau et canaux

Ainsi, dans ces classes des polygones à la surface inférieure à l'UMC pourront être préservés dans la limite du raisonnable. Il convient de supprimer les éventuels micro-polygones.

Remarque : les frontières extérieures au projet (dans la zone tampon de 1 Km) ne sont pas destinées à être complétées sur le côté opposé. Le photo-interprète évitera autant que possible de dessiner de très petits polygones au niveau de cette limite.

La **largeur minimale de collecte** (LMC) correspond à la plus petite largeur cartographiable. Normalement, aucun couloir ou élément de réseau ne doit être de largeur inférieure à celle-ci. Il est entendu que ce respect de la LMC ne s'entend pas ponctuellement, mais sur une longueur significative.

On peut distinguer deux cas de figure.

Les LMC des réseaux

Les réseaux (routiers, ferrés et hydrographiques) sont généralement produits par création de zones tampon sur du linéaire. Le respect de la LMC pour ces classes doit donc être géré dès leur création en ne produisant pas de zones tampon d'une largeur inférieure à la LMC. Pour autant, cela ne règle pas tous les problèmes, car des espaces entre les réseaux peuvent se générer. Il est alors important de corriger ces cas.



Dans cet exemple, la jonction de deux réseaux forme une « épine » très fine entre les réseaux. Parfois, ces épines sont indétectables à l'œil nu mais sont dangereuses pour les requêtes SIG ou les mises à jour.



Ici, on a le même cas, plus visible mais toujours inférieur à la LMC.

C'est un biais de l'assemblage SIG de plusieurs couches vectorielles. Normalement, ces deux éléments devraient être supprimés afin que les deux lignes se rejoignent.



Ici, on a un rétrécissement : les polygones ne se recoupent jamais, mais forment un couloir. Ces cas sont dangereux car on peut coder deux entités différentes en même temps sans s'en rendre compte.



Ici, un mauvais assemblage de couches forme une épine.



Ici, l'intégration du réseau dans un polygone existant forme une langue inférieure à la LMC dans c polygyne.

OCS GE2

Ces artefacts, créés pendant des procédures SIG, sont parfois peu visibles à l'échelle de travail et sont donc difficiles à repérer.

Les LMC des autres classes

Si elles doivent être vérifiées sur les autres postes que les réseaux, il convient toutefois d'accorder un certain degré de tolérance de polygones non conformes si l'on se réfère à l'ensemble de la couche. De fait, pendant le travail de PIAO il est difficile de contrôler la LMC en tout point et il est fort probable que des polygones ne la respecteront pas. En revanche, les assemblages de couches doivent être maîtrisés et contrôlés.



Sur les cas présentés ci-dessus, il s'agit d'une jonction non réalisée entre le squelette et la couche forêt. Il en résulte un micro-couloir.

Cependant, les problèmes de LMC peuvent être faits par le photo-interprète lui-même, soit par inattention, soit pour rassembler deux polygones. Autant cela peut être accepté pour un couloir très ponctuel, pour assembler deux parties afin de dépasser l'UMC, autant cela représente une erreur dès lors que c'est inutile.



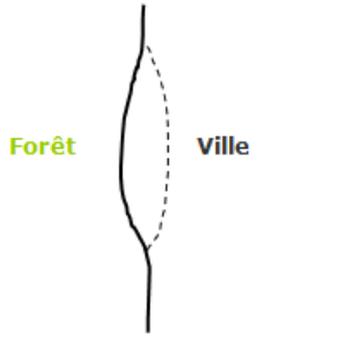
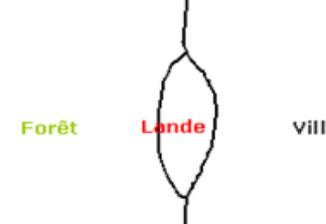
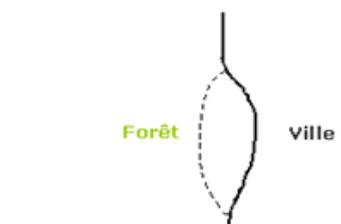
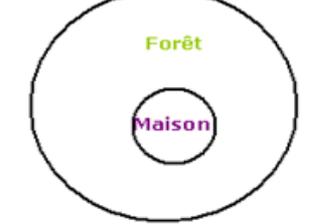
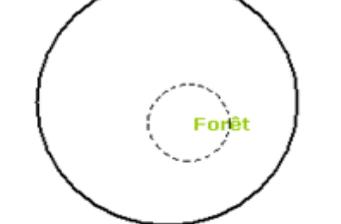
Création d'un couloir entre deux polygones. Il est de plus inutile puisque les deux parties sont bien supérieures à l'UMC.

Règles de l'agrégation

Le regroupement des surfaces ou agrégation est parfois nécessaire afin de respecter les UMC. Cela implique de pouvoir intégrer dans sa cartographie des entités trop petites pour avoir valeur d'information ou des entités, dont l'utilisation ne correspondent à aucun thème de la nomenclature.

Le photo-interprète a alors le choix d'assimiler ce polygone à un voisin, ou éventuellement de réajuster le contour pour atteindre la surface minimum en fonction des différences notables de surface, en prenant soin toutefois de ne pas introduire d'erreurs de reconnaissance.

Les exemples ci-après résument le principe de regroupement qui tend à assimiler un polygone non conforme à son voisin le plus proche au niveau thématique.

	
<p>L'entité Champ est trop petite pour être cartographiée. Elle est située entre deux entités de thèmes différents</p>	<p>On a agrégé les entités liées aux espaces anthropisés, ainsi Champ est intégré à Ville</p>
	
<p>L'entité Lande est trop petite pour être cartographiée. Elle est située entre deux entités de thèmes différents</p>	<p>On a agrégé les entités liées aux espaces naturels, ainsi Lande est intégrée à Forêt</p>
	
<p>L'entité Maison est trop petite pour être cartographiée. Elle est située au cœur d'une entité de thème différent</p>	<p>On a agrégé l'entité la plus petite à celle qui l'englobe. Ce résultat vaut pour n'importe quelle entité</p>



Exemple d'agréations. À gauche, la surface en herbe dans la forêt est trop petite pour l'UMC, elle est donc assimilée à la forêt. À droite, une ripisylve est trop fine pour la LMC, elle est donc intégrée dans la zone arbustive naturelle.

Règles de segmentation

La segmentation d'un territoire suit au moins deux principes qui garantissent la justesse du tracé et l'homogénéité du découpage.

La segmentation logique

Le respect de la segmentation logique concerne la délimitation d'entités différentes séparées par un élément trop étroit pour être considérées comme un polygone appartenant au squelette. Ce problème récurrent en PIAO pose la question de la délimitation de deux entités distinctes séparées par un élément structurant. La solution généralement acceptée est de considérer le milieu du chemin ou de la haie comme limite.



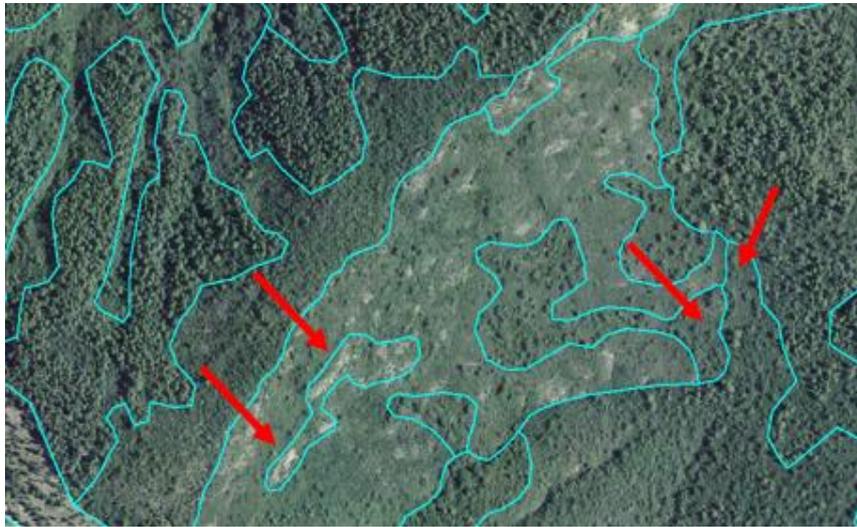
Ici, le photo-interprète essaye de faire passer ses limites de polygones par les chemins du parcellaire agricole. Cela peut également être une haie, un fossé, un mur, etc.

La segmentation homogène

Certaines entités d'occupation du sol présentent des transitions floues. Dans ce cas de figure, le photo-interprète reste seul juge pour trouver la limite de plus grande vraisemblance. Il s'agit de cartographier des ensembles les plus homogènes et de trouver une limite logique et cohérente.

OCS GE2

Dans les milieux naturels, cela peut s'avérer difficile de trouver une limite entre deux entités. L'exemple le plus fréquent est le passage d'une forêt à une forêt ouverte par l'appréciation de la densité.



À gauche, la limite tente de regrouper de petites entités majoritairement rocheuses. À droite, la densité d'arbre est le critère mais son appréciation est différente d'une personne à une autre.

L'homogénéité est parfois trompeuse. Dans les milieux naturels, beaucoup de paramètres peuvent jouer sur la perception :

- L'ombre
- L'humidité
- La saison
- ...



Les différences d'ombrages pourraient faire penser qu'il y a une différence de hauteur d'arbres à droite et à gauche. Or, ce n'est pas le cas

Pour éviter ces mauvaises interprétations, seuls un encadrement, une formation continue, et un retour d'expérience permet d'homogénéiser le travail d'équipe.

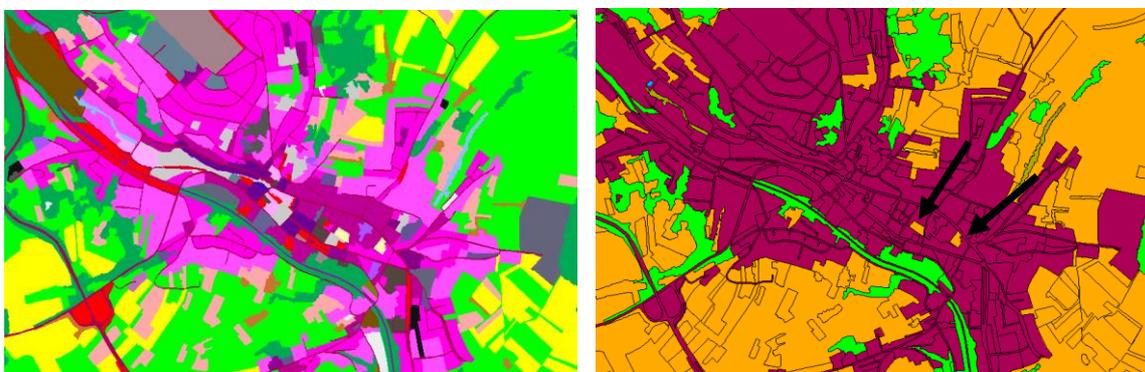
Par ailleurs, il ne faut pas oublier que les ortho-photos sur un département ne sont pas toutes acquises en même temps. Il en résulte un assemble de dates, voire de saisons différentes qui peuvent tromper le photo-interprète.

Règles concernant l'imbrication

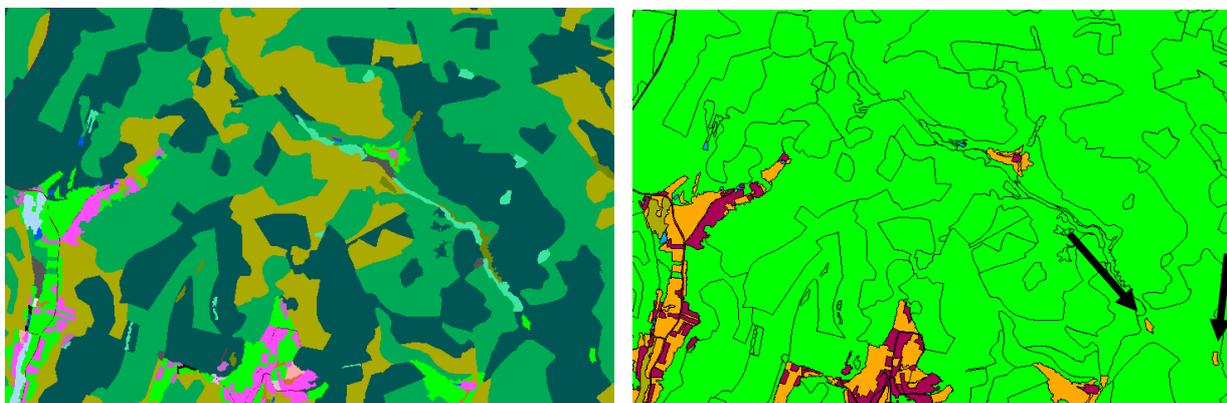
Le principe d'une nomenclature emboîtée veut que pour déterminer une occupation du sol l'on parte du niveau 1, si on peut affiner la caractérisation on passe au niveau 2, et ainsi de suite. Dans la pratique, on code généralement avec le niveau le plus élevé.

Toutefois, le photo-interprète doit garder en tête ce fonctionnement et comprendre que le niveau 4 fait partie d'un sous-groupe au niveau 1 qui divise le territoire en secteurs urbanisés, agricoles et (semi) naturels. Ainsi, s'il code le verger présent sur le terrain d'un particulier, il intègre du territoire agricole dans la trame urbaine, or, ce n'est pas la réalité.

Faire un affichage de l'OCS avec cinq couleurs en fonction du niveau 1 permet en général de voir s'il y a une cohérence géographique dans la répartition de l'occupation du sol. Une classe isolée doit attirer l'œil. Il convient alors de vérifier si ces cas sont une réalité ou une erreur d'appréciation de la nomenclature emboîtée.



Dans cet exemple, le passage d'un affichage au niveau 4 à un passage au niveau 1 fait apparaître des fragments d'agricole dans l'urbain.



Dans cet exemple, le passage d'un affichage au niveau 4 à un passage au niveau 1 fait apparaître des fragments d'agricole dans le naturel.

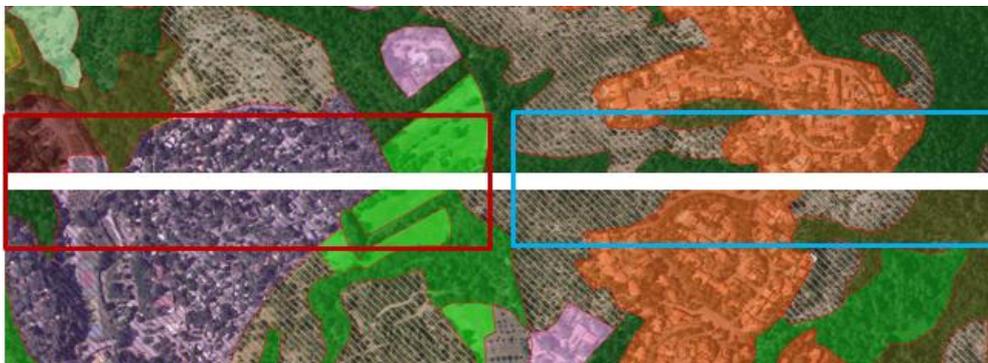
Il est donc important de toujours utiliser la nomenclature dans son entièreté et non de se focaliser sur le niveau le plus élevé. L'imbrication de niveaux dans la nomenclature est une règle importante pour l'analyse postérieure de la base de données.

Règles de jonctions

Le problème des raccordements entre polygones voisins de chaque côté de la limite d'un département est fréquent. En effet, si les photo-interprètes n'ont pas pris en compte ce qui était déjà fait dans le département voisin avant de s'attaquer à la photo-interprétation de la bordure, il est fort probable que des problèmes de raccordement se créent.

Ces discontinuités entre départements sont difficiles à identifier lorsque le travail est déjà fait. Il ne reste que la solution de passer en revue visuellement la zone de jonction des deux départements. Toutefois, ce genre de problème peut être évité en gérant correctement le projet.

Dans l'exemple ci-dessous, on voit que pris séparément, les deux blocs sont bien photo-interprétés en respectant les postes, les UMC et les règles de PIAO.



Exemple de la limite entre deux blocs (haut et bas)

Mais lorsque l'on fait la jonction, apparaissent parfois des artefacts de jointure. Ce n'est pas systématique et parfois la jonction reste acceptable comme dans l'exemple ci-dessous.



Dans le cas ci-dessus (cadre rouge), malgré l'absence d'échange entre les photo-interprètes, la transition est acceptable

Parfois en revanche, on voit nettement les artefacts se dessiner et même des erreurs apparaîtront comme au niveau de la flèche bleue (exemple ci-après) ou l'un des photo-interprètes n'a pas indiqué le bon code d'OCS.



Dans le cas ci-dessus en revanche (cadre bleu), elle ne l'est pas, car les photo-interprètes n'ont respectivement pas tracé les petits bouts de polygones qu'il devrait y avoir de chaque côté. Par ailleurs, au niveau de la flèche, on voit que l'un a classé la forêt en « résineux » et l'autre en « mixte » ce qui peut éventuellement se justifier si on ne tient pas compte de la partie voisine.

De fait, il s'agit souvent d'une interprétation différente, mais pas obligatoirement fautive si l'on considère la règle d'agrégation des petites entités aux plus importantes. Reprendre ces limites est facile, mais peut être chronophage.



Ici, les limites au niveau des flèches rouges ont été corrigées et l'on remarque que la zone de jonction n'est plus réellement visible.

Dans un projet aussi important, réparti sur 10 départements, il est important de s'assurer que les jonctions entre les différentes OCS ne comporteront pas de décalages ou d'incohérences thématiques.

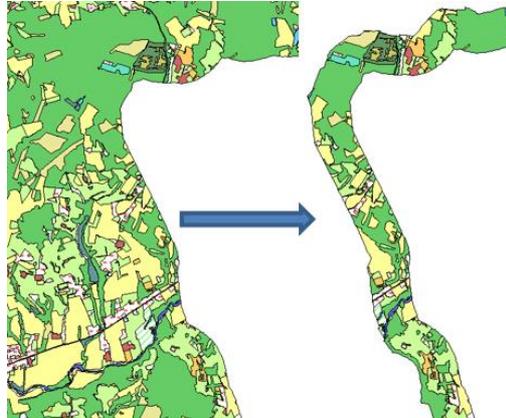
Comme les départements sont produits selon des contraintes de disponibilité de données, l'ordre de production implique qu'ils ne sont pas toujours voisins. Il est donc important de mettre en place une procédure permettant de minimiser les erreurs de jonction. Normalement, si le prestataire suit les recommandations de production, cette étape ne devrait poser aucun problème.

Deux méthodes sont envisageables :

À] Le débordement. C'est la plus simple. Il est demandé de déborder la photo-interprétation au-delà de la limite en dessinant autant que possible ses polygones dans leur entièreté. Ainsi, pour la production du département jointif, le photo-interprète pourra s'appuyer sur de vraies limites d'OCS et non une limite administrative ou de zone tampon. Toutefois il est probable que certains polygones trop grands nécessitent une reprise de limite, car non dessinés complètement lors du débordement.

OCS GE2

B] L'accrochage. Cette méthode est moins facile à mettre en place, car elle dépend du logiciel de PIAO et de la méthode de travail. Il s'agit d'extraire une zone tampon au niveau de la limite du département finalisé.

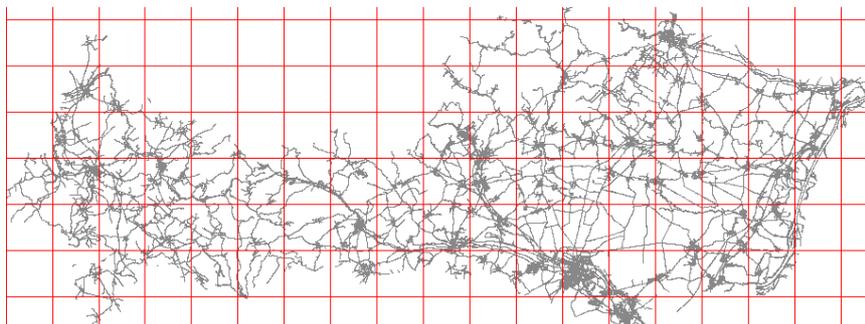


Extraction d'un bloc au niveau de la limite

Le travail sur le département voisin s'attache alors reprendre l'interprétation en reprenant les limites existantes des polygones sans modifier les limites au niveau de la bordure extérieure. La jonction entre les départements après assemblage est parfaite.

Intégration d'une grille dans le squelette

Le squelette présente la particularité de former peu de polygones, parfois même un seul pour le réseau routier, mais ces polygones sont alors très grands, couvrent toute la zone et sont de forme complexe. De fait, certains logiciels SIG atteignent leurs limites de gestion à l'affichage et renvoient une erreur. Pour éviter ce genre de problème, il est possible d'intégrer des limites fictives dans ces polygones afin de les tronçonner. C'est d'ailleurs ce qui est fait dans la majorité des couches à grands polygones de la BD-Topo. Le maillage doit néanmoins être relativement grand pour ne pas générer trop d'intersections.



Recoupement des réseaux par une grille régulière.

Par ailleurs, les intersections étant aléatoires, elles génèrent un nombre de micro-polygones. Il conviendra de définir une taille limite pour ces derniers et de les supprimer par utilisation d'un processus SIG adapté.

Utilisation des données exogènes

Afin de bien comprendre la démarche de classement des photo-interprètes lorsque le polygone nécessite la confirmation d'une donnée exogène, il est important d'en noter la source lorsque celle-ci n'est pas l'une des données ci-dessous :

- BD-Topo
- Scan 25
- RPG
- BD-Forêt
- BD parcellaire ou cadastre
- Base tempêtes
- BD-Alti
- Image IRC

Ainsi, le recours à Google earth / map / street ou au géoportail ou toute autre source de données devra être spécifiée dans le champ « comment ».

Par ailleurs, afin de savoir si le commentaire concerne les millésimes 2018 ou 2010, il est considéré que lorsque rien n'est précisé, il concernera 2018. En revanche, pour le millésime 2010, un « 2010 » devra précéder le commentaire.

Note : ce procédé devra également être utilisé pour les simples commentaires ou doutes du photo-interprète.

Gestion des commentaires

Il s'agit pas d'une règle, mais d'un biais imposé par la présence des champs « comment » et « doute ». En effet, le contrôle des UMC implique qu'il ne peut y avoir deux polygones jointifs de même code (niv4 et niv5). Or, ces derniers ne sont pas fusionnés lorsqu'un processus de suppression de ces limites est réalisé en prenant en compte l'ensemble attributs (hors ID unique). Il conviendra donc de prendre en compte ce biais dans le contrôle et l'élimination des polygones inférieurs à l'UMC.

Il est recommandé de ne pas ajouter de commentaires sur les polygones de grande taille et notamment du réseau, mais de se limiter à des cas particuliers ponctuels.

ATTENTION : il faut noter que laisser les commentaires de la couche 2019 sur la mise à jour 2010 génère des erreurs d'UMC, car le champ commentaire introduit une fausse information dans la dissolution. Ces commentaires obsolètes sont donc à éliminer et réciproquement.

Gestion des tables attributaires

La table attributaire des fichiers Shape file est stockée sous forme de fichier .bdf. En dehors de la géométrie (délimitation des polygones : fichier .Shp) c'est le second fichier le plus volumineux. C'est aussi le seul fichier sur lequel il est possible de réduire la taille en octets.

En effet, pour le shape file, un choix de 4 types de valeurs est disponible en fonction de la dimension du nombre, de son signe et de la présence ou non d'une virgule. Ces différents types n'occupent pas la même place.

OCS GE2

Le tableau ci-dessous résume le poids et la dimension pour chaque type :

Type de données	Plage stockable	Taille (en octets)
Entier court	-32 768 à 32 767	2
Entier long	-2 147 483 648 à 2 147 483 647	4
Réel simple (nombre à virgule flottante à simple précision)	environ -3,4E38 à 1,2E38	4
Réel double (nombre à virgule flottante à double précision)	environ -2,2E308 à 1,8E308	8

De même, pour les champs « texte », le nombre de caractères va conditionner la place. Un caractère correspond à 1 octet. Si 100 caractères sont réservés pour un champ, chaque enregistrement occupe donc 100 octets, même si le champ est vide.

Afin d'optimiser la taille des fichiers, le CQE recommande certaines adaptations par rapport aux livraisons actuelles :

- Le millésime est codé sur 160 caractères, or il ne s'agit que d'une date (année). Un entier court (4) serait optimal.
- Le lib_n1 est codé sur 160 caractères, or l'intitulé le plus long fait 44 caractères. Une chaîne de 45 caractères serait optimale.
- Le lib_n2 est codé sur 160 caractères, or l'intitulé le plus long fait 59 caractères. Une chaîne de 60 caractères serait optimale.
- Le lib_n3 est codé sur 160 caractères, or l'intitulé le plus long fait 50 caractères. Une chaîne de 50 caractères serait optimale.
- Le lib_n4 est codé sur 160 caractères, or l'intitulé le plus long fait 47 caractères. Une chaîne de 50 caractères serait optimale.
- Le lib_n5 est codé sur 160 caractères, or l'intitulé le plus long fait 20 caractères. Une chaîne de 20 caractères serait optimale.
- Les codes cod_n1 à cod_n5 sont codés en Réel double. Un entier court (5) serait optimal.
- Le doute (indice) est codé en Réel double. Un entier court (1) serait optimal.
- Les champs surf_m, surf_ha et perimetre sont codés en Réel double. Un Réel simple (9,2) / (9,6) / (12,2) serait optimal.
- Le champ source est codé sur 160 caractères, or il ne contient que 20 caractères. En prévoyant une marge pour d'autres informations, une chaîne de 70 caractères serait optimale.
- Le champ gid est codé en Réel double. Un entier long (9) serait optimal.
- Les champs comment peut rester sur 160 caractères.

Le gain serait de 863 octets par polygone. Par exemple, le bdf du fichier OCS_GE2_D67_ZONE1_2010.dbf fait actuellement 197132288 octets (197 Mo) pour 137810 polygones.

D'après le calcul, le gain de taille pour cet exemple serait de $863 \times 137810 = 118930030$ octets (118 Mo). On voit donc que pour les couches départementales, le gain de taille devrait avoisiner les 350 Mo, ce qui n'est pas négligeable.

Ces recommandations valent également pour la couche d'évolution qui comprend plus de champs.

OCS GE2

Par ailleurs, se pose la question de mettre systématiquement les intitulés des niveaux 1 à 3 dans les tables attributaires. Si l'utilisateur n'en a pas l'usage, sachant qu'il aura à sa disposition le champ "code" de chaque niveau pour ses éventuelles requêtes SIG, il est inutile de maintenir cette structure. Une table Excel complète pourrait être fournie à part, permettant à ceux qui le désirent de faire une jonction afin de récupérer ces intitulés.

Futures mises à jour

Dans d'éventuelles futures mises à jour de cette base de données, la plupart de ces règles de production devront être respectées, mais certaines d'entre elles n'ont plus lieu d'être, car elles ne concernent que la production initiale de la base d'occupation du sol.

Ainsi, l'intégration d'une grille dans le squelette n'est plus à faire ou à modifier. De même, le processus destiné à mieux gérer les jonctions de département n'est plus obligatoirement nécessaire, mais un soin particulier devra néanmoins être apporté à la délimitation des évolutions au niveau des bordures.