

Le 20 juin 2014

Pôle Observation / Félix Bougé – Marion Pinault

CONTROLE QUALITE DES DONNEES D'OCCUPATION DU SOL PRODUITES SUR LE SCOT DU BASSIN DE VIE D'AVIGNON

Rapport de contrôle de qualité

TABLE DES MATIERES

1) CONTEXTE ET COMMANDE	2
2) SPECIFICATIONS TECHNIQUES DES DONNEES COMMANDEES	3
3) LE CONCEPT DE QUALITE	5
5) MATERIEL ET METHODE DU CONTROLE QUALITE EXTERNE	7
6) PREMIERE LIVRAISON DES DONNEES D'OCCUPATION DU SOL ET CONTROLES PRELIMINAIRES DES DONNEES OCSOL	8
7) RESULTATS DU CONTROLE QUALITE EXTERNE ET REPRISES EN VUE D'UNE SECONDE LIVRAISON	11
8) CONTRIBUTIONS DES MEMBRES DU COMITE TECHNIQUE AU CONTROLE QUALITE DES DONNEES D'OCCUPATION DU SOL	16

1) CONTEXTE ET COMMANDE

Entre 1999 et 2010, l'aire urbaine d'Avignon a connu une augmentation de population de 76% et une extension de 136% de sa surface (publication AURAV octobre 2012). C'est la croissance la plus forte de France métropolitaine. Cette évolution rapide illustre une forte artificialisation (consommation d'espaces naturels, IFEN) du territoire qui s'étend au centre du réseau de villes Avignon, Orange, Carpentras et Cavaillon.

Les lois Grenelle de l'environnement affirment la nécessité de limiter la consommation d'espaces agricoles et naturels et d'intégrer des connaissances plus fines sur les continuités écologiques. Les documents d'urbanisme et de planification (SCOT, PLU, Cartes Communales, PDU) sont donc amenés à intégrer des éléments de connaissance objectifs sur ces questions.

Il apparaît donc prioritaire de mieux connaître et mesurer l'occupation du sol pour en suivre ses évolutions.

L'Agence d'Urbanisme Rhône Avignon Vaucluse (AURAV) a passé, en décembre 2013, un marché avec un prestataire (Alisé Géomatique) pour la production de bases de données d'occupation du sol sur le territoire du SCoT du Bassin de Vie d'Avignon (SCoT BVA) soit 667 km². Ce marché comprenait également trois tranches conditionnelles permettant la production des mêmes données (mêmes spécifications) sur d'autres territoires de SCoT de Vaucluse (SCoT Arc Comtat Ventoux, SCoT Cavaillon Coustellet L'Isle sur le Sorgue, SCoT du Pays d'Apt).

La méthode de production de cette cartographie d'occupation du sol a été élaborée en collaboration avec les utilisateurs finaux de la donnée, les partenaires institutionnels régionaux et l'appui technique du Centre Régional d'Information Géographique de PACA (CRIGE PACA).

Cette commande comprend :

- la production d'une cartographie d'occupation du sol pour deux dates 2001 et 2010, initialement prévu en tranche ferme du marché ;
- la production d'un linéaire de haies en milieu agricole et naturel et d'une trame arborée en milieu urbain, en option du marché ;
- la production d'une mise à jour 2013 basée sur une image PLEIADE (50 cm de résolution) en avenant au marché initial.

Seule la méthode pour contrôler la qualité des données d'occupation du sol 2010 sera proposée dans ce rapport méthodologique sachant que cette donnée a servi de référence pour la production des référentiels 2001, 2013 et des linéaires de haies. La qualité des autres bases de données sera appréciée visuellement mais le contrôle qualité réalisé sera moins exhaustif.

L'objet de ce contrôle qualité est bien de mesurer la qualité des données produites par le prestataire pour vérifier si elles correspondent bien aux spécifications décrites dans le cahier des charges.

A la suite de ce contrôle qualité, des allers retours avec le bureau d'études permettront de discuter des problèmes rencontrés et de demander, si cela était nécessaire, à ce qu'un nouveau jeu de données corrigées soit livré.

Une phase de maintenance corrective débutera ensuite et durera un an après la date de relivraison des données, permettant aux utilisateurs finaux de prendre le temps de tester la donnée et ses limites afin de faire remonter au bureau d'études les dernières erreurs ponctuelles nécessitant une reprise de la base.

2) SPECIFICATIONS TECHNIQUES DES DONNEES COMMANDEES

Référentiel photo :

Propriétaire	Etendue géographique	Date de prise de vue	Précision
Groupement de commande (CRIGE-PACA, CG 84)	84 + 4 communes du 30 (Saze, Les Angles, Rocheford du Gard, Villeneuve lez Avignon)	Juillet 2010	20 cm
BD Ortho - IGN	30	2001, 2006, 2010	50 cm
BD Ortho - IGN	84	2001, 2005, 2009	50 cm

+ Référentiel 2013 : images PLEIADE (Projet PUC - ASTRIUM, mars-avril 2013)

Méthode :

La méthode consiste à digitaliser des zones homogènes sur des photos aériennes prises à des dates différentes et à définir pour chaque portion de l'espace numérisé une classe de description dite « sémantique » (cf. Annexe 1 – nomenclature à 60 classes de niveau 4). La commande a été volontairement orientée vers une méthode de photo-interprétation dite « manuelle » afin d'atteindre la qualité souhaitée (80% poste à poste et 90% en global). D'éventuels pré-traitements automatiques pourront être mobilisés comme support de ce travail de photo-interprétation mais ne devront en aucun cas constituer une donnée pré-finalisée.

Unité Minimale de Collecte (UMC) :

Elle correspond à la plus petite unité spatiale cartographiée. Elle varie entre 0 et 500 m² pour le territoire artificialisé et 500 à 1000 m² pour les espaces naturels. *Les polygones situés sur les bords du SCOT pourront avoir une UMC inférieure à l'UMC demandée.*

Nomenclature de la donnée d'occupation du sol 2010 :

La nomenclature retenue présente 60 postes de niveau 4 imbriqués en 5 niveaux. Seul le poste « parking » a été détaillé au niveau 5. La nomenclature retenue est compatible avec le référentiel d'occupation du sol régional (ocsol CRIGE PACA) jusqu'au niveau 3, lui-même compatible avec les référentiels d'occupation du sol nationaux et internationaux (ocsol grande échelle de l'IGN en cours de réflexion et Corine Land Cover).

Nomenclature des données d'occupation du sol 2001 et 2013 :

La nomenclature des données 2001 et 2013 sera moins précise car les images 2001 (BD-Ortho) et 2013 (PLEIADE) présentent des résolutions bien moins bonnes (50 cm) que celle de 2010 (20 cm) ainsi qu'une date de prise de vue limitante pour 2001 (prise de vue en fin d'hiver). La nomenclature

de ces données compte 18 classes imbriquées en 3 niveaux (Annexe 2). Pour les espaces artificialisés, la nomenclature 2010 est maintenue jusqu'au niveau 3. Pour les espaces agricoles et naturels, la nomenclature est maintenue jusqu'au niveau 2 car l'interprétation de ces espaces est trop compliquée du fait de la résolution des images. Ces données serviront principalement à réaliser un suivi de l'occupation du sol entre 2001 et 2013.

Echelle de saisie à l'écran:

La saisie des objets à l'écran sera réalisée au 1/3000^{ème} en autorisant un passage à 1/5000^{ème} pour les territoires agricoles et naturels tout en gardant une précision suffisante pour la caractérisation du bâti diffus en zone agricole et naturelle.

L'un des objectifs de ces données est de constituer un support pour l'analyse de l'occupation du sol à l'échelle communale notamment à travers les PLU (consommation d'espaces, « dents creuses », potentiel de renouvellement urbain, diagnostic agricole, etc.)

Fiabilité :

La fiabilité de la photo interprétation voulue est de 90% en global (tous postes confondus). Une qualité d'au moins 80% est souhaitée pour chacun des postes de la nomenclature.

Topologie :

Aucune erreur topologique ne sera acceptée.

Délais et livrables :

Les délais de production des données étaient de 5 mois pour la tranche ferme + 1 mois pour les options. La production de la donnée 2013, en avenant au marché initial, a quant à elle été estimée à 1 mois environ par le bureau d'études.

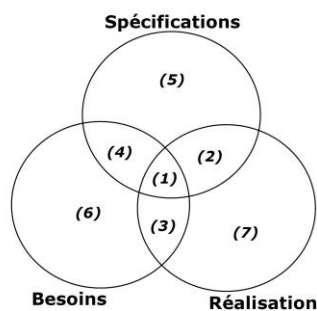
La livraison des données d'occupation du sol 2001 et 2010 doit s'accompagner de celle des documents suivants :

- dictionnaire de postes complet ;
- rapport méthodologique détaillant les règles de saisie ;
- contrôle de qualité interne réalisé.

3) LE CONCEPT DE QUALITE

On attribue généralement deux sens à la notion de qualité d'un produit :

- ▶ Pour le producteur, la qualité désigne la conformité du produit à ses spécifications. La notion d'erreur est donc liée à la violation des règles fixées dans les spécifications. On parle de **qualité interne**.
- ▶ Pour les utilisateurs, la qualité désigne l'adéquation du produit aux besoins et aux attentes des utilisateurs. On parle dans ce cas de **qualité externe**. **C'est cette qualité**



- (1) *Qualité maîtrisée* : besoins bien traduits et réalisés.
- (2) *Sur-qualité* : on a fabriqué ce qui avait été demandé mais dont le client n'avait pas besoin.
- (3) *Qualité plus* : on a fabriqué quelque chose qui n'était pas demandé, mais qui répond à un besoin.
- (4) *Défaut* : le besoin a été défini et la réalisation n'en a pas été faite
- (5) *Double illusion* : on définit quelque chose dont on n'a pas besoin et qui ne sera pas fabriqué
- (6) *Insatisfaction* : le besoin existe mais n'a pas su être défini et la réalisation n'en a pas été effectuée
- (7) *Gaspillage* : on a fabriqué quelque chose dont on n'a pas besoin et qui n'a pas été demandé

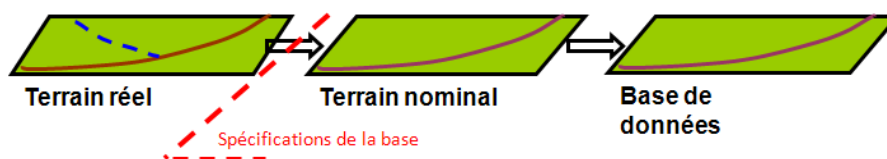
(Source : IGN, mission qualité)

que nous évaluerons dans ce document.

L'ISO définit la qualité comme « l'ensemble des propriétés et des caractéristiques d'un produit ou service qui confère l'aptitude à satisfaire des besoins exprimés ou implicites » (ISO 8402)

Dans le cas précis d'un **contrôle de qualité externe portant sur une donnée d'occupation du sol**, l'objectif n'est pas de comparer la base de données produite au terrain réel mais bien à un terrain dit « nominal » qui est une représentation de la réalité à un temps T. Ce sont les spécifications qui définissent ce que l'on retient du « monde réel » et comment on le représente dans la base (terrain nominal).

Notre contrôle qualité doit donc s'attacher à refaire le travail du photo-interprète en respectant les spécifications techniques demandées dans le cahier des charges (référentiel image, échelle de saisie à l'écran, données exogènes, connaissance partielle du terrain, etc.).



4) RETOURS D'EXPERIENCE SUR DIFFERENTS CONTROLES QUALITE

Données	Méthode	Points forts	Points faibles
Modèle d'occupation du Sol pour la Région Alsace	Contrôle prioritaire sur les postes ayant un taux de fiabilité producteur faible de manière générale	Méthode centrée sur les postes de l'ocsol ayant la fiabilité utilisateur la plus faible Cela permet de gagner du temps.	Méthode qui se base sur les résultats du contrôle qualité du producteur, Certains postes seront moins inventoriés que les autres.
Modèle d'occupation du Sol pour la Région PACA	Contrôle sur des zones définies à dire d'expert avec une surface d'environ 15000 ha	Choix de zones représentatives de l'OCSOL régionale	Durée exponentielle de la photo-interprétation et difficulté à obtenir une numérisation de qualité en peu de temps (temps d'adaptation, accroche-objet)
Service régulier de la BD Forêt pour la région PACA	Création de deux bases de données structurées pour répondre aux spécifications.	Possibilité de vérifier la validité de la géométrie	La base de données Forêt a des spécifications différentes de la BD Ocsol (emprise, échelle, nomenclature, UMC)
Occupation du Sol à grande échelle de la communauté du pays d'Aix (CPA)	Contrôle réalisé au 1/5000 Contrôle automatisé en temps réel sur SIG Contrôle de la pertinence de la photo – interprétation par échantillonnage	Travail de terrain fouillé avec un focus sur les terrains en friches. Création d'un applicatif dédié au contrôle	Pas de calcul d'un indicateur global pour la qualité géométrique de l'occupation du sol.

La connaissance et la compréhension de diverses expériences nous a permis d'élaborer notre propre méthode adaptée à l'examen de nos données d'occupation du sol 2010.

5) MATERIEL ET METHODE DU CONTROLE QUALITE EXTERNE

Notre contrôle qualité externe(CQE) s'est appuyé sur un premier niveau de contrôle portant sur :

- ▶ Les règles de formatage du jeu de données (format, structure) ;
- ▶ La validité topologique des données ;
- ▶ La validité des attributs (cohérence logique, validité sémantique).

Compte tenu du fait que ce sont des méthodes manuelles qui ont été déployées pour la production des données d'occupation du sol 2001 et 2010 par Alisé Géomatique, le contrôle qualité externe sera réalisé par l'AURAV en privilégiant des contrôles par photo-interprétation complétés par des vérifications de terrain.

a) Contrôles par photo-interprétation

Le contrôle de la qualité géométrique et sémantique de la donnée a été réalisé en deux étapes :

- ▶ Préalablement, un choix aléatoire de polygones par échantillonnage stratifié pour chacun des postes de la nomenclature 2010 ;
- ▶ Vérification de la qualité sémantique et géométrique de ces polygones un à un par photo-interprétation (tolérance de 1 mm à l'écran au 1/3 000ème)

Le tirage aléatoire stratifié de polygones a été réalisé de manière à ce qu'au moins une des deux conditions suivantes soit vérifiée :

- Le taux d'échantillonnage du poste est supérieur à 2% du nombre total de polygones de la classe ;
- Le nombre de polygone échantillonné est supérieur à 20. Pour les postes présentant moins de 20 polygones, l'intégralité des polygones de la classe seront donc vérifiés.

L'échantillonnage a été réalisé en tenant compte de l'effectif total de chacune des classes afin qu'il soit statistiquement le plus significatif possible tel que détaillé ci-dessous :

- a. Les 13 postes avec plus de 1000 polygones ont été inventoriés avec un taux compris entre 2 et 5% ;
- b. Les 6 postes comprenant entre 500 et 1000 polygones ont été inventoriés avec un taux d'échantillonnage de 10% ;
- c. Les 19 postes présentant entre 130 et 500 polygones ont été inventoriés à 15% ;
- d. Les 10 postes avec un nombre de polygones compris entre 50 et 130 ont été inventoriés avec des pourcentages allant de 20 à 40% ;
- e. Les 11 autres postes comprenant moins de 50 polygones ont été inventoriés intégralement ;
- f. Le poste « zone incendiée » n'est pas présent, il ne sera donc pas inventorié.

Matrice de confusion : tableau de contingence entre le terrain nominal (en colonne) et le modèle d'occupation du sol (en ligne)

Indice de fiabilité : pourcentage des éléments de la couche d'occupation du sol répondant entièrement aux exigences des spécifications

b) Contrôles sur le terrain

Le travail de terrain va permettre de confronter la photo-interprétation à la réalité du terrain. Au cours du travail de photo-interprétation, les polygones ayant un classement difficile sont signalés par un signe dans une colonne dédié. Cela permet ensuite de prévoir une sortie sur le terrain pour vérifier de visu l'occupation du sol à l'endroit en question en fonction des clés disponibles dans le dictionnaire fourni par le prestataire.

Le travail de terrain interviendra :

- avant le contrôle qualité (quelques jours) pour se mettre dans les conditions du photo-interprète
- pendant et après l'étape de contrôle qualité pour mener des vérifications sur des secteurs qui poseraient des problèmes de photo-interprétation à partir de l'ortho-photo 2010 (quelques jours).

Durant ce travail de terrain, il faudra tenir compte du fait que nous sommes en 2014 et que la donnée a été produite à partir d'une image datant de 2010 (ex : un verger n'a pas pu apparaître en moins de 4 ans, les plantations agricoles ont probablement changé depuis pour les cultures annuelles, etc.). De plus, cette vérification a été faite en mai, une période où la végétation est à un stade avancé qui ne correspond pas forcément aux images sachant que l'ortho-photo de 2001 était prise en fin d'hiver et que celle de 2010 datait de début juillet.

Matériel de terrain :

- Un appareil photo avec GPS
- Un périphérique GPS pour tablette tactile
- Une tablette tactile
- Qgis 2.0

6) PREMIERE LIVRAISON DES DONNEES D'OCCUPATION DU SOL ET CONTROLES PRELIMINAIRES DES DONNEES OCSOL

a) Première livraison

Alisé Géomatique a livré les données d'occupation du sol 2001 et 2010 le 24 avril 2014 soit à peine 4 mois après le lancement du marché (lancement fin décembre 2013). Les délais de production de la tranche ferme sont donc largement respectés, les délais souhaités pour la production de la tranche ferme étant de 5 mois pour l'offre de base et de 1 mois supplémentaire pour les options.

Les différents documents demandés ont également été livrés accompagnés du :

- dictionnaire de postes complet où les nouveaux postes ajoutés à la nomenclature sont détaillés ;
- rapport méthodologique détaillant les « secrets de fabrication » des données par le prestataire (extraction des réseaux, règles de saisie, etc.) ;
- du contrôle de qualité interne réalisé par le prestataire.

Le bureau d'études Alisé Géomatique, a remis son contrôle qualité interne au moment de la livraison de ces données. La production de la donnée 2001 s'appuyant sur la donnée 2010 a permis au prestataire, en amont de son contrôle qualité, de mettre en évidence et de modifier quelques incohérences. Un seul poste, « bâti mixte », a été signalé comme ne respectant pas les spécifications de qualité visées dans les spécifications (80% minimum poste par poste). Le prestataire propose donc de reprendre intégralement ce poste dans la seconde livraison qui sera faite des données 2001 et 2010 suite à notre contrôle qualité externe.

Ce contrôle qualité interne a servi de support à notre contrôle qualité externe détaillé ci-après.

Cette partie présente les résultats du contrôle qualité externe mené à l'AURAV d'avril à juin 2014. Celui-ci est composé d'une présentation des contrôles préliminaires, de la présentation de l'échantillon sélectionné et de la matrice de confusion pour les données 2010.

La donnée d'occupation du sol 2010 comprend 60 postes et un total de 39508 polygones tandis que la donnée 2001 contient seulement 18 postes et 27387 polygones. La nomenclature 2001 est simplifiée par rapport à celle de 2010 et la précision est moindre à cause de plusieurs contraintes :

- la moindre résolution des images,
- des prises de vue en hiver qui ne permettent pas de distinguer les cultures agricoles
- des problèmes d'ortho rectification des images aériennes de 2001 en particulier sur le Gard et dans le secteur d'Orange
- l'absence de données complémentaires ou de terrain.

Pour toutes ces raisons, le balayage de l'ensemble de la donnée 2010 sert de base à la production 2001. C'est pourquoi nous avons centré le contrôle qualité sur les données 2010. Des contrôles complémentaires ont été effectués sur les données 2001 et sur le shape des changements entre 2010 et 2001, ceux-ci viennent confirmer la qualité de toutes les données livrées par Alisé Géomatique et aucun nouveau problème ne vient s'ajouter à ceux repérés pour la donnée de 2010.

b) Contrôles préliminaires

o Formatage des jeux de données

Fichiers	Date	Type fichier	Symbo logie	Taille	Ouverture	Nb colonnes	Nb lignes	Cases vides
OCCSOL _ALL_20 10	24/04/20 14	Shape file	OUI	108 Mo	OK	14	39508	0
OCCSOL _ALL_20 01	24/04/20 14	Shape file	OUI	115 Mo	OK	10	27387	0
CHANG_ 2001_201 0	24/04/20 14	shape file	NON	21,8 Mo	OK	15	10510	0

Le formatage des fichiers est correct. Les fichiers ne contiennent pas de case vide. Ils couvrent l'ensemble du territoire souhaité et sont proposés dans le format demandé.

○ Validité topologique

Fichiers	Surface totale	Trous	Super positions	Ne doit pas se superposer	Ne doit pas avoir de discontinuité	Type vérification
OCCSOL_ALL_2010	710,7 km ²	1	0	OK	OK	Arcgis
OCCSOL_ALL_2001	704 km ²	3	0	Ok	Ok	Arcgis
CHANG_2001_2010	74,41 km ²	-	0	Ok	-	Arcgis

La validité topologique a été vérifiée avec les outils fournis par Arcgis pour les géodatabases. Il n'y a pas de superpositions, pas de discontinuités et pas d'autres erreurs topologiques.

○ Valeurs aberrantes et respect des modalités

Fichiers	Valeurs aberrantes	Nb modalité Niv1	Nb modalité Niv2	Nb modalité Niv3	Nb modalité Niv4
OCCSOL_ALL_2010	0	4	13	26	57
OCCSOL_ALL_2001	0	4	12	18	-

Les tables attributaires des données d'occupation du sol ne présentent pas de valeurs aberrantes et les modalités de choix sur le poste de niveau 4 sont respectées.

NB : 3 classes de niveau 4 n'ont pas été utilisées ainsi que 5 classes de niveau 5.

○ Unité minimale de collecte de la donnée ocsol 2010

Postes des polygones ayant une surface inférieure à l'UMC	Polygones découpés par les limites du SCOT	Polygones avec une erreur (surface comprise dans la marge acceptée de 10%)	Polygones avec une erreur (surface non comprise dans la marge d'erreur acceptée de 10%)
Bâti collectif	0	1	2
Espaces ouverts de loisirs	1	0	0
Céréales	4	0	0
Terres en interculture	10	1	1
Serres	2	2	0
Oliveraies	2	4	1
Réseau routier	2	0	1
Agricole délaissé	10	0	1
Formation semi arbustive	0	1	0
Pelouse et pâturage	3	7	0
Ripisylve	0	0	3
Plans d'eau	1	3	1
TOTAL	35	19	10

Les polygones situés en bordure du périmètre du SCOT peuvent présenter des surfaces inférieures à l'UMC du fait du découpage arbitraire des limites communales et ne doivent

donc pas être considérés comme des erreurs d'UMC. Au total, seuls 10 polygones présentent donc une surface inférieure à l'UMC. **L'UMC est donc globalement très bien respectée.**

7) RESULTATS DU CONTROLE QUALITE EXTERNE ET REPRISES EN VUE D'UNE SECONDE LIVRAISON

La partie qui suit présente les résultats du contrôle sur un échantillon stratifié par poste des polygones de la couche d'occupation du sol 2010 sélectionnés aléatoirement. Chacun des polygones inventorié a été comparé à l'ortho photographie de l'année et confronté à la définition du poste de niveau 5 dans le dictionnaire des données fourni par Alisé Géomatique.

Les postes du modèle d'occupation du sol 2010 sont très détaillés. Nous avons un total de 60 postes présent sur le SCOT du Bassin de Vie d'Avignon au niveau 5 de la nomenclature. Ceux présentent des profils très différents pour les effectifs et la surface totale qu'ils représentent. Le tableau à la page suivante permet de se rendre compte de ces différences.

○ Présentation des 60 postes de niveau 5 de l'OCSOL 2010

Intitulés des postes	Effectifs	Part du total	Surface (ha)	Part de la surface totale
113 10-Espace de bâti diffus en espace agricole	4947	13%	11659	3,41%
112 10-Bâti individuel dense	3753	9%	9 136	4,56%
21130-Terres en interculture	3500	9%	40 14	5,65%
11220-Bâti individuel lâche	2746	7%	3592	3,39%
24 110-Friches agricoles et délaissés en milieu agricole	2205	6%	3440	2,74%
22 110-Vignes	2110	5%	3329	16,41%
2 1120-Cultures légumières et maraichères de plein champ et horticulture	1848	5%	3243	4,20%
2 1110-Cultures céréalières	1629	4%	2984	12,86%
23 110-Prairies agricoles	1580	4%	28 10	5,05%
3 1110-Forêts de feuillus	1556	4%	2463	3,95%
12 130-Espaces d'activité économique	1513	4%	2424	1,99%
222 10-Vergers et petits fruits	1423	4%	24 16	3,40%
14 130-Terrain vague en milieu urbanisé	1087	3%	2407	0,61%
322 10-Pelouses et pâturages naturels	859	2%	1969	2,77%
12243-Autres parkings	744	2%	1947	0,24%
32 110-Formations ouvertes arbustives et/ou arborées	625	2%	1728	2,43%
11240-Bâti collectif	596	2%	14 14	0,50%
12 140-Espaces d'équipements collectifs	518	1%	1253	0,68%
2 12 10-Serres	504	1%	8 16	1,15%
32 120-Formations semi-ouvertes arbustives et/ou arborées	477	1%	702	1,76%
223 10-Oliveraies	471	1%	637	0,34%
32 130-Formations arbustives fermées	363	1%	569	4,84%
12 150-Equipements eaux, électricité, télécommunications et déchets	331	1%	504	0,39%
14220-Espaces ouverts de sports et de loisirs	274	1%	485	0,80%
11320-Espace de bâti diffus en espace naturel	260	1%	437	0,24%
4 1120-Ripisylves	239	1%	362	0,99%
12 120-Espaces commerciaux	225	1%	353	0,23%
13 130-Chantiers	224	1%	333	0,29%
11120-Tissu urbain aéré	219	1%	280	0,16%
14 110-Parcs verts urbains	215	1%	242	0,19%
12 160-Terrain vague en zone d'activité	207	1%	234	0,33%
42 110-Cours et voies d'eau	196	0%	216	3,47%
33 110-Plages, dunes et sable	165	0%	207	0,71%
12241-Grands parkings commerciaux	165	0%	174	0,16%
1110-Tissu urbain compact	162	0%	168	0,19%
13 120-Décharges	147	0%	163	0,20%
422 10-Plans d'eau	141	0%	163	0,23%
3 1130-Forêts mélangées	135	0%	162	0,47%
3 1120-Forêts de conifères	126	0%	162	0,90%
22220-Pépinières (arbres, arbustes)	121	0%	145	0,23%
32 140-Formations principalement arborées	109	0%	143	0,20%
142 10-Espaces bâtis de sports et de loisirs	107	0%	139	0,07%
12220-Réseaux ferroviaires et espaces associés	83	0%	135	0,51%
122 10-Réseaux routiers et espaces associés	75	0%	133	4,68%
11330-Habitat temporaire	71	0%	116	0,07%
12 110-Espaces industriels	68	0%	116	0,20%
12242-Grands parkings - stationnement	60	0%	93	0,08%
332 10-Roches nues	52	0%	72	0,13%
11250-Bâti mixte	49	0%	68	0,05%
11230-Bâti individuel dans parc paysager	44	0%	61	0,10%
14 140-Cimetières	40	0%	56	0,09%
14 120-Places	40	0%	49	0,01%
4 1130-Roselières	34	0%	47	0,10%
13 110-Extraction de matériaux	18	0%	44	0,23%
124 10-Aéroports	14	0%	32	0,30%
12230-Gares routières et/ou ferroviaires	14	0%	8	0,01%
123 10-Haltes fluviales	10	0%	7	0,01%
4 1110-Zones humides indifférenciées	9	0%	7	0,06%
224 10-Lavandes/ lavandins	4	0%	2	0
TOTAL	39507	100%	71070	100%

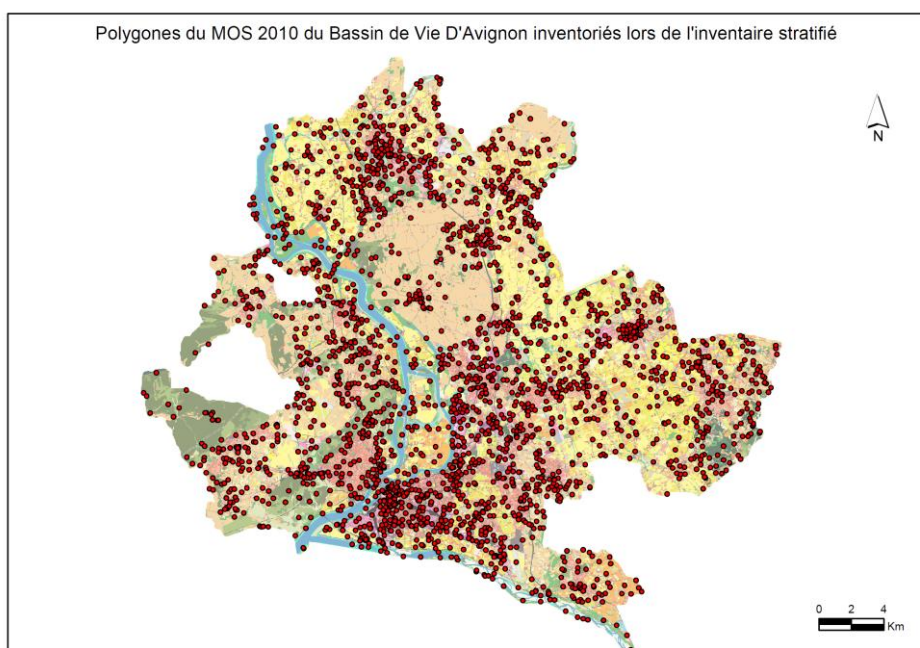
La nomenclature ocsol 2010 est relativement détaillée avec 60 postes de niveau 5. Les postes présentant les plus grandes emprises au sol en 2010 sont les postes « vigne » avec 116 km², « cultures céréalières » avec 91 km², « terres en intercultures » avec 40 km² et « prairies agricoles » avec 35 km².

Le poste « espace bâti en milieu agricole » représente une surface relativement faible mais concerne 13% des polygones de la couche d'occupation du sol. A l'inverse, 10 postes (en rouge dans le tableau ci-dessus) contiennent moins de 50 polygones.

Notre méthode d'échantillonnage a donc été volontairement stratifiée pour s'adapter aussi bien à des postes présentant de très gros effectifs (près de 5 000 polygones) qu'à des postes à très faible effectif (moins de 10 polygones).

La donnée d'occupation du sol 2010 comprend **39 503 polygones** pour une surface totale de **711 km²** (soit le périmètre du SCOT BVA auquel s'ajoute une bande de 200 mètres au-delà de cette limite).

Répartition spatiale des zones inventoriées en tirage aléatoire stratifié



○ Contrôle sémantique de la donnée ocsol 2010 : matrice de confusion

Le tableau qui suit présente la matrice de confusion finale du contrôle qualité réalisé par l'AURAV. Cette matrice se lit en ligne. La diagonale présente les taux de fiabilité par poste à l'issue du CQE. Les cases en dehors de la diagonale représentent les polygones mal classés et leur part par rapport à l'ensemble des polygones inventoriés. Les colonnes situées à droite reprennent le nombre de polygones par poste contrôlés ramenés en taux d'échantillonnage. Cette matrice de confusion est construite sur le même modèle que celle fournie par le prestataire dans son contrôle qualité interne.

o Analyse de la matrice de confusion et reprises en vue d'une 2^{nde} livraison des données d'occupation du sol 2010 et 2001

La matrice de confusion montre que pour la donnée 2010, **le prestataire répond aux spécifications du cahier des charges et dépasse même les exigences du cahier des charges avec une fiabilité globale (overall accuracy) de 97 % et un indice Kappa de même valeur**, soit 7 points de plus que ce qui avait été annoncé par le prestataire à la passation du marché. Parmi les 2 551 polygones inventoriés, 2 477 ont bien été classés. Tous les postes ont été inventoriés à hauteur d'au moins 20 polygones par classe et le taux de polygones bien classés dépasse 80 % poste par poste.

Le poste « Pelouses et pâturages naturels » présente le taux de fiabilité le plus faible (81%) après inventaire de 10 % des polygones de cette classe.

Ces confusions sont liées à un manque de précision dans la définition du poste en amont de la production. Alisé Géomatique a donc facturé à l'AURAV la reprise de ce poste. Le poste a donc été repris intégralement en réaffectant les espaces enherbés associés aux infrastructures routières et ferroviaires au poste « terrains vagues en milieu urbanisé » qui sera rebaptisé « terrains vagues en milieu urbanisé et espaces associés aux infrastructures routières et ferroviaires ». Pour les équipements (base militaire, prison, etc.) présentant de grands espaces enherbés, il a été acté que la partie enherbée soit inclus à l'emprise de l'équipement. Pour les aéroports, de la même façon, les espaces enherbés seront inclus à l'emprise de l'aéroport.

Le poste « Bâti mixte » présente également une fiabilité inférieure aux autres postes avec un taux de fiabilité de 82 %. Cette erreur a déjà été signalée dans le contrôle de qualité interne remis par le prestataire et sera donc corrigée par le bureau d'études en 2^{nde} livraison.

Enfin les postes « Terres en interculture » et « Bâti individuel en parc paysager » présentent quelques confusions mais les taux de fiabilité de ces postes restent largement acceptables avec des taux de fiabilité supérieurs à 90%.

Tous les autres postes de la nomenclature présentent des taux de fiabilité excellents, bien supérieurs à 90%.

o Contrôles géométriques de la donnée ocsol 2010

Une vérification géométrique a été effectuée pour l'ensemble des 791 polygones inventoriés. Nous n'avons remarqué que très peu d'erreurs géométriques. Le contrôle qualité externe montre que les limites géométriques des réseaux et des parcelles ont été scrupuleusement respectées lors de la photo-interprétation. Le tableau suivant illustrent les quelques oublis présents dans la géométrie.

ID	NEW_ID	NEW_NIV5	SURF (m²)
1092	1	14120- Places	733,68
2746	2	11240-Bâti collectif	3,71
3452	3	14120- Places	616,33
2744	4	12230- Gares routières et/ou ferroviaires	822,53
14483	5	12243-Autres-Parkings	8,91

Ce contrôle géométrique montre que seulement 4 polygones ont dû être modifiés sur les 791 inventoriée pour une surface totale modifiée de 2 185 m². **Les quelques polygones mis en évidence seront repris par le prestataire dans la seconde livraison. La qualité géométrique des données produites est donc excellente.**

8) CONTRIBUTIONS DES MEMBRES DU COMITE TECHNIQUE AU CONTROLE QUALITE DES DONNEES D'OCCUPATION DU SOL

Les membres du comité technique ocsol (CG 84, Région PACA, Syndicat Mixte (SM) du SCoT BVA, les EPCI du SCOT BVA, la CC Pays Vaison Ventoux, le SM du SCOT de l'Arc Comtat Ventoux, SM du SCOT du bassin de Cavaillon Coustellet l'Isle-sur-la-Sorgues, SM du SCOT du Pays d'Apt, CRIGE-PACA, Chambres consulaires, etc.) ont été invités à apporter leur contribution au contrôle qualité externe. En tant qu'acteurs de terrain et futurs utilisateurs des données, il nous a semblé normal de les associer à ce stade de la production.

Leur examen des données a reposé sur un balayage visuel de la couche d'occupation du sol sur leur territoire d'intervention. Ces retours viennent compléter l'approche statistique déployée par l'AURAV pour réaliser le contrôle qualité externe des données d'occupation du sol 2010.

Le Grand Avignon, la Communauté de communes des Pays de Rhône et Ouvèze (CCPRO), la Communauté de communes des Sorgues du Comtat (CCSC) ainsi que le CRIGE-PACA nous ont fait part de leurs remarques.

Ces retours ont été particulièrement utiles sur les secteurs emblématiques du territoire qui présentaient parfois quelques imprécisions (Palais des papes, Rocher des Doms, Fort de Villeneuve). Les retours du CRIGE ont également permis de mettre en évidence de petites incohérences avec la méthode de production des données d'occupation du sol moyenne échelle régionale mais globalement les données produites sur le SCOT BVA sont conformes au référentiel ocsol PACA en cours de mise à jour.

ANNEXE 1 : NOMENCLATURE DE L'OCSOL 2010, VERSION FINALE

COULEUR1	typo_niv1	COULEUR2	typo_niv2	COULEUR3	typo_niv3	COULEUR4	typo_niv4	code_niv5	typo_niv5	umet				
1	Territoires artificialisés	11	Zones urbanisées	11	Tissu urbain continu	11	Tissu urbain compact	11110	11110-Tissu urbain compact	500				
	Territoires artificialisés		Zones urbanisées		Tissu urbain continu		Tissu urbain aéré	11120	11120-Tissu urbain aéré	500				
	Territoires artificialisés		Zones urbanisées		Tissu urbain discontinu		Bâti individuel dense	11210	11210-Bâti individuel dense	500				
	Territoires artificialisés		Zones urbanisées		Tissu urbain discontinu		Bâti individuel lâche	11220	11220-Bâti individuel lâche	500				
	Territoires artificialisés		Zones urbanisées		Tissu urbain discontinu		Bâti individuel dans parc paysager	11230	11230-Bâti individuel dans parc paysager	500				
	Territoires artificialisés		Zones urbanisées		Tissu urbain discontinu		Bâti collectif	11240	11240-Bâti collectif	500				
	Territoires artificialisés		Zones urbanisées		Tissu urbain discontinu		Bâti mixte	11250	11250-Bâti mixte	500				
	Territoires artificialisés		Zones urbanisées		Bâti diffus		Bâti diffus	11310	11310-Espace de bâti diffus en espace agricole	500				
	Territoires artificialisés		Zones urbanisées		Bâti diffus		Espace de bâti diffus en espace naturel	11320	11320-Espace de bâti diffus en espace naturel	500				
	Territoires artificialisés		Zones urbanisées		Bâti diffus		Habitat temporaire	11330	11330-Habitat temporaire	250				
	Territoires artificialisés		12		Zones industrielles ou commerciales et réseaux de communication		12	Zones industrielles et commerciales	12	Espaces industriels	12110	12110-Espaces industriels	500	
	Territoires artificialisés				Zones industrielles ou commerciales et réseaux de communication			Zones industrielles et commerciales		Espaces commerciaux	12120	12120-Espaces commerciaux	500	
	Territoires artificialisés	Zones industrielles ou commerciales et réseaux de communication		Zones industrielles et commerciales	Espaces d'activité économique	12130		12130-Espaces d'activité économique		500				
	Territoires artificialisés	Zones industrielles ou commerciales et réseaux de communication		Zones industrielles et commerciales	Espaces d'équipements collectifs	12140		12140-Espaces d'équipements collectifs		250				
	Territoires artificialisés	Zones industrielles ou commerciales et réseaux de communication		Zones industrielles et commerciales	Equipements eaux, électricité, télécommunications et déchets	12150		12150-Equipements eaux, électricité, télécommunications et déchets		250				
	Territoires artificialisés	Zones industrielles ou commerciales et réseaux de communication		Zones industrielles et commerciales	Terrain vague en zone d'activité	12160		12160-Terrain vague en zone d'activité		1000				
	Territoires artificialisés	Zones industrielles ou commerciales et réseaux de communication		Réseaux routiers et ferroviaires et espaces associés	Réseaux routiers et espaces associés	12210		12210-Réseaux routiers et espaces associés		(5m)				
	Territoires artificialisés	Zones industrielles ou commerciales et réseaux de communication		Réseaux routiers et ferroviaires et espaces associés	Réseaux ferroviaires et espaces associés	12220		12220-Réseaux ferroviaires et espaces associés		(5m)				
	Territoires artificialisés	Zones industrielles ou commerciales et réseaux de communication		Réseaux routiers et ferroviaires et espaces associés	Gares routières et/ou ferroviaires	12230		12230-Gares routières et/ou ferroviaires		250				
	Territoires artificialisés	Zones industrielles ou commerciales et réseaux de communication		Réseaux routiers et ferroviaires et espaces associés	Parings	12241		12241-Grands parkings commerciaux		250				
	Territoires artificialisés	Zones industrielles ou commerciales et réseaux de communication		Réseaux routiers et ferroviaires et espaces associés	Parings	12242		12242-Grands parkings - stationnement		250				
	Territoires artificialisés	Zones industrielles ou commerciales et réseaux de communication		Réseaux routiers et ferroviaires et espaces associés	Parings	12243		12243-Autres parkings		250				
	Territoires artificialisés	Zones industrielles ou commerciales et réseaux de communication	Zones portuaires	Haltes fluviales	12310	12310-Haltes fluviales	1000							
	Territoires artificialisés	Zones industrielles ou commerciales et réseaux de communication	Aéroports	Aéroports	12410	12410-Aéroports	500							
	Territoires artificialisés	13	Activités extractives et de stockage et production d'énergie	13	Activités extractives	13	Extraction de matériaux	13110	13110-Extraction de matériaux	500				
	Territoires artificialisés		Activités extractives et de stockage et production d'énergie		Activités extractives		Décharges	13120	13120-Décharges	500				
	Territoires artificialisés		Activités extractives et de stockage et production d'énergie		Activités extractives		Chantiers	13130	13130-Chantiers	500				
	Territoires artificialisés	14	Espaces verts artificialisés et zones de loisirs	14	Espaces verts urbains	14	Parcs verts urbains	14110	14110-Parcs verts urbains	250				
	Territoires artificialisés		Espaces verts artificialisés et zones de loisirs		Espaces verts urbains		Places	14120	14120-Places	250				
	Territoires artificialisés		Espaces verts artificialisés et zones de loisirs		Espaces verts urbains		Terrain vague en milieu urbanisé	14130	14130-Terrain vague en milieu urbanisé	250				
	Territoires artificialisés		Espaces verts artificialisés et zones de loisirs		Espaces verts urbains		Cimetières	14140	14140-Cimetières	250				
	Territoires artificialisés		Espaces verts artificialisés et zones de loisirs		Equipements sportifs et de loisirs		Espaces bâtis de sports et de loisirs	14210	14210-Espaces bâtis de sports et de loisirs	250				
	Territoires artificialisés		Espaces verts artificialisés et zones de loisirs		Equipements sportifs et de loisirs		Espaces ouverts de sports et de loisirs	14220	14220-Espaces ouverts de sports et de loisirs	500				
	2	Territoires agricoles	21	Terres arables	21	Terres arables autres que serres	21	Cultures céréalières	21110	21110-Cultures céréalières	1000			
		Territoires agricoles		Terres arables		Terres arables autres que serres		Cultures légumières et maraichères de plein champ et horticulture	21120	21120-Cultures légumières et maraichères de plein champ et horticulture	500			
		Territoires agricoles		Terres arables		Terres arables autres que serres		Terres en interculture	21130	21130-Terres en interculture	500			
		Territoires agricoles	22	Cultures permanentes	22	Serres	22	Serres	22120	22120-Serres	500			
		Territoires agricoles		Cultures permanentes		Vignobles		Vignes	22110	22110-Vignes	1000			
		Territoires agricoles		Cultures permanentes		Vignobles		Vignes en terrasse	22120	22120-Vignes en terrasse	1000			
		Territoires agricoles		Cultures permanentes		Arboriculture autre que oliviers		Vergers et petits fruits	22210	22210-Vergers et petits fruits	1000			
		Territoires agricoles		Cultures permanentes		Arboriculture autre que oliviers		Pépinières (arbres, arbustes)	22220	22220-Pépinières (arbres, arbustes)	1000			
		Territoires agricoles		Cultures permanentes		Oliveraies		Oliveraies	22310	22310-Oliveraies	1000			
Territoires agricoles		Cultures permanentes		Plantes aromatiques		Lavandes/ lavandins		22410	22410-Lavandes/ lavandins	1000				
Territoires agricoles		23		Prairies		23		Prairies	23	Prairies agricoles	23110	23110-Prairies agricoles	1000	
Territoires agricoles				Prairies				Surfaces essentiellement agricoles interrompues par des espaces naturels importants		Friches agricoles et délaissés en milieu agricole	24110	24110-Friches agricoles et délaissés en milieu agricole	1000	
Territoires agricoles				Zones agricoles hétérogènes				Forêts		Forêts de feuillus	31110	31110-Forêts de feuillus	1000	
3		Forêts et milieux semi-naturels		31		Forêts		31	Forêts	31	Forêts de conifères	31120	31120-Forêts de conifères	1000
		Forêts et milieux semi-naturels				Forêts			Forêts mélangées		31130	31130-Forêts mélangées	1000	
		Forêts et milieux semi-naturels	Forêts		Forêts	Forêts	31140		31140-Forêts		1000			
		Forêts et milieux semi-naturels	32	Milieux à végétation arbustive et/ou herbacée	32	Maquis et garrigues	32	Formations ouvertes arbustives et/ou arborées	32110	32110-Formations ouvertes arbustives et/ou arborées	1000			
		Forêts et milieux semi-naturels		Milieux à végétation arbustive et/ou herbacée		Maquis et garrigues		Formations semi-ouvertes arbustives et/ou arborées	32120	32120-Formations semi-ouvertes arbustives et/ou arborées	1000			
	Forêts et milieux semi-naturels	Milieux à végétation arbustive et/ou herbacée		Maquis et garrigues		Formations arbustives fermées		32130	32130-Formations arbustives fermées	1000				
	Forêts et milieux semi-naturels	Milieux à végétation arbustive et/ou herbacée		Maquis et garrigues		Formations principalement arborées		32140	32140-Formations principalement arborées	1000				
	Forêts et milieux semi-naturels	Milieux à végétation arbustive et/ou herbacée		Pelouses et pâturages naturels		Pelouses et pâturages naturels		32210	32210-Pelouses et pâturages naturels	1000				
	Forêts et milieux semi-naturels	33		Espaces ouverts, sans ou avec peu de végétation		33		Plages, dunes et sables	33	Plages, dunes et sable	33110	33110-Plages, dunes et sable	1000	
	Forêts et milieux semi-naturels			Espaces ouverts, sans ou avec peu de végétation				Roches nues		Roches nues	33210	33210-Roches nues	1000	
	Forêts et milieux semi-naturels			Espaces ouverts, sans ou avec peu de végétation				Zones incendiées		Zones incendiées	33310	33310-Zones incendiées	1000	
	4	Surfaces en eau et zones humides		41		Zones humides intérieures		41	Zones humides intérieures	41	Zones humides indifférenciées	41110	41110-Zones humides indifférenciées	1000
		Surfaces en eau et zones humides				Zones humides intérieures			Ripisylves		41120	41120-Ripisylves	1000	
		Surfaces en eau et zones humides				Zones humides intérieures			Roselières		41130	41130-Roselières	1000	
		Surfaces en eau et zones humides		42		Eaux continentales		42	Cours d'eau et voies d'eau	42	Cours et voies d'eau	42110	42110-Cours et voies d'eau	(5m)
		Surfaces en eau et zones humides	Eaux continentales		Plans d'eau	42210	42210-Plans d'eau		500					

ANNEXE 2 : STRUCTURE DE LA TABLE DU CONTROLE QUALITE « VERIF OCSOL 2010 FINAL.dbf »

Table

verif_OCSOL2010_final

FID	Shape *	ID	NIV5 10	LIB5 10	AREA HA	INVF	INVF VERIF	INVF NEW5	INVF COMM	NIV5 NUM o
0	Polygone ZM	16	0	11110-Tissu urbain compact	1,883	1	1	0		11110
1	Polygone ZM	17	0	11110-Tissu urbain compact	0,52	0	0	0		11110
2	Polygone ZM	18	0	11110-Tissu urbain compact	0,173	0	0	0		11110
3	Polygone ZM	19	0	11110-Tissu urbain compact	0,341	0	0	0		11110
4	Polygone ZM	20	0	11110-Tissu urbain compact	0,12	0	0	0		11110
5	Polygone ZM	21	0	11110-Tissu urbain compact	1,914	0	0	0		11110
6	Polygone ZM	22	0	11110-Tissu urbain compact	0,219	1	1	0		11110
7	Polygone ZM	23	0	11110-Tissu urbain compact	0,132	0	0	0		11110
8	Polygone ZM	0	0	11110-Tissu urbain compact	0,085	0	0	0		11110
9	Polygone ZM	1	0	11110-Tissu urbain compact	0,351	0	0	0		11110
10	Polygone ZM	2	0	11110-Tissu urbain compact	0,1	0	0	0		11110
11	Polygone ZM	3	0	11110-Tissu urbain compact	0,237	0	0	0		11110
12	Polygone ZM	4	0	11110-Tissu urbain compact	3,056	0	0	0		11110
13	Polygone ZM	5	0	11110-Tissu urbain compact	0,293	0	0	0		11110
14	Polygone ZM	6	0	11110-Tissu urbain compact	0,663	0	0	0		11110
15	Polygone ZM	7	0	11110-Tissu urbain compact	0,59	1	1	0		11110
16	Polygone ZM	8	0	11110-Tissu urbain compact	2,165	1	1	0		11110
17	Polygone ZM	9	0	11110-Tissu urbain compact	0,54	0	0	0		11110
18	Polygone ZM	10	0	11110-Tissu urbain compact	0,116	0	0	0		11110
19	Polygone ZM	11	0	11110-Tissu urbain compact	0,262	0	0	0		11110
20	Polygone ZM	12	0	11110-Tissu urbain compact	0,134	0	0	0		11110
21	Polygone ZM	13	0	11110-Tissu urbain compact	0,323	0	0	0		11110
22	Polygone ZM	14	0	11110-Tissu urbain compact	0,19	0	0	0		11110
23	Polygone ZM	15	0	11110-Tissu urbain compact	1,697	0	0	0		11110
24	Polygone ZM	24	0	11110-Tissu urbain compact	0,133	0	0	0		11110
25	Polygone ZM	25	0	11110-Tissu urbain compact	0,152	0	0	0		11110
26	Polygone ZM	26	0	11110-Tissu urbain compact	0,537	0	0	0		11110

ID : identifiant du polygone

LIB5_10 : Numéro et libellé du poste du polygone

INVF (INventaire Final) : 1 = le polygone fait partie de l'échantillon stratifié 0 : le polygone ne fait pas parti de l'échantillon

INVF_VERIF : .0 : le polygone n'a pas été inventorié, 1 : le polygone est conforme à l'OCSOL nominale, 2 : le polygone n'est pas conforme

INVF_NEW5 : Nouveau numéro de poste attribué aux polygones non conformes après le contrôle de qualité

INVF_COMM : Commentaire sur les polygones non conforme à la situation nominale

NIV5_NOM_o : Ancien numéro de poste attribué au polygone