

La segmentation et le programme PYRAM : l'assemblage de la carte des physionomies

Le programme PYRAM, développé historiquement par l'IGN, permet la réalisation d'une segmentation à partir des dalles d'orthophotos en proche infrarouge de 1 x 1 km d'emprise. Le processus de segmentation s'appuie sur des critères radiométriques et géométriques pour identifier des zones homogènes dans l'image de manière automatique (champs, forêts, bâtiments, etc.). Ce processus permet de produire les différents niveaux de découpage ou « niveaux de coupes » qui sont utilisés pour la production de la carte des physionomies. Les différents niveaux de coupes produits sont hiérarchiquement imbriqués. Cela signifie qu'il est possible de reconstituer les limites des niveaux de coupes supérieurs à partir de celles des niveaux inférieurs. Pour chaque niveau de coupes choisi, une nouvelle image est produite dans laquelle tous les pixels appartenant à une même zone homogène se voient attribuer une couleur identique. Le résultat de la segmentation est donc une image (donnée en mode raster). Seuls trois niveaux de coupe sont utilisés. La figure 1 illustre les trois niveaux de segmentation.

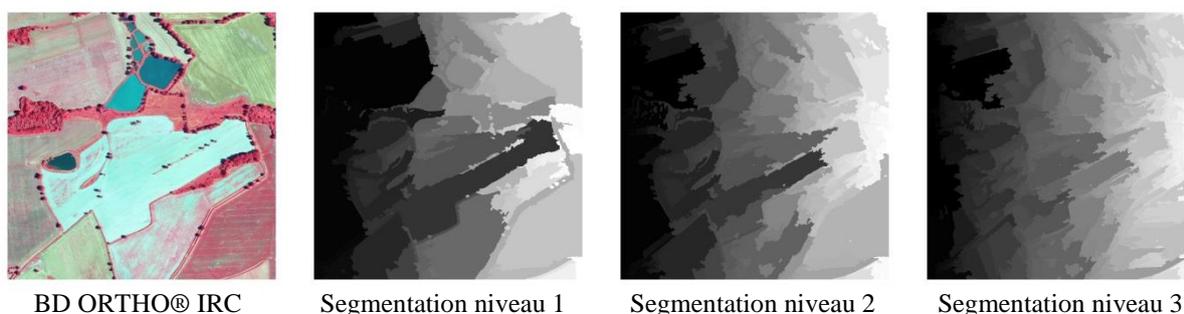


Figure 1 : Exemple de différents niveaux de segmentation raster d'une même dalle de 1 x 1 km de la BD ORTHO® en proche infra-rouge (le niveau de coupe le plus fin à gauche)

Une fois la segmentation réalisée, l'image est vectorisée par un logiciel développé par l'IGN. Les zones homogènes issues de la segmentation sont transformées en objets vectoriels (polygones) dont les contours correspondent aux limites de zone. Cette transformation vectorielle a pour but de faciliter la qualification physionomique des différentes zones issues du découpage. L'attribution de cette information nomenclaturale se fait à travers l'étape de classification. Celle-ci permet d'établir une correspondance entre la signature spectrale et la texture d'un objet (ici une zone homogène), et un type de milieu (culture, prairie, forêt, etc.). La segmentation est réalisée sur une image infrarouge couleur (IRC) plus appropriée pour l'étude de la végétation (voir l'IF numéro 25, en ligne : http://inventaire-forestier.ign.fr/spip/IMG/pdf/IF_25_proche_infrarouge_2.pdf). Dans cette image les couleurs rouge, vert et bleu, représentent respectivement les canaux « proche infrarouge, rouge et vert ». L'intensité de la réflexion de la lumière par le couvert végétal dans le rouge et le proche infrarouge est généralement considérée comme étant liée au degré d'activité chlorophyllienne des plantes (débourrage, état sanitaire, etc.). En fonction de la date de prise de vue, l'intensité de l'activité chlorophyllienne pourra également révéler certaines conditions environnementales ou pratiques culturales (fauche, pâture, etc.). Les figures 2a et 2b représentent le tracé, au format vecteur, des limites des trois niveaux de coupe de la segmentation sur la BD ORTHO® IRC de l'IGN.



La segmentation et le programme PYRAM : l'assemblage de la carte des physionomies

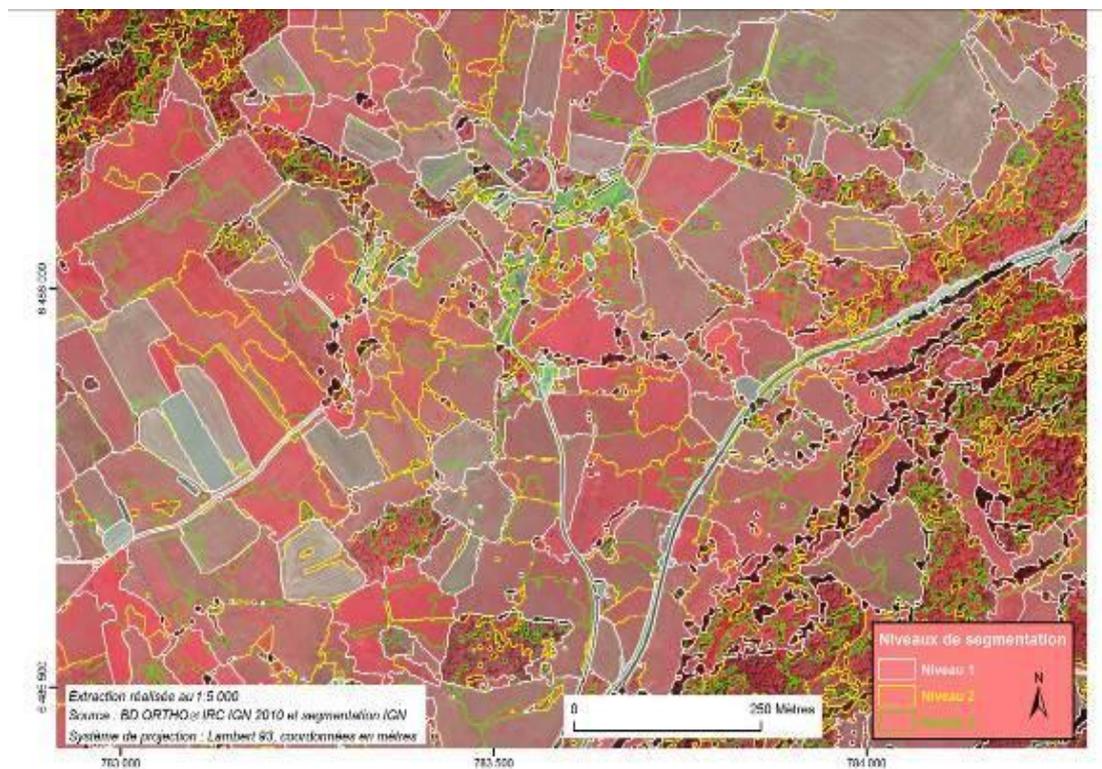
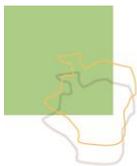


Figure 2a : Illustration des trois niveaux de découpage utilisés. Les niveaux 2 (polygones jaune) et 3 (vert) sont intéressants à conserver pour proposer d'éventuels redécoupages de polygones de niveau 1 (blanc), spécialement au niveau de l'hygrosère qui ressort, selon les secteurs étudiés, dans les différents niveaux de la segmentation.



La segmentation et le programme PYRAM : l'assemblage de la carte des physionomies

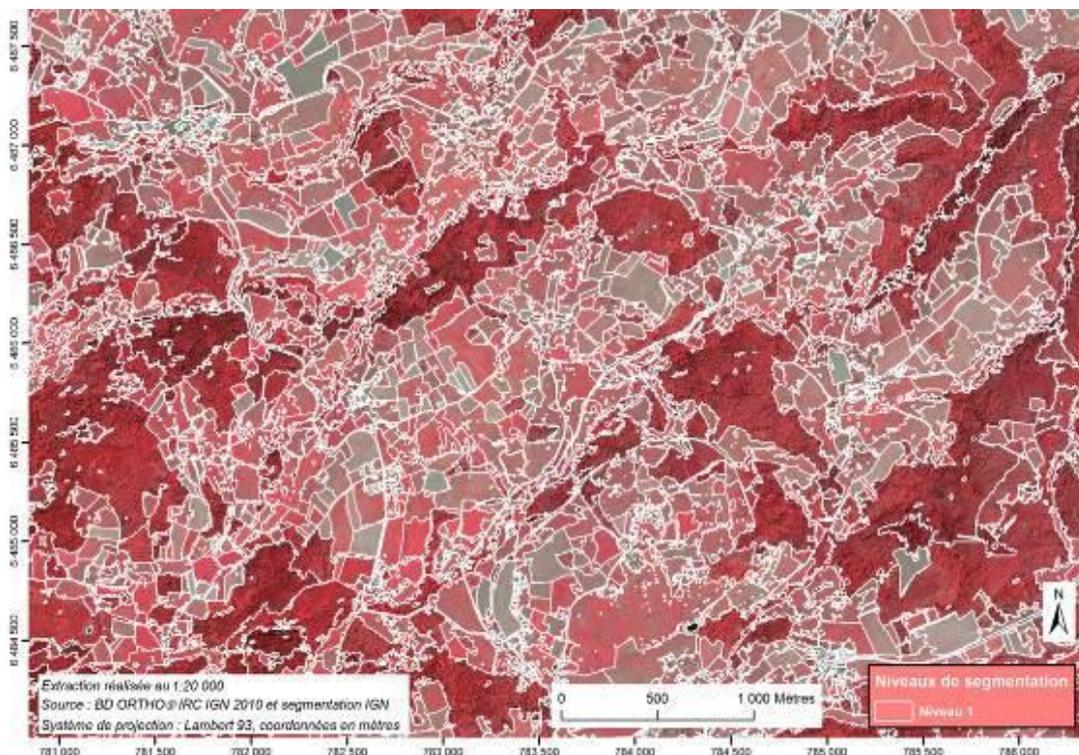


Figure 2b : Illustration du niveau 1 de découpage utilisé. Cette couche correspond à l'information de base utilisée pour la réalisation de la pré-cartographie.

Après vectorisation de la segmentation de niveau 1, un ensemble de masques est utilisé pour simplifier la couche résultante. Des post-traitements sont ensuite appliqués, ils portent *a minima* sur les points suivants :

- corrections topologiques ;
- suppression des polygones de taille inférieure à 2 500 m² (0,25 ha) par fusion avec le polygone voisin avec lequel ils partagent la plus grande frontière ;
- suppression des polygones fins et allongés par fusion avec le polygone voisin avec lequel ils partagent la plus grande frontière ;
- correction de la forme de certains polygones en bord de route.

L'utilisation de la segmentation IGN comme information de base pour la production de la carte des physionomies pour les différents milieux garantit une interopérabilité géométrique. Cette dernière facilite également la constitution de la carte des physionomies assemblée (assemblage de tous les milieux).

Réalisation des masques

Dans la suite de ce document, on considère la production des différentes couches d'information géographique à l'échelle d'un département.



La segmentation et le programme PYRAM : l'assemblage de la carte des physionomies

L'ensemble des masques produits sont destinés à être soustraits à la segmentation vectorisée afin de faciliter la manipulation de cette dernière pour la classification des milieux ouverts. Des masques sont produits pour les occupations du sol suivantes :

- infrastructures de transport (le masque est réalisé par l'IGN à partir de données issues de la BD TOPO®) ;
- surface en eau (le masque est directement extrait de la BD TOPO®) ;
- bâtis (le masque est réalisé par les équipes du CEREMA Sud-Ouest et EVS-ISTHME sur la base de données issues pour partie de la BD TOPO® et pour partie de la BD MAJIC de la DGFIP) ;
- forêt fermée (le masque est directement extrait du fond forestier de l'IGN) ;
- vergers (le masque est issu de la BD-TOPO®) ;
- vignes (le masque est directement issu de la BD-TOPO®).

Après soustraction des masques infrastructures de transport, surfaces d'eaux, bâti et forêt, les surfaces restantes sont théoriquement des milieux ouverts. Afin d'obtenir la carte des physionomies assemblée finale, on ajoute à ces milieux ouverts, le domaine de la forêt fermée issu du fond forestier.

Masque des infrastructures de transport

Le masque des infrastructures de transports (routes et voies ferrés) sera, à terme, issu d'un autre projet de l'IGN : l'occupation du sol à grande échelle ou OCS GE.

Ce projet a pour but de répondre aux lois de 2010 portant engagement pour l'environnement (dites Grenelle 2 et Loi de modernisation de l'agriculture LMAP). Cette couche d'occupation du sol à grande échelle a pour vocation le suivi de la consommation d'espaces, notamment de l'artificialisation des sols et de l'évolution des surfaces, urbaines, agricoles ou forestières, et vise la production d'indicateurs fiables, homogènes et comparables sur l'ensemble du territoire. L'IGN en tant qu'institution nationale doit être garant du continuum national tout en assurant un emboîtement local/régional/national potentiel.

Le projet OCS GE s'inscrit dans le sillage des travaux conduits par le groupe de travail mandaté par le ministère chargé de l'écologie et auquel l'IGN participe, pour la définition d'une nomenclature nationale d'occupation du sol et de préconisations de bonnes pratiques pour la production de données d'occupation du sol (groupe de travail OCS GE du CNIG : http://cnig.gouv.fr/?page_id=8069).

Cette couche OCS GE doit couvrir à terme l'ensemble du territoire métropolitain ainsi que les départements d'outre-mer. Plus d'informations sont disponibles en ligne à l'adresse : <http://professionnels.ign.fr/ocsge>.

La couche des réseaux de transport de l'OCS GE « ossature », est constituée à partir des tronçons qui composent le réseau routier principal (liaisons entre métropoles, entre départements et entre villes, essentiellement via les autoroutes, les routes nationales et les routes départementales) et de ceux qui composent les voies ferrées en service (voir <http://cnig.gouv.fr/wp-content/uploads/2014/12/PNOCSGE-10-d%C3%A9cembre-20141.pdf>). Ces tronçons sont convertis en surfaces par application d'un tampon (buffer) de taille variable en fonction de l'importance de la voie. Les éléments surfaciques contenus dans la BD TOPO® sont ajoutés (surfaces de routes) pour créer l'ossature de l'OCS GE. Un enrichissement est ensuite pratiqué, des tronçons de route



La segmentation et le programme PYRAM : l'assemblage de la carte des physionomies

supplémentaires seront ajoutés en utilisant également les spécifications de l'ossature de l'OCS GE (annexe).

Pour les départements qui ne disposent pas d'une ossature OCS GE, une couche spécifique est réalisée avec les mêmes spécifications. Toutefois, en raison de contraintes matérielles et temporelles, cette couche ne subit pas les traitements suivants qui sont appliqués à l'ossature de l'OCS GE :

- comblement des espaces vides inférieurs à 10 m de large entre réseaux. L'espace est comblé par de la surface asphaltée (routes). Les voies ferrées ne sont pas élargies sauf pour combler des espaces entre deux tronçons de voies ferrées ;
- les giratoires et tous les terrains situés entre deux tronçons routiers dont la surface est inférieure à 500 m² sont transformés en surface asphaltée.

Masque des zones bâties

Le masque des zones bâties utilisé pour la réalisation de la carte des physionomies, est issu des travaux menés par les équipes du CEREMA Sud-Ouest et d'EVS-ISTHME. À terme, c'est le poste « CS1.1 – Surfaces anthropisées » de l'OCS GE qui sera utilisé.

Ce masque des zones bâties est élaboré à partir de deux bases de données complémentaires :

- les bâtiments et les surfaces anthropisées de la BD TOPO® ;
- la base de données cadastrales MAJIC (DGFIP).

Fichiers fonciers (communément appelés fichiers MAJIC)

Il s'agit d'une base de données cadastrales qui se compose du plan et de la documentation littérale. Cette base est issue des données MAJIC (Mise À Jour des Informations Cadastreales), nom de l'application interne à la DGFIP (Direction Générale des Finances publiques) pour gérer et calculer l'impôt foncier. Ces informations sont mises à jour en continu dans tous les CDIF (Centre des Impôts Fonciers) de France. La source de ces fichiers est fiscale et les informations majoritairement déclaratives. Ce sont principalement les services de l'État, les collectivités et les organismes chargés d'une mission de service public qui ont accès à cette base de données, sur demande et sous conditions. Cette base nécessite par ailleurs un lourd pré-traitement avant usage.

Pour l'élaboration du masque urbain, les parcelles sont filtrées pour ne sélectionner que celles qui ont au moins un bâtiment à l'hectare.

Le processus d'élaboration du masque se déroule en plusieurs étapes.

La première étape consiste à croiser la couche BD TOPO® bâtiments et surfaces d'activité avec la couche de segmentation réalisée par l'IGN, considérant qu'un polygone de la segmentation est classé urbain quand il comprend au moins 1 % de sa surface bâtie et quand le centroïde d'un polygone de la segmentation intersecte une surface « anthropisée » de la BD TOPO®.



La segmentation et le programme PYRAM : l'assemblage de la carte des physionomies

La deuxième étape a pour objet de croiser la couche de segmentation avec les « parcelles bâties » des fichiers fonciers. Dans ce cas, tout polygone de la segmentation intersectant une parcelle bâtie et dont la surface est occupée par au moins 40% de la parcelle bâtie est considéré comme « Urbain ».

Catégories sélectionnées de la BD TOPO® pour l'élaboration du masque urbain

Couche des bâtiments :

Bâti_indifférencié
Bâti_industriel
Bâti_remarquable
Réservoir
Construction_surfacique
Construction_légère
Gare
Poste_transformation

Couche des surfaces « anthropisées » :

Piste_aérodrome
Terrain_sport
Cimetière
Surface_route (attributs « Péages » et « Parking »)
Aire_triage
Surface_activité

Point de vigilance :

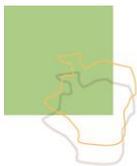
Les seuils de classement en catégorie urbaine d'un polygone de la segmentation dont la couche du bâti BD TOPO® représente 1 % de sa surface, ou 40 % pour le fichier foncier, sont des seuils plus ou moins arbitraires appliqués de façon empirique et non statistique. Ceux-ci doivent en tout état de cause être choisis après vérification minutieuse de leurs résultantes cartographiques et ne doivent surtout pas être appliqués *in extenso* pour tous les départements.

Masque des forêts

Le masque forêt est issu du domaine de la forêt fermée de la BD Forêt V2. En 2015, la BD Forêt V2 couvre près des 2/3 du territoire métropolitain (61 départements réalisés sur 96). L'existence, d'une cartographie des formations forestières d'une telle ampleur, a logiquement conduit à intégrer cette donnée très en amont dans le processus de réalisation de la pré-cartographie. La carte forestière est donc utilisée dans un premier temps en tant que masque et permet de simplifier la segmentation de base (suppression des polygones de la segmentation compris dans les polygones de la carte forestière). Cette étape est d'autant plus importante que les forêts présentent une très forte hétérogénéité de surface (succession de houppiers, d'ombres...). Cela engendre un découpage très important de l'orthophoto par le processus de segmentation utilisé, y compris au niveau le plus grossier (niveau 1).

Les espaces décrits par la BD Forêt V2 sont essentiellement les espaces naturels et semi-naturels : forêts fermées ; forêts ouvertes ; peupleraies ; landes (maquis, garrigues...) ; formations herbacées (pelouses alpines, pelouses pastorales...). Les objets présents dans la base sont définis par une surface supérieure ou égale à 5 000 m² (50 ares), selon les seuils suivants :

- exclusion des surfaces dont l'utilisation du sol est exclusivement agricole ;
- largeur de l'objet d'au moins 20 mètres ;
- taux de couvert de la végétation (arborescente et arbustive) supérieur ou égal à 10 %.



La segmentation et le programme PYRAM : l'assemblage de la carte des physonomies

Ces seuils proviennent de la définition internationale de la forêt, en vigueur à l'heure actuelle (source FAO 2004, Global forest resources assessment update 2005: Terms and definitions. FRA working paper 83/E, FAO, Rome.).

La discrimination des essences présentes en forêt se fait ensuite selon les règles suivantes :

- trois classes de taux de couvert des arbres capables d'atteindre une hauteur supérieure à 5 mètres à maturité *in-situ* : inférieur à 10 % (non forêt), entre 10 et 40 % (forêt ouverte), supérieur ou égal à 40 % (forêt fermée) ;
- un seuil de 75 % de couvert des arbres ayant accès à la lumière pour distinguer un peuplement en composition pure ou en essence pure.

Une description précise des postes de légende de la BD Forêt V2 est disponible dans le « Descriptif de contenu de la BD Forêt V2 » sur le site de l'IGN :

(http://professionnels.ign.fr/sites/default/files/DC_BDFORET_v2.pdf).

Principes de réalisation de la BD Forêt V2 :

La première étape de réalisation de la BD Forêt V2 s'appuie sur l'utilisation de la BD ORTHO[®] en infrarouge fausses couleurs et est commune à la production du thème végétation de la BD TOPO[®]. Il s'agit d'une phase semi-automatique qui utilise notamment des outils d'analyse d'images et qui se base sur la segmentation IGN. C'est lors de cette première étape, que sont saisies, à partir des photos aériennes, les limites de la forêt et des landes. De même sont saisis les autres éléments arborés constitutifs de la couche végétation de la BD TOPO[®] : les bosquets (appelés bois dans la nomenclature BD TOPO[®]), les haies et les vergers. Ces éléments arborés ne sont pas pris en compte dans la suite du processus de réalisation de la BD Forêt V2, qui ne s'intéresse à l'espace forestier qu'au sens de la définition internationale de la forêt. Ils seront néanmoins extraits de la BD TOPO[®], pour une réutilisation dans le cadre de la production cartographique (couche « Haies »)

La deuxième étape permet d'affiner et de préciser l'information (re-précision de tracés en conformité avec les définitions, distinction des essences) par photo-interprétation. La création de ces couches d'informations repose donc sur l'interprétation de photographies aériennes infrarouges d'une résolution de 50 centimètres. Les statistiques de l'inventaire forestier ainsi que des vérifications terrain améliorent la fiabilité et la qualité globale de la base de données.

Le produit final est mis en cohérence avec les réseaux routiers, hydrographiques et ferrés de la BD TOPO[®] (limite ajustée à l'axe des réseaux). Cette démarche a fait l'objet d'une attention particulière, notamment en lien avec le programme OCS GE de l'IGN.

Le descriptif de contenu de la BD Forêt V2 aborde notamment les questions de sémantique, de référentiels cartographiques et d'attribut.

Traitements de la segmentation et des masques

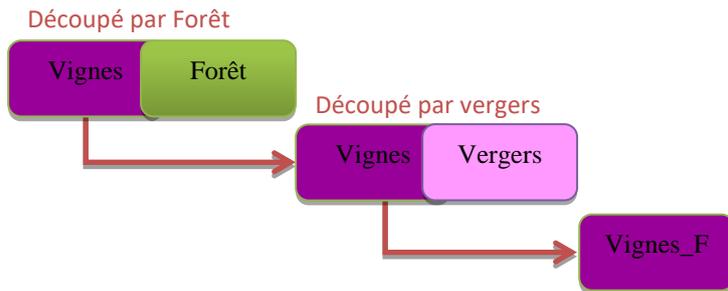
De façon à assurer la cohérence entre les contours de la segmentation et les différents masques ainsi que la cohérence topologique entre les masques, différents traitements sont réalisés.



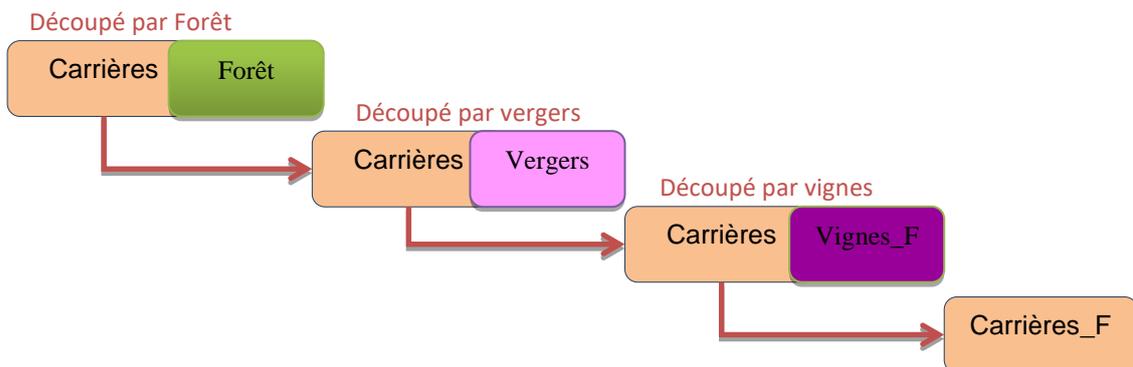
La segmentation et le programme PYRAM : l'assemblage de la carte des physionomies

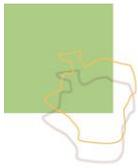
Ces traitements consistent notamment en un découpage des couches les unes par rapport aux autres. Ainsi, tous les masques (bâti, forêt, vergers et vignes) sont découpés par le masque des infrastructures de transport (routes et voies ferrées) et par les surfaces en eau. Les autres traitements pour chacun des masques se déroulent de la façon suivante :

Masque vigne :



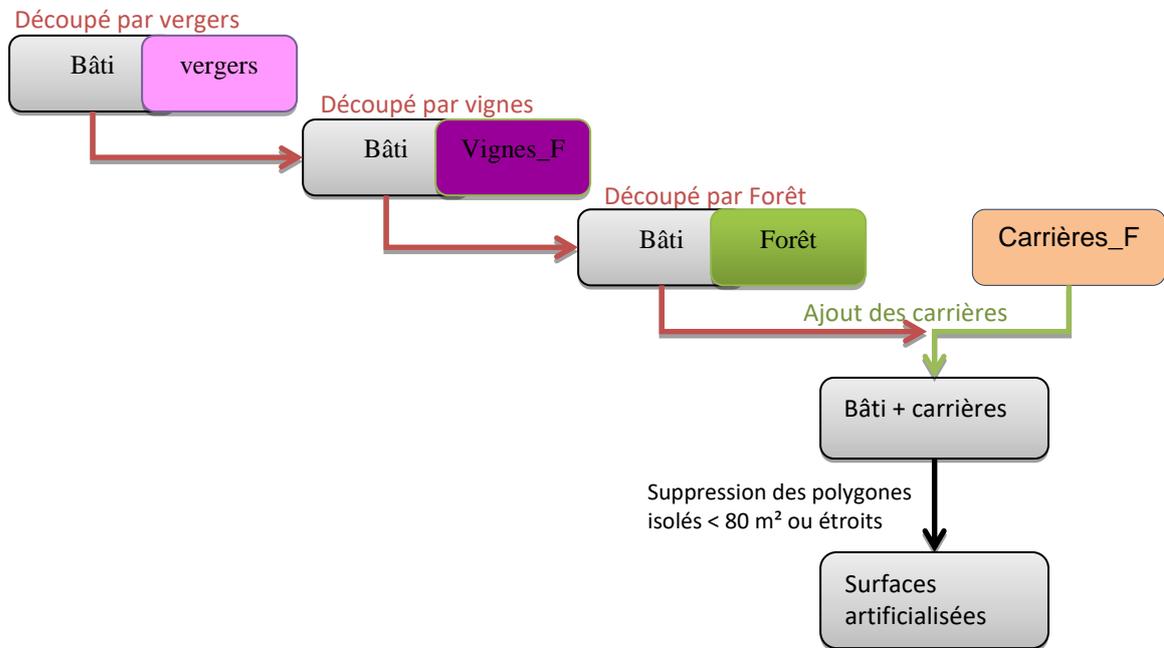
Préparation des carrières :



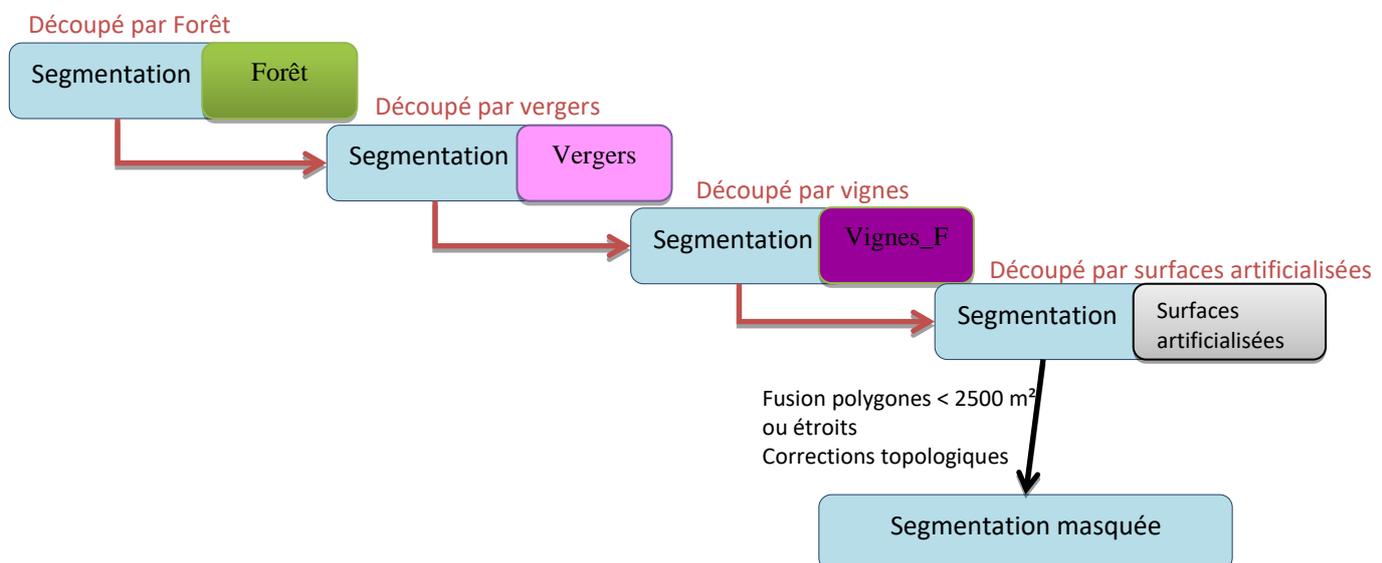


La segmentation et le programme PYRAM : l'assemblage de la carte des physionomies

Masque bâti :



Segmentation :



Le processus de découpage de la segmentation par les masques privilégie les données obtenues par photo-interprétation par rapport aux données issues de traitements automatiques. Suite à l'ensemble des traitements, la segmentation masquée est envoyée aux organismes assurant la qualification des polygones (segments) par télédétection. Elle sera ensuite assemblée avec la forêt issue du fond forestier et les surfaces en eau pour constituer la carte des physionomies.



La segmentation et le programme PYRAM : l'assemblage de la carte des physionomies

Assemblage de la carte des physionomies

Après l'étape de télédétection, les milieux ouverts sont assemblés avec les masques des forêts fermées et les surfaces en eau. Quelques post-traitements supplémentaires sont effectués à cette étape :

- Les surfaces d'eau (hors bassin) d'une superficie inférieure à 2 500 m² sont ponctualisées, les polygones sont fusionnés avec le voisin avec lequel ils partagent la plus grande frontière dans la carte des physionomies. La couche de ponctuels ainsi créée contient toutes les surfaces en eau hors bassin de moins de 2 500 m². Les informations sémantiques de la couche de ponctuels renseignent le régime des eaux (permanent ou intermittent) et la superficie de la surface d'eau correspondant au polygone d'origine.
- Les surfaces d'eau correspondant à des bassins (bassin de rétention, d'épuration, pisciculture...) sont basculées dans la couche des surfaces artificialisées.
- Les polygones de segmentation classés dans la catégorie « artificialisés » par le processus de la classification automatique des milieux ouverts sont basculés dans la couche surface artificialisée (cela se produit lorsque des bâtiments ne sont pas présents dans la BD Topo).

À noter : les polygones voisins qualifiés de manière identique ne sont pas fusionnés sauf dans le cas de la forêt fermée. Ce choix a été retenu afin d'accélérer le processus de cartographie dans la mesure où il est plus simple de fusionner des polygones existant pour obtenir le contour d'une UVC que de découper un polygone. Ceci est particulièrement vrai lors de la phase de cartographie *in situ*.

Au final, 4 couches d'information sont accessibles:

- La carte des physionomies ;
- la couche départementale des surfaces artificialisées ;
- la couche départementale des voies de communication (routes et voies ferrées) ;
- une couche départementale d'information ponctuelle contenant les points des surfaces d'eau de moins de 2 500 m².

Rédaction : Loïc Commagnac, Bénédicte Maisonneuve (IGN)



La segmentation et le programme PYRAM : l'assemblage de la carte des physionomies

Annexe : Spécifications de l'ossature de l'OCS GE

Pour les tronçons routiers, les critères de « buffer » sont les suivants:

Critère	Taille du « buffer » (en mètres)
Largeur de chaussée supérieure à 5 mètres	Largeur/2 + 1
Largeur de chaussée inférieure à 5 mètres et nombre de voies supérieur ou égale à 2	Nombre de voies x 1,75
Largeur de chaussée inférieure à 5 mètres et nombre de voies non renseigné	2,5

Pour les tronçons ferrés, les critères de « buffer » sont les suivants :

Critère	Taille du « buffer » (en mètres)
1 voie	3,5
2 voies	5,5
3 voies	8,5
4 voies	10,5
Nombre de voies non indiqué	2,5

(les critères ont été évalués en fonction des normes RFF d'éloignement des voies)