

Note méthodologique sur la construction des unités urbaines

1 La méthode de calcul des unités urbaines

La notion d'unité urbaine repose sur la **continuité du bâti** et le **nombre d'habitants**. Le calcul des unités urbaines se déroule en quatre grandes étapes, qui sont détaillées dans la suite :

1. la création des taches de bâti continu ;
2. le calcul de la population des taches et la sélection des taches de plus de 2 000 habitants ;
3. la création des unités urbaines proprement dites par passage des taches aux communes ;
4. Une vérification manuelle de la phase automatique afin de tenir compte de la stabilité temporelle des concepts et des imperfections locales des données utilisées.

Dans ce qui suit, seule la phase automatique avant vérification manuelle est explicitée.

1.1 Création des taches de bâti continu

La construction des taches de bâti continu en France est réalisée par un traitement automatique des données sur le bâti issues de la BD TOPO¹ de l'Institut national de l'information géographique et forestière (IGN). Tous les bâtiments à usage d'habitation sont considérés pour réaliser les taches de bâti continus. D'autres bâtiments sont également pris en compte comme certains espaces publics (cimetières, stades, aéro-dromes...) et les terrains industriels et commerciaux (usines, zones d'activités, centre commerciaux...) qui ont été traités comme des habitations avec la règle des 200 mètres pour relier des zones de construction habitées. De plus, la construction des taches tient compte des coupures du tissu urbain par des cours d'eau en l'absence de ponts.

Pour l'hydrographie, on ne conserve que les surfaces hydrographiques qui ont une intersection avec un tronçon hydrographique se trouvant au niveau du sol (qui ne se trouve donc ni sur un pont ni dans un tunnel) ou dont la position est inconnue. Cette méthode permet d'obtenir des surfaces correspondant à des cours d'eau en surface suffisamment larges pour ne pas pouvoir être traversés à pied. Certaines surfaces hydrographiques non sélectionnées sont des grands lacs ou étangs non traversés par une rivière mais qui coupent déjà les taches du fait de leur taille.

En ce qui concerne les ponts, on ne conserve que ceux qui croisent un des éléments hydrographiques sélectionnés. En effet, dans la BD TOPO certains ponts sont des ponts de routes enjambant d'autres routes et ce type de ponts ne permet pas de créer de continuité de bâti. Cependant, comme on ne dispose pas de la nature détaillée de ce que supporte le pont, il est possible qu'on prenne en compte par exemple des ponts qui ne supporte qu'une voie ferrée (et qui donc ne nous intéressent pas pour la continuité du bâti) ou même le cas extrême d'un aqueduc passant au-dessus d'un cours d'eau au sol. Dans certains cas nous disposons d'une information un peu plus détaillée sur la nature du pont, ce qui nous permet d'exclure une partie des aqueducs et ponts canaux.

1. "La BD TOPO® est une description vectorielle 3D (structurée en objets) des éléments du territoire et de ses infrastructures, de précision métrique, exploitable à des échelles allant du 1 : 2 000 au 1 : 50 000.", <http://professionnels.ign.fr/bdtopo>

Au niveau des bâtiments, une exclusion des serres agricoles s'est avérée nécessaire dans la mesure où on constatait que pour des éléments semblables (des serres agricoles en plastique), certaines étaient décrites dans la BD TOPO comme des éléments légers tandis que d'autres non. Cela amenait à prendre en compte des serres en plastique dans la constitution de certaines unités urbaines et dans certains cas à les étendre artificiellement.

Fonctionnement général de l'algorithme Avec les moyens informatiques dont nous disposons il est facile de créer des taches de bâti continu à partir de la description géométrique des bâtiments. Chaque bâtiment est modélisé par un polygone pour lequel les coordonnées géographiques de chaque sommet sont connues. Le logiciel R, grâce par exemple au package *sf*, permet de manipuler ce genre d'objets géographiques facilement. Ainsi pour créer des taches de bâti continu à partir de la donnée des polygones formant les bâtiments il suffit de calculer la zone tampon de 100m autour de chaque bâtiment et de fusionner les zones tampons qui se recoupent. Ces opérations élémentaires de création de zones tampons, fusion de polygones et détection de polygones se recoupant sont implémentées par des fonctions du package *sf*. La figure 1 donne une illustration de l'enchaînement de ces opérations de base.

La véritable difficulté réside ici dans le volume des données à traiter. Les opérations élémentaires décrites ci-dessus deviennent très coûteuses en ressources computationnelles lorsque le nombre d'éléments à considérer augmente. En réalité c'est surtout l'opération de détection des zones tampons qui se recoupent qui est la plus coûteuse de toute puisque le nombre de calculs à effectuer et la mémoire vive nécessaire sont proportionnels au carré du nombre d'éléments à traiter. Avec les ressources informatiques à disposition et en utilisant le package R il est impossible de traiter l'ensemble du territoire français d'un seul coup. C'est pourquoi on a mis en place une approche **itérative par échelle croissante** : il s'agit de réaliser les calculs sur des petites zones géographiques de façon indépendantes² et de recombinaison ensuite les résultats entre eux pour avoir les résultats sur une zone plus grande, et ainsi de suite jusqu'à avoir traité tout le territoire.

On crée les taches de bâti continu par itération sur des zones de plus en plus grandes : carreaux de taille croissante, départements, régions et enfin France entière. On construit d'abord, pour un département donné, les taches pour chaque petit carreau et on fusionne les taches qui se touchent aux limites de ces carreaux (voir la figure 2). On obtient ainsi les taches au niveau du département. De la même façon, pour une région, on crée les taches sur tous ses départements et on fusionne sur les bords. Enfin pour obtenir les taches sur la France entière on calcule les taches par région et on fusionne les taches aux frontières de ces régions. On passe ici par le niveau administratif (départements et régions) car c'est ainsi que sont stockées les données de la BD TOPO.

Au sein d'une sous-zone particulière, les taches loin de la frontière ne peuvent pas recouper des taches d'une autre sous-zone. Au contraire les taches proches de la frontière sont susceptibles de fusionner avec d'autres taches de l'autre côté de la frontière. En se restreignant aux taches proches de la frontière on limite ainsi considérablement le nombre d'éléments à considérer pour le calcul de la fusion des taches. C'est en procédant de cette manière qu'il est possible de limiter à la fois la mémoire vive utilisée et le nombre de calculs à effectuer.

Compte tenu des différences de systèmes de projection, il faut réaliser séparément les calculs pour la France métropolitaine et les DOM.

La prise en compte des coupures par les cours d'eau

La définition des unités urbaines inclut la prise en compte de la coupure des taches de bâti continu par des éléments naturels : cours d'eau, gravières et dénivelé important. Seuls la coupure par des cours d'eau à ici été gérée de façon automatique. Il s'agit de ne pas relier dans une même tache deux bâtiments distants

2. Ceci rend les calculs parallélisables mais nous n'avons pas exploité cette possibilité.

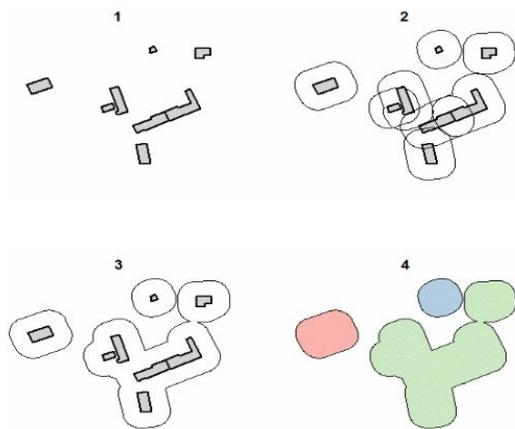


FIGURE 1 – Construction des taches de bâti continu à partir des contours de bâtiments.

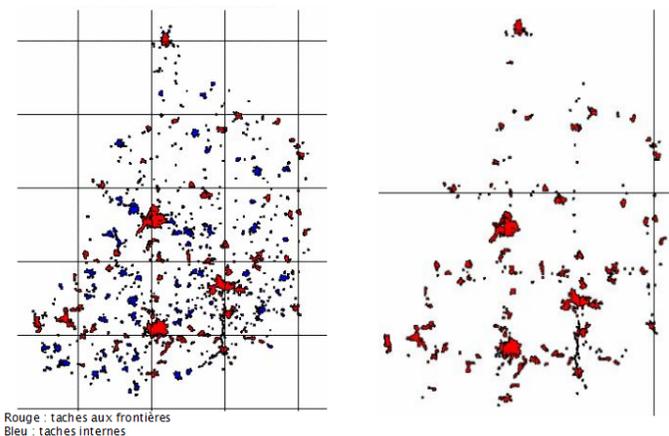


FIGURE 2 – Approche par carreaux d'échelle croissante

de moins de 200m l'un de l'autre mais situés de part et d'autre d'un cours d'eau en l'absence de pont sur ce cours d'eau permettant de relier les deux bâtiments (voir figure 3).

Pour cela, la solution choisie consiste à :

1. créer les taches en intégrant les ponts dans le bâti (seulement les ponts traversant un cours d'eau) ;
2. soustraire la surface des ponts à celle des rivières ;
3. et soustraire aux taches les éléments obtenus en 2.

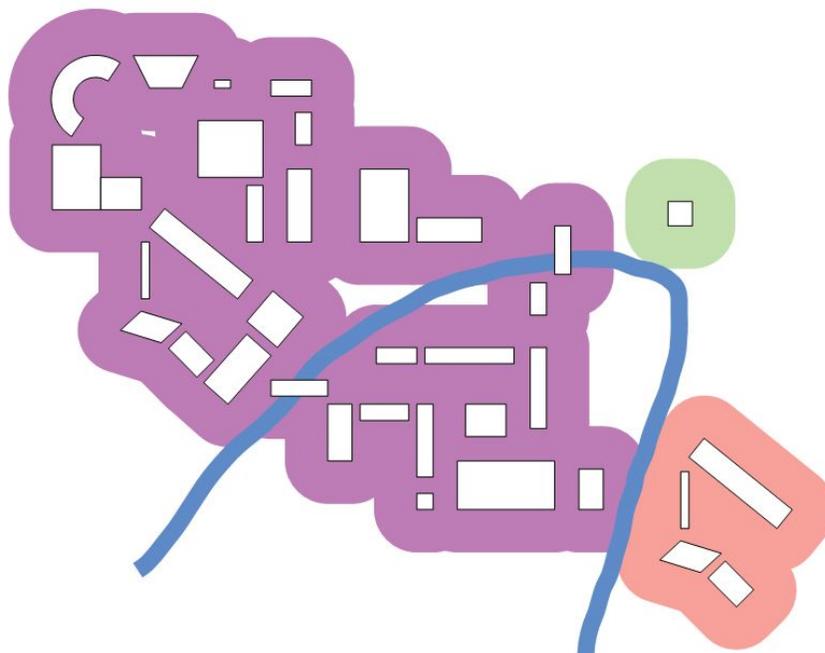


FIGURE 3 – Prise en compte des rivières et des ponts

Note : La zone bleue représente la rivière. Les polygones blancs représentent les bâtiments (y compris les ponts). Les taches de bâti continu sont représentées en violet, vert et rose clair. La tache rose est séparée de la tache violette car bien que les bâtiments soient distants de moins de 200 mètres de part et d'autre de la rivière, aucun pont n'est présent à moins de 200 mètres pour les relier.

Les taches internationales

Les données utilisées pour construire les *taches internationales* (qui s'étendent de part et d'autre d'une frontière nationale) sont celles d'OpenStreetMap³ téléchargées à partir du site Geofabrik⁴.

Comme pour la BD TOPO, seuls certains éléments cartographiques d'OpenStreetMap sont utilisés pour les calculs. Nous avons conservé les éléments qui semblaient les plus proches des éléments utilisés dans la BD TOPO. Nous avons donc utilisé :

- tous les éléments du thème "buildings". Cela correspond à des bâtiments individuels ou des groupes de bâtiments connectés.
- les éléments de type cimetière dans le thème "landuse".
- les éléments classés comme des ponts dans le thème "roads".
- les éléments de type "water" (lac, étang, mare, réservoir, lit d'un cours d'eau, etc.), "reservoir" (bassin de retenue) et "river" (les rivières les plus large, supérieur a 12m) dans le thème "water".

Pour éviter d'avoir un trop grand volume de données et donc des temps de calculs trop longs, on ne sélectionne que des éléments se trouvant dans des zones frontalières (pays frontaliers et même dans certains cas des régions de ces pays) : Andore, la Belgique, la Suisse, Monaco, le Luxembourg, l'Espagne, le Nord-Ouest de l'Italie, et les lands de Sarre, Bade-Wurtemberg et Rhénanie-Palatinat en Allemagne.

À partir des données sélectionnées, l'algorithme fonctionne de la même façon que pour les taches construites sur le territoire français, mais sans les itérations sur les départements et les régions, c'est-à-dire qu'on ne fait que des itérations sur les carreaux de taille croissante. Une fois les taches calculées en France et à l'étranger, on fusionne les taches et on ne conserve que celles qui ont une partie en France.

1.2 La population des taches

A l'issue de l'étape précédente, on dispose des taches bâti continu (règle des 200 mètres entre chaque bâtiment) s'étendant sur le territoire français ainsi que de celles qui s'étendent de part et d'autre des frontières nationales. Comme on l'a vu dans la définition, il faut également appliquer à ces taches un critère de population (2 000 habitants dans la tache). La population des taches est calculée en additionnant les population de la partie française et de la partie étrangère (lorsque la tache s'étend au-delà de la frontière). Pour la partie française, la population est mesurée à l'aide des fichiers démographiques sur les logements et les individus (Fidéli) qui permet de calculer le nombre d'individus géolocalisés qui se trouvent dans la tache. Pour la partie étrangère, on utilise le fichier GEOSTAT 2011⁵ qui contient des informations carroyées de population nationales provenant de 29 pays membres de l'UE et de l'AELE (Association européenne de libre-échange) dans une grille de carreaux de 1 km de côté. Dans le cas de cette source, comme on ne dispose pas de la localisation précise de la population, on va estimer la population en supposant qu'elle se trouve entièrement dans l'intersection entre les carreaux et les taches puisque c'est là que se trouvent les bâtiments. On calcule alors les populations comme indiqué sur le figure 4

3. OpenStreetMap® est une base de données géographique. C'est "un ensemble de données ouvertes, disponibles sous la licence libre Open Data Commons Open Database License", <https://www.openstreetmap.org/copyright>

4. <http://www.geofabrik.de/geofabrik/geofabrik.html>

5. "L'initiative GEOSTAT a été prise conjointement par Eurostat et les instituts nationaux de statistique afin d'établir une infrastructure de données et de production pour les statistiques géospatiales. [...] GEOSTAT a comme objectif principal d'aider les INS à mettre en place leurs données, méthodes et systèmes de production pour parvenir à un recensement entièrement géocodé en 2021. ", <https://ec.europa.eu/eurostat/fr/web/gisco/gisco-activities/integrating-statistics-geospatial-information/geostat-initiative>

<https://ec.europa.eu/eurostat/fr/web/gisco/geodata/reference-data/population-distribution-demography/geostat> pour le téléchargement des données.

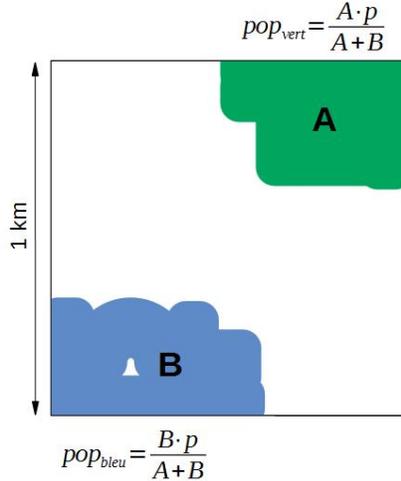


FIGURE 4 – Estimation de la population des taches avec la grille de densité

Note : Les zones bleues et vertes correspondent à des taches de bâti continu. A correspond à l'aire de la surface verte et B à l'aire de la surface bleue, et p est la population du carreau englobant de 1 km de côté. On calcule pop_{vert} , la population de la tache verte et pop_{bleu} , la population de la tache bleue.

1.3 Passage des taches aux communes et composition communale des unités urbaines

Composition des unités urbaines

On appelle *zone agglomérée* une tache de bâti continu (pas de coupure de plus de 200 mètres entre deux constructions) qui compte plus de 2 000 habitants. Et on appelle *fraction communale* la partie d'une zone agglomérée se trouvant dans une même commune (une zone agglomérée peut en effet s'étendre sur plusieurs communes). Les communes appartenant à une unité urbaine sont alors :

1. celles qui contiennent entièrement une zone agglomérée ;
2. celles qui ne sont pas dans le cas précédent et dont une des fractions comprend plus de 50 % de sa population ;
3. celles qui ne sont pas dans l'un des deux cas précédents et dont au moins une des fractions compte plus de 2 000 habitants.

Dans le premier et le dernier cas, la commune constitue à elle seule une unité, c'est ce qu'on appelle une *ville isolée*. Dans le deuxième cas, la commune ainsi que toutes les communes qui sont dans le même cas et qui partagent la même tache constituent une *agglomération multicommunale*. On ne conserve ensuite que les unités dont la population se trouvant dans la zone agglomérée est supérieure à 2 000 habitants.

Selon cette procédure, il se peut alors qu'une ville isolée partage la même tache de bâti continue qu'une unité urbaine multicommunale, sans pour autant en faire partie. De plus, il se peut que la majorité de la population d'une commune dite urbaine se trouve hors d'une zone agglomérée.

Correction des unités urbaines non-connexes

A l'issue du processus décrit ci-dessus, on peut obtenir des unités urbaines non-connexes, c'est-à-dire composées de plusieurs ensembles non contigus de communes. Cela n'est en général pas souhaitable. Une unité urbaine non-connexe peut par exemple apparaître lorsque la jonction entre une commune et le reste de l'unité se fait via une commune isolée ou non urbaine partageant la même zone agglomérée, ou lorsque deux communes sont séparées par un bras de mer mais reliées par un pont. Ces deux cas sont différents, dans

le premier il semble plus pertinent de scinder l'unité en deux, tandis que dans le deuxième on souhaiterait plutôt conserver l'unité telle quelle (voir les figures 5 et 6).

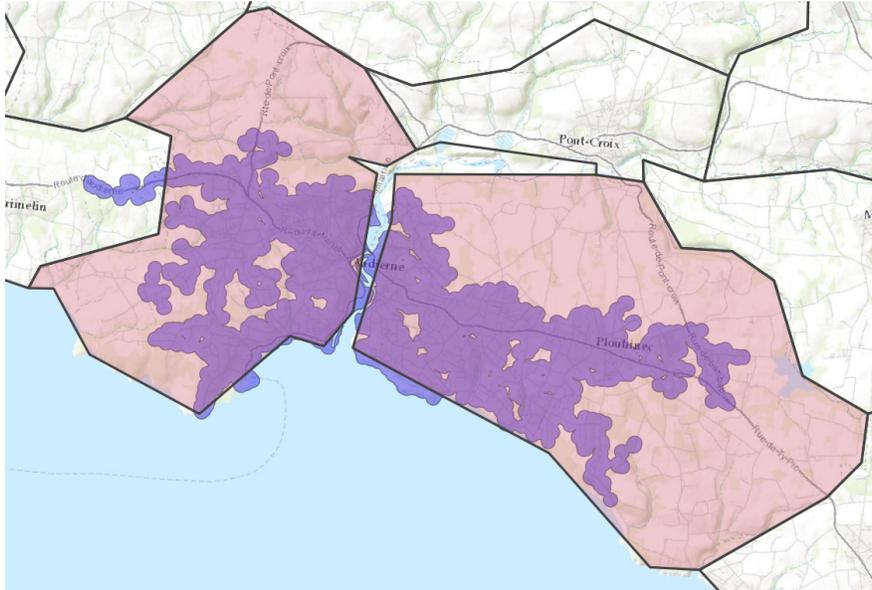


FIGURE 5 – Exemple d'une unité non contiguë conservée.

Note : Il s'agit de l'unité de Plouhinec - Audierne (en rose sur l'image) composée des communes de Plouhinec et d'Audierne. Cette unité n'est pas contiguë car les deux communes se trouvent de part et d'autre de l'embouchure d'un fleuve (Le Goyen) mais leur tache (en violet sur l'image) est reliée par un pont qui passe au-dessus du Goyen.

Source : Esri World Topographic Map pour fond de carte.

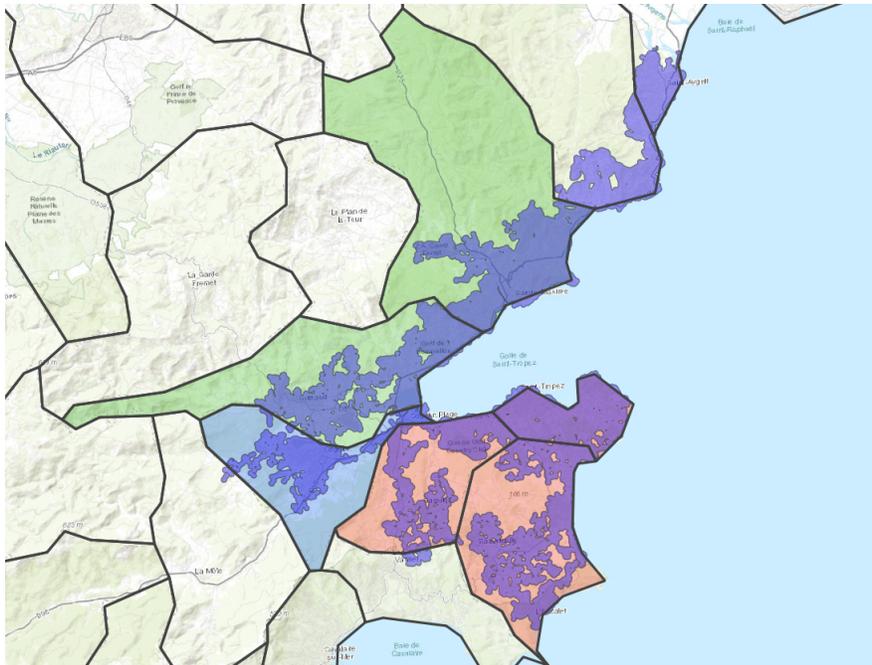


FIGURE 6 – Exemple d'une unité non contiguë séparée en deux.

Note : Il s'agit des unités urbaines des Saint-Tropez (en rose) et de Sainte-Maxime (en vert). Ces unités partagent la même tache mais sont séparées par l'unité de Cogolin (en bleu) qui est isolée.

Source : Esri World Topographic Map pour fond de carte.

Un algorithme permet de repérer de façon automatique les unités urbaines non connexes. Pour cela, on construit pour chaque unité multicommunale, un graphe de ses communes reliant les communes qui se touchent. On détermine ensuite le nombre de composantes connexes du graphique et si celui-ci est supérieur à un alors l'unité est non connexe. Les unités ainsi repérées sont alors analysées manuellement afin de savoir s'il est souhaitable de scinder l'unité et de construire un fichier qui indique pour chaque commune à quel sous-ensemble de l'unité on souhaite l'affecter. On peut alors recalculer les unités urbaines (filtrer les nouvelles unités dont la population urbaine est inférieure à 2 000 habitants, recalculer les identifiants d'unités, etc.).

Détermination du type de commune urbaine

Une fois les unités urbaines constituées, on classe les communes en trois types : *ville-centre*, *banlieue* ou *ville isolée*.

Lorsqu'une unité urbaine est constituée de plusieurs communes, on la désigne sous le terme d'agglomération multicommunale. Les communes qui la composent sont soit ville-centre, soit banlieue. Si une commune représente plus de 50 % de la population de l'agglomération multicommunale, elle est seule ville-centre. Sinon, toutes les communes qui ont une population supérieure à 50 % de celle de la commune la plus peuplée, ainsi que cette dernière, sont villes-centres. Les communes urbaines qui ne sont pas villes-centres constituent la banlieue de l'agglomération multicommunale. Une agglomération multicommunale peut n'être constituée que de villes-centres.

Lorsqu'une unité urbaine est constituée d'une seule commune, on la désigne sous le terme de ville isolée.