

La production de données carroyées à l'Insee : aspects méthodologiques

Séminaire *Statistique et géographie*

Jean Finot – Ali Hachid
(Division Méthodes et référentiels géographiques - DMCSI)



Le carroyage

Plan de la présentation :

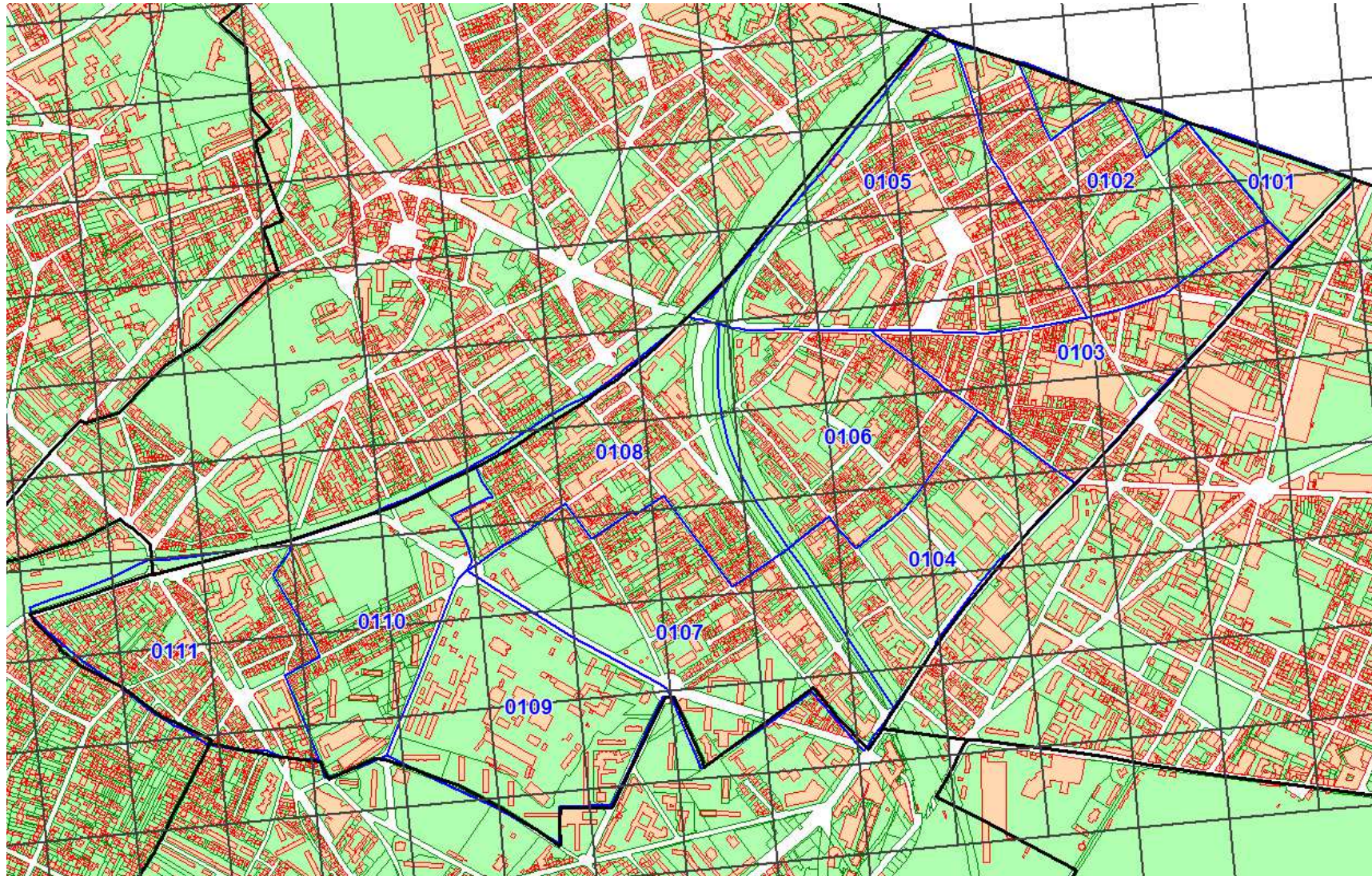
1. Généralités
2. Éléments de géodésie
3. Le plan cadastral
4. Comment carroyer
5. Du carreau au rectangle
6. Gestion du risque de rupture de la confidentialité
7. Les variables
8. Mise à disposition du carroyage

Généralités

Définition :

- Le carroyage est un découpage du territoire...
- ...en mailles régulières et fines.
- Ce maillage n'est donc aucunement lié à un quelconque autre découpage du territoire notamment administratif (communes)
- Il ne prend en compte aucune réalité du terrain (topographie)
- Traditionnellement, il peut s'agir de carreaux de 100, 200, 250, 500 mètres ou d'1km de côté
- A l'Insee, carroyage de la source fiscale sur des carreaux de 200 mètres

Exemple : carroyage de Malakoff



Généralités

Avantages :

- Découpage stable dans le temps
- Permet d'élaborer des diagnostics sur des territoires à façon
- Permet de cibler les zones d'intervention des politiques publiques (nouveaux contours de la politique de la ville)

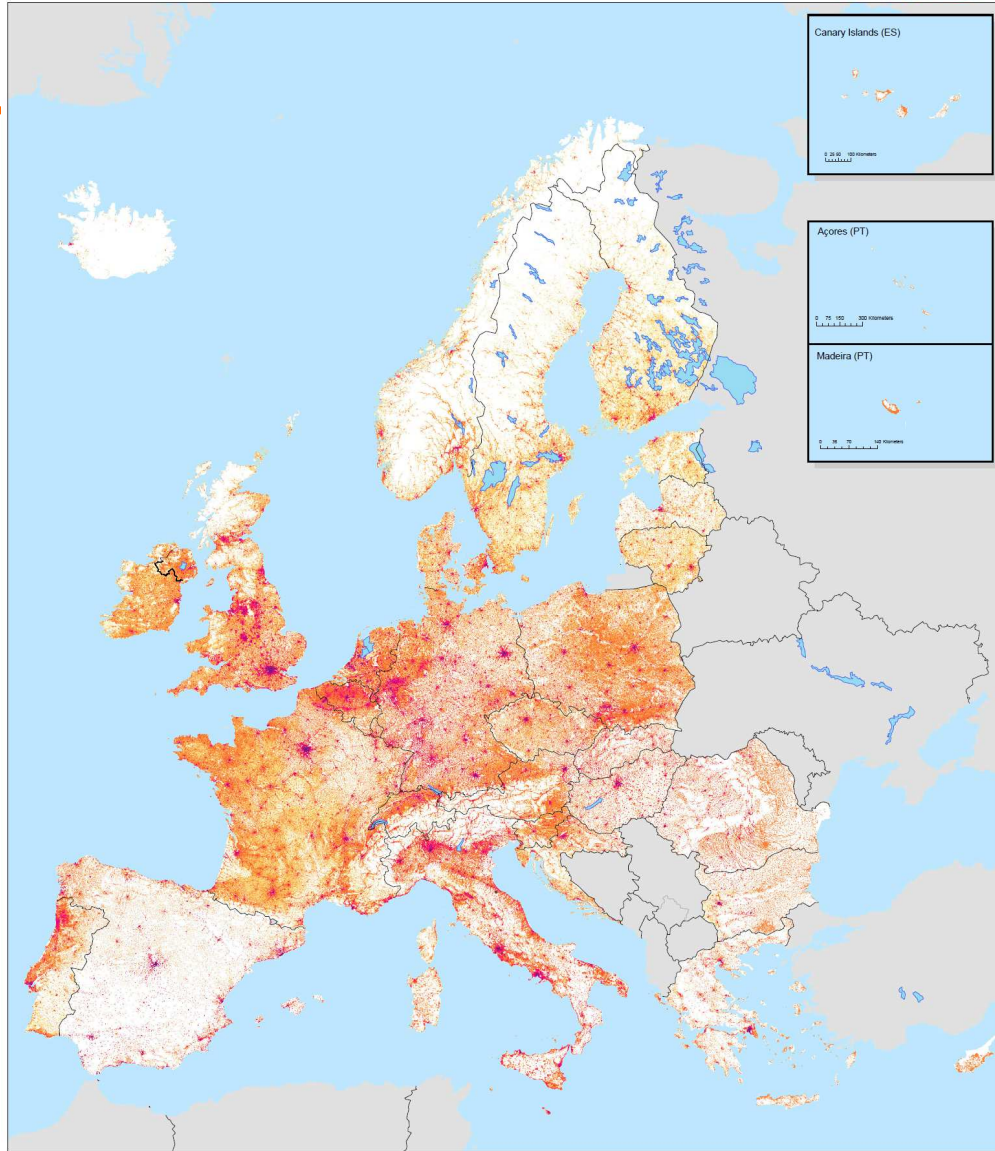
Cependant :

- Nécessite un processus de géolocalisation précis pouvant être lourd et coûteux
- Nécessite un traitement du secret statistique et du secret fiscal

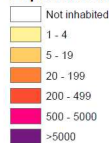
Généralités

Contexte international :

- Des initiatives pour promouvoir la production harmonisée de statistiques carroyées et en premier lieu les effectifs de population
- Au niveau européen, le projet GEOSTAT lancé en 2010 :
 - Renforcer l'intégration d'information géographique et statistique
 - Promouvoir le carroyage
 - Partager les expériences
 - Production d'un jeu de données carroyées sur les recensements européens de 2011



Population density



Data sources:

National grid data from census 2011 information, aggregated to 1km² grid cells:
Belgium, Bulgaria, The Czech Republic, Estonia, Finland, France, Germany, Ireland, Liechtenstein, Lithuania
The Netherlands, Norway, Poland, Portugal, Slovenia, Spain, Sweden, Switzerland.

Disaggregated data produced by the Austrian Institute of Technology (AIT):
Austria, Croatia, Cyprus, Denmark, Greece, Hungary, Iceland, Italy, Latvia, Luxembourg, Malta, Romania, Slovakia, United Kingdom.
Source data: population per LAU2 (2011) (© Eurostat), EEA Factbook Service Processor (© Eurostat), Europe as of 2009 (© European Environment Agency),
LAU2 administrative boundaries (© Eurogeographics), Corine land cover 2006 (© European Environment Agency), Open Street Map data 2011 (© Geotane © OpenStreetMap).

© Eurogeographics for the Administrative Boundaries
Map © European Forum for Geostatistics. Further information: <http://www.efgs.info>



Généralités

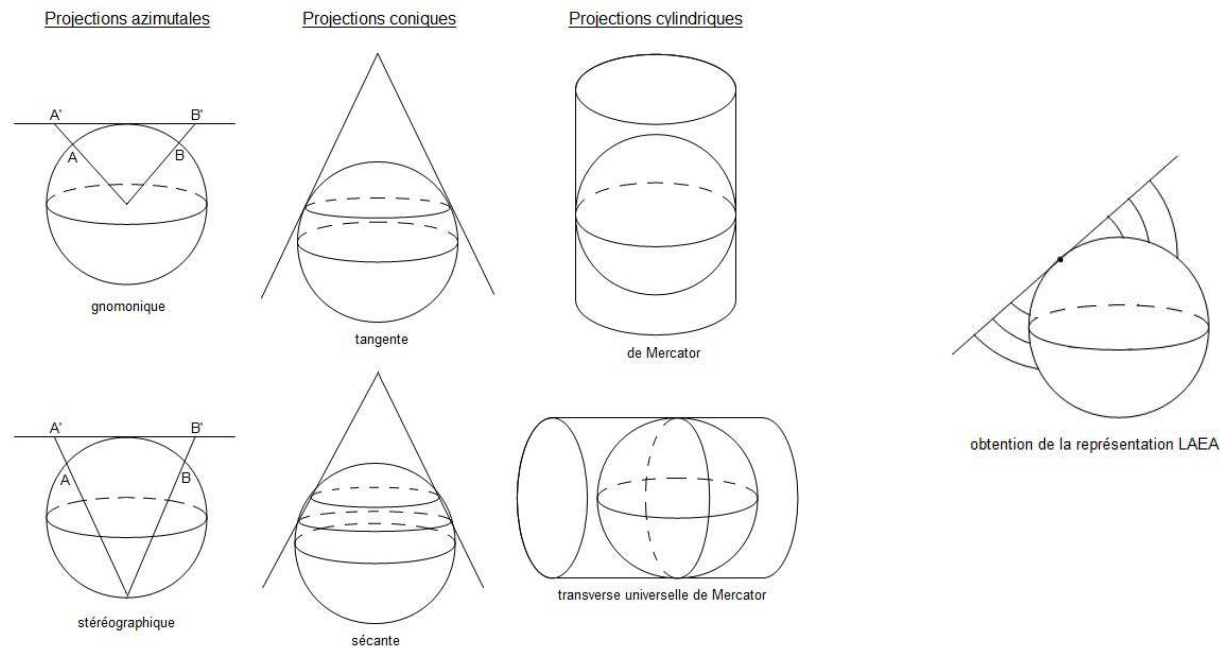
La matière première :

- Disposer (ou produire) des coordonnées géographiques pour chaque unité statistique
- Dans un même système de projection cartographique
- Pour cela, la source de données doit comporter des informations localisantes : adresse, parcelle cadastrale
- La source fiscale retravaillée (revenus fiscaux localisés - RFL) fournit la référence cadastrale de chaque local
- Le plan cadastral donne les coordonnées (en Lambert 93) de chaque parcelle (au centre)
- Processus moins coûteux qu'une géolocalisation par l'adresse

Éléments de géodésie

➤ Géodésie :

- science qui a pour objet l'étude de la forme, des dimensions et du champ de pesanteur de la Terre
- étudie les problèmes de représentation plane de la surface terrestre
- par projection sur une surface développable c'est à dire dépliable sur un plan (cône, cylindre, plan)

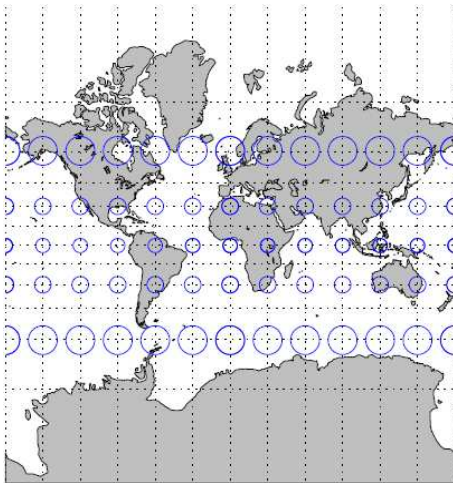


Éléments de géodésie

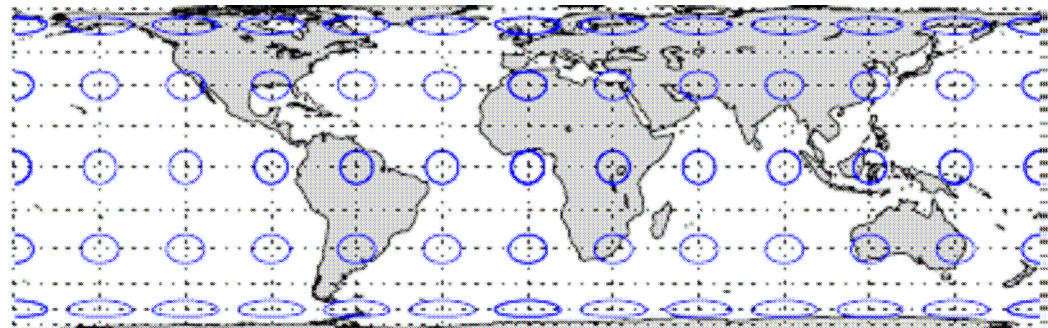
➤ Propriétés des projections:

- Projection conforme: conserve les angles (ex. projection conique de Lambert, projection cylindrique de Mercator)
- Projection équivalente: conserve les surfaces (ex. Lambert azimuthal equal area LAEA)
- Les 2 propriétés ne sont pas réconciliables
- Pas de projection, pas de carte parfaite qui soit partout exacte (les déformations sont inévitables)

Projection cylindrique conforme de Mercator



Projection cylindrique équivalente de Lambert



Éléments de géodésie

- Propriétés des projections (suite) :
 - Il faut choisir la meilleure projection en fonction de l'usage
 - Pour un usage statistique, la LAEA est recommandée au niveau européen car distord moins les surfaces et donne donc une représentation plus exacte des cartes de densité
 - Les carreaux sont découpés sur une projection LAEA
 - Conversion préalable des coordonnées en LAEA (algorithmes fournis par l'IGN)

Éléments de géodésie

- Cas des DOM :
 - La projection LAEA est adaptée pour l'Europe continentale
 - En Martinique et à la Réunion, le carroyage est établi à partir des systèmes locaux de projection cartographique :
 - En Martinique, projection « Fort-Desaix » basée sur l'ellipsoïde International Hayford 1909 et une projection Mercator (UTM fuseau 20 Nord)
 - A La Réunion, projection « RGR92 » basée sur l'ellipsoïde IAG-GRS 1980 et une projection Mercator (UTM fuseau 40 Sud)
 - Pas de données carroyées en Guadeloupe et en Guyane, en raison d'une moins bonne reconstitution des ménages fiscaux dans la source RFL

Le plan cadastral

Le cadastre :

- 2 formats existants :
 - vectorisé (80% des communes) : tous les objets (parcelle, bâti, plaque adresse...) sont disponibles sous format vecteur utilisable dans un SIG
 - raster : seul un scan de la feuille cadastrale est livré accompagné des coordonnées des centres de parcelle



Le plan cadastral

2 versions disponibles :

- Le plan cadastral produit par la DGFIP
 - Traitements conséquents pour lire chaque couche du cadastre (parcelles, bâti, zones communales, étiquettes...)
 - Conversion des coordonnées fournies dans diverses projections (Lambert Carto, Lambert 9 zones, coordonnées locales)
- La BD PARCELLAIRE® de l'IGN :
 - Assemblage du plan cadastral de la DGFIP
 - Mise en cohérence topologique (problèmes de chevauchement, de trous, cohérence avec les couches du RGE)
 - Processus de vectorisation interne à l'IGN
- Les 2 sources sont exploitées sur plusieurs millésimes pour maximiser la géolocalisation à la parcelle

Comment carroyer ?

1^{ère} étape : la géolocalisation

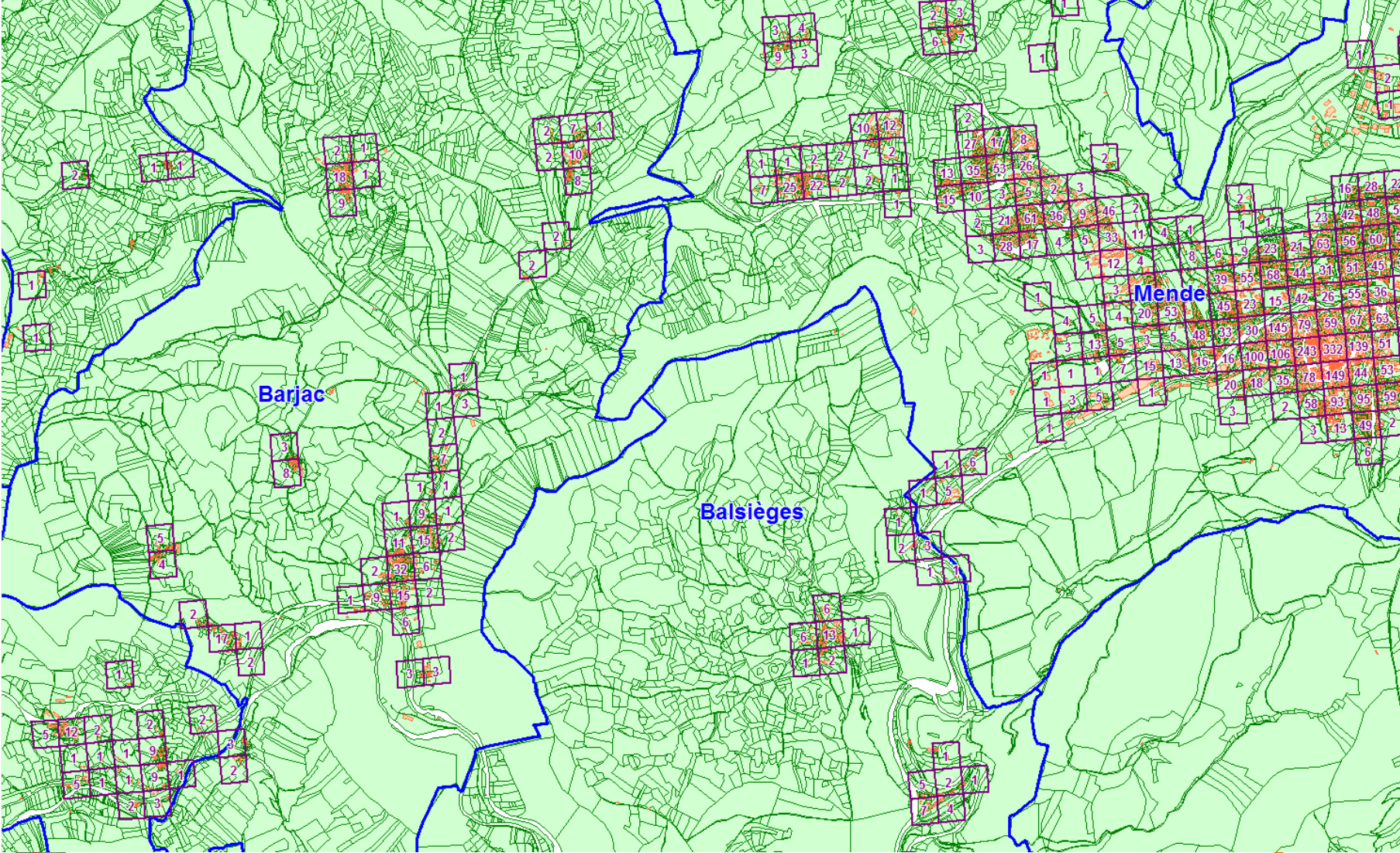
- Géolocalisation à la parcelle :
 - 99,6% des locaux TH en 2010 sont géolocalisés au centre de leur parcelle
- Imputation de coordonnées pour les locaux non géolocalisés à la parcelle (parcelle inexistante ou non renseignée)
- Du plus fin au plus grossier :
 - Interpolation à la voirie selon le même côté
 - Position au centre de la voirie dans la section cadastrale
 - Position au centre de la section
 - Position au centre de la voie
 - En ultime recours, position au centre de la commune

Comment carroyer ?

2^{ème} étape : agrégation au carreau

- L'appartenance d'un local de coordonnées x,y à un carreau se déduit simplement de ses coordonnées : $\text{Partie.Entière}(x/m)*m$ et $\text{Partie.Entière}(y/m)*m$ donnent la position du coin inférieur gauche du carré d'appartenance de m mètres de côté
- On obtient en Métropole 2 278 213 carreaux de 200 m habités :
 - Dont 825 731 carreaux d'un ménage
 - Dont 1 815 800 de moins de 11 ménages
- Nécessité de traiter le secret statistique et fiscal :
 - secret fiscal : pas de diffusion sous le seuil de 11 ménages fiscaux
 - secret statistique : pas de diffusion de statistiques permettant d'en déduire qu'un ensemble d'individus (ou de ménages) possèdent à plus de 80% une caractéristique jugée « sensible »

Comment carroyer : exemple en Lozère



Du carreau au rectangle

- Une solution :
 - Agréger les carreaux en rectangles d'au moins 11 ménages (seuil de diffusion des statistiques fiscales)
- Autres solutions ailleurs :
 - Diffuser sur des carreaux plus grands
 - Remonter le niveau de population des petits carreaux à un seuil donné
 - Non-diffusion des carreaux trop peu peuplés
 - Permutations de populations entre carreaux

Du carreau au rectangle

- Processus descendant de constitution des rectangles :
 - La Métropole est divisée en 36 grands carrés de taille identique (pour des raisons d'optimisation de temps de traitement)
 - Chaque grand carré est découpé en 2 soit horizontalement soit verticalement par son centre de gravité (pondéré par la population) pour former 2 rectangles
 - Les rectangles ainsi créés sont à leur tour découpés horizontalement ou verticalement selon la même méthode
 - Et ainsi de suite... jusqu'à ce qu'un découpage supplémentaire entrainerait le non respect de la règle du seuil de diffusion

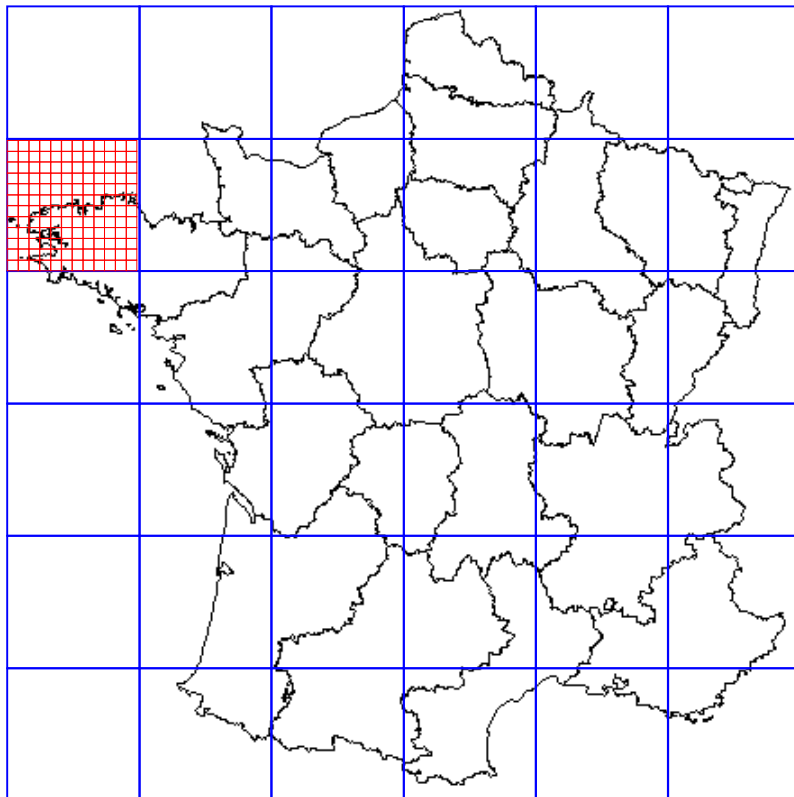
Du carreau au rectangle

Constitution des rectangles (suite) :

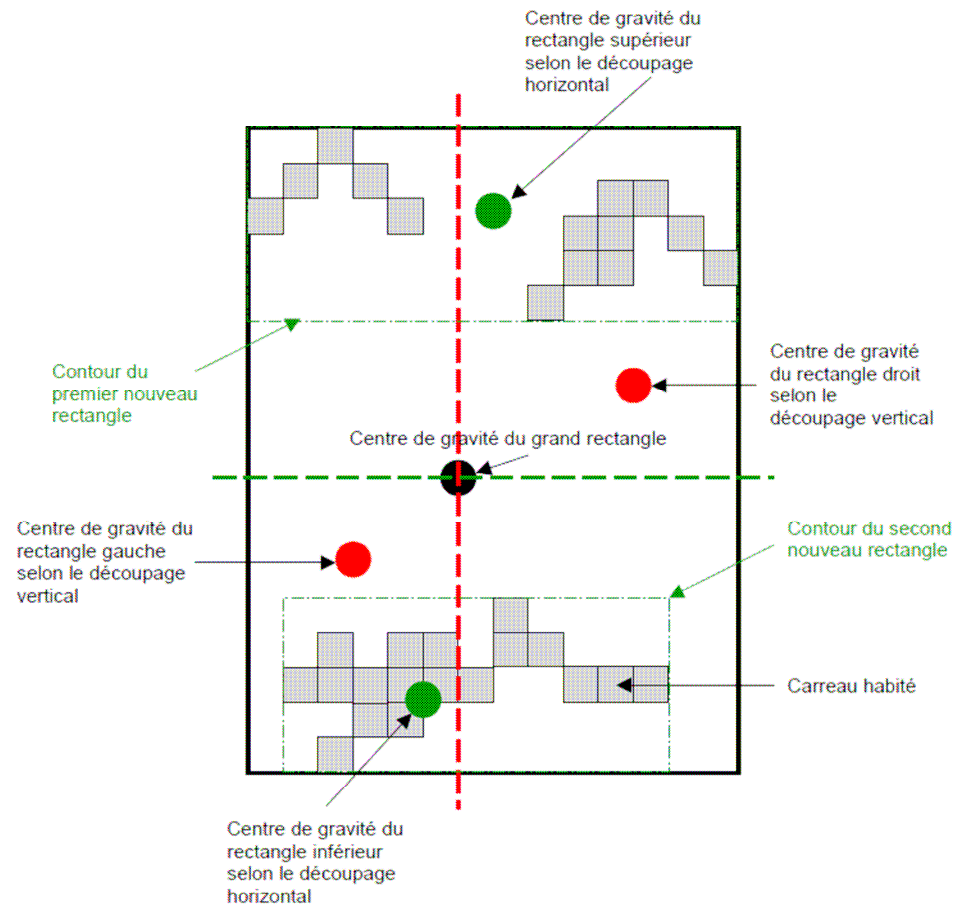
- À chaque étape on arbitre entre une scission horizontale ou verticale
- En choisissant celle qui produit 2 rectangles au sein desquels les carreaux habités sont les moins dispersés
- La dispersion d'un rectangle est mesurée par la somme des distances au carré entre son centre de gravité et ses carreaux habités, pondérées par la population

Du carreau au rectangle

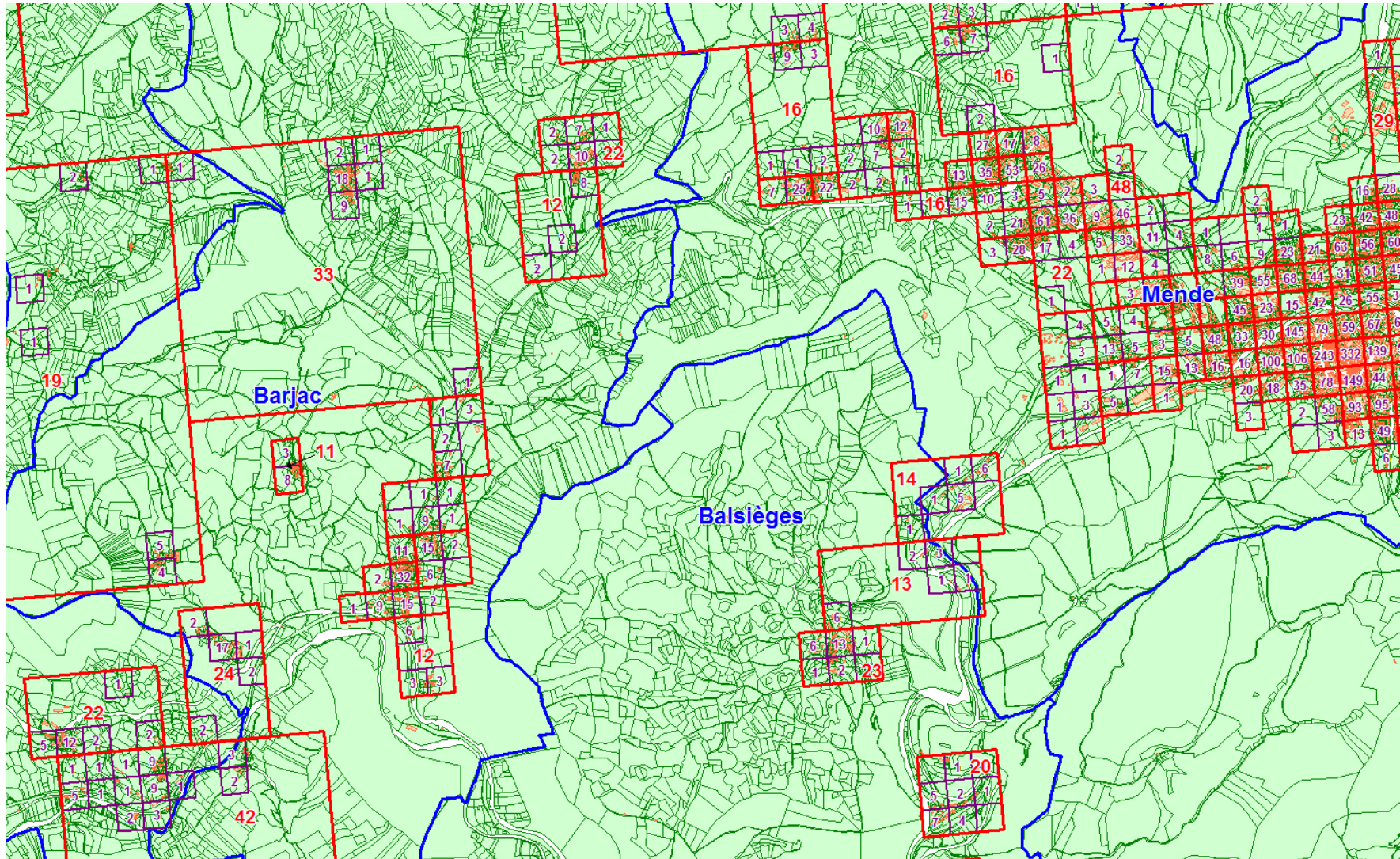
Au départ, 36 grands carrés
composés de petits carreaux



Traitement par grand carré :
à chaque étape, des rectangles plus petits



Du carreau au rectangle



Du carreau au rectangle

Résultats :

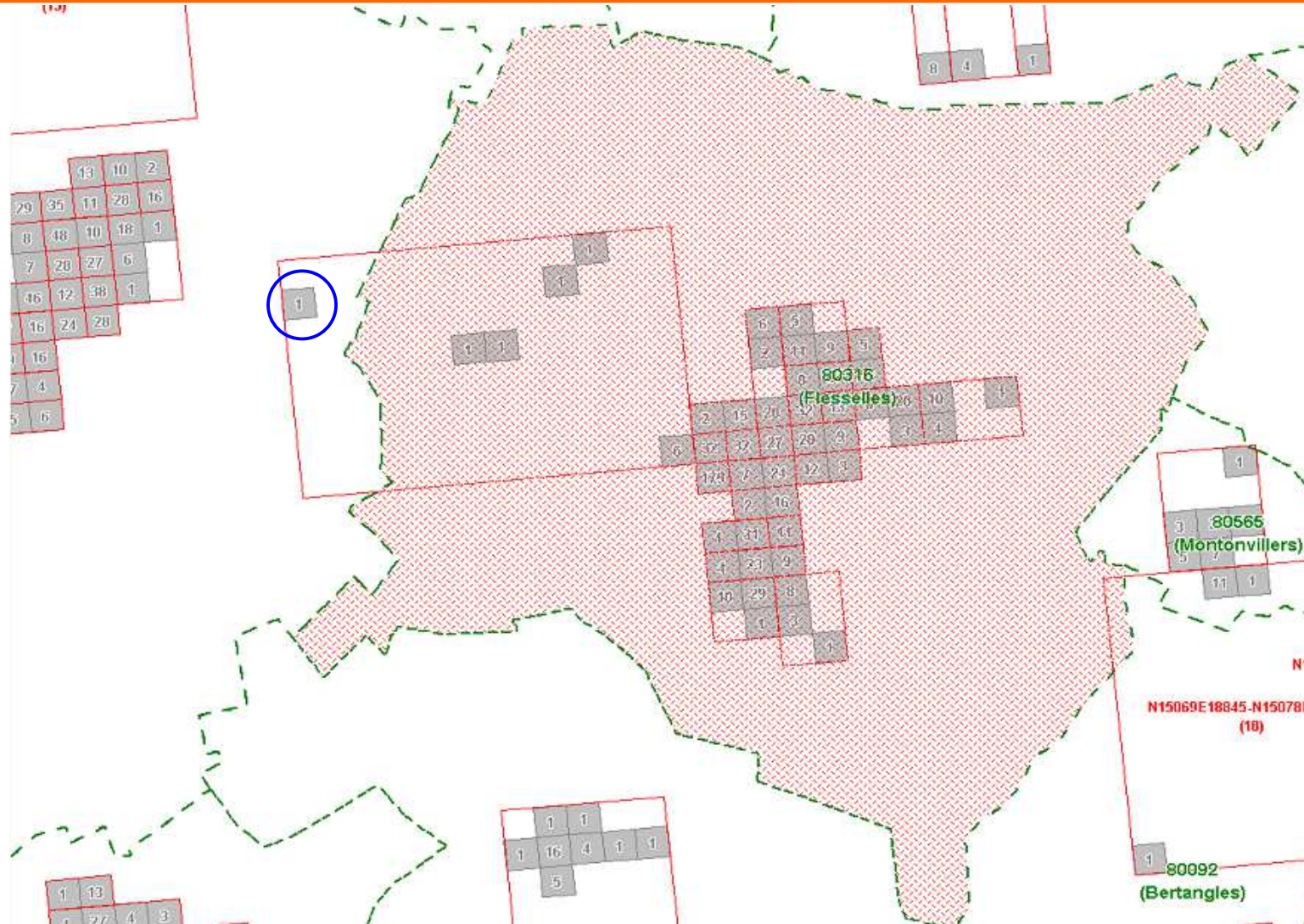
- On passe de 2 278 213 carreaux à 698 659 rectangles
- La moitié des rectangles compte de 11 à 21 ménages
- 41% des carreaux initiaux de 11 ménages ou plus sont regroupés dans des rectangles pour satisfaire la contrainte

	effectifs	% carreaux
Total carreaux habités en métropole	2 278 213	100%
Carreaux de 11 ménages ou plus	462 413	20,3%
<i>dont carreaux de 11 ménages ou plus diffusés tels quels</i>	<i>273 459</i>	<i>12,0%</i>
<i>dont carreaux de 11 ménages ou plus regroupés avec d'autres carreaux dans des rectangles</i>	<i>188 954</i>	<i>8,3%</i>
Carreaux de moins de 11 ménages regroupés avec d'autres carreaux dans des rectangles	1 815 800	79,7%
Total rectangles en métropole (y compris 3.)	698 659	

Risque de rupture de confidentialité

- Problème de la rupture de confidentialité par différenciation avec d'autres niveaux géographiques de diffusion des statistiques fiscales
 - à la commune
 - à l'Iris
- 153 cas de rupture identifiés en Métropole avec le niveau communal
- Mise à blanc du revenu du plus petit rectangle de ces communes

Risque de rupture de confidentialité



Les variables « non sensibles »

- Nombre d'individus (seule variable diffusée au carreau)
- Nombre d'individus de 0-3 ans, 4-5 ans, 6-10 ans, 11-14 ans, 15-17 ans, 25 ans et plus
- Nombre de ménages
- Nombre de ménages de 5 personnes et plus
- Nombre de ménages présents depuis 5 ans ou plus dans leur logement actuel
- Nombre de ménages en logement individuel, en logement collectif
- Surface cumulée des résidences principales

La variable « indicateur de revenus »

Principes

Définir un indicateur sommable

Ne pas diffuser de trop bas revenus ou de trop hauts revenus

Méthode : le revenu annuel par u.c. winsorisé

- Un revenu par u.c. d'un ménage supérieur au 8^{ème} décile de la distribution (29 336 € en métropole) est descendu à ce seuil
- Un revenu par u.c. d'un ménage inférieur à 40% de la médiane de la distribution (7 500 € en métropole) est remonté à ce seuil
- Une winsorisation « flottante » est utilisée si les revenus de tous les ménages d'un rectangle sont winsorisés

L'indicateur diffusé

La somme des revenus par u.c. (winsorisés) des individus : le « revenu par u.c. » d'un individu est le revenu par u.c. du ménage auquel il appartient

Les variables sensibles

Principe : on ne souhaite pas indiquer que, dans un rectangle, une très grande majorité (plus de 80 %) de ménages, ou d'individus, possède une caractéristique jugée sensible.

Variables concernées :

- Nombre d'individus de 65 ans et plus
- Nombre d'individus de 75 ans et plus
- Nombre de ménages d'une personne
- Nombre de ménages propriétaires
- Nombre de ménages dont le revenu par u.c. est inférieur au seuil de bas revenu (60% de la médiane de la distribution)

Pour respecter ce principe, dans certains cas l'effectif réel n'a donc pu être diffusé. Ceci est signalé à l'utilisateur de la façon suivante : à chacune de ces variables sensibles est associée dans le fichier diffusé une variable indicatrice, valant 1 si le nombre « n » diffusé doit être compris comme « n ou plus », et 0 sinon.

Les variables sensibles

Résultats :

- Au final très peu de rectangles sont impactés par le traitement des variables sensibles (sauf pour le statut d'occupation)

Variable	Nombre de rectangles traités	% de rectangles traités
Individus de 65 ans et plus	317	0,05%
Individus de 75 ans et plus	59	0,01%
Ménages d'une personne	1 587	0,23%
Ménages propriétaires (+80%)	417 059	59,69%
Ménages propriétaires (-20%)	32 058	4,59%
Ménages au-dessous du seuil de bas revenu	688	0,10%

Mise à disposition

- Mise à disposition sur insee.fr – rubrique « données locales »
- Couches cartographiques des carreaux et des rectangles en projection Lambert II étendue
- Au niveau des carreaux, diffusion uniquement du nombre d'individus
- Au niveau des rectangles, diffusion de toutes les variables, avec les « indicatrices de traitement » des variables sensibles
- Utilisation des carreaux par ventilation des statistiques au rectangle au prorata de la population de ses carreaux

Mise à disposition

Institut national de la statistique et des études économiques
Mesurer pour comprendre

Mobile | Actualités | Agendas | Contactez-nous | Aide | FAQ | Liens | English | Home page

Chercher sur le site

Accueil | **Thèmes** | Bases de données | Publications et services | Régions | Définitions et méthodes | Accès par public | L'Insee et la statistique publique

Thèmes

- ▶ Agriculture
- ▶ Commerce
- ▶ Comptes nationaux - Finances publiques
- ▶ Conditions de vie - Société
- ▶ Conjoncture
- ▶ Économie
- ▶ Enseignement - Éducation
- ▶ Entreprises
- ▶ Industrie - IAA - Construction
- ▶ Population
- ▶ Revenus - Salaires
- ▶ Santé
- ▶ Services - Tourisme - Transports
- ▼ Territoire
 - ▶ Pays, villes d'Europe et du monde
 - ↳ Régions, départements et villes de France
 - ▶ Environnement
 - ▶ Travail - Emploi

Accueil > Thèmes > Territoire > Régions, dé... > Données car... > Données carroyées à 200 mètres

■ Données carroyées à 200 mètres

Mise à jour : 20 novembre 2013

18 variables sur la structure par âge des individus, les caractéristiques des ménages (locataire/propriétaire, etc.) et les revenus au 31 décembre 2010.

Afin de respecter la règle de diffusion des données sur les revenus fiscaux des ménages, aucune information statistique (à l'exception du nombre total d'individus) n'est diffusée sur des carreaux de moins de 11 ménages. Ces carreaux de faibles effectifs sont donc regroupés en rectangles de taille plus importante et satisfaisant à cette règle des 11 ménages minimum.

Par ailleurs, un certain nombre de variables considérées comme « à risque » ont été traitées afin que tout risque de rupture de confidentialité soit évité.

L'utilisation correcte de ces données carroyées impose donc une lecture attentive de la [documentation complète sur les données carroyées à 200 mètres](#).

:: Téléchargement des données et des fonds de carte

Chaque fichier compacté (Format ZIP) contient :

- un fichier de données en format dBase (DBF)
- un fond de carte en format de transport MapInfo (MIF/MID)

Les carreaux décrits dans les fichiers sont uniquement ceux portant une population non nulle.

[Liste et description des variables](#) (Format PDF - 63 Ko)

■ Tables des carreaux

- [Pour la métropole](#) (Format ZIP - 108 Mo)
- [Pour la Martinique](#) (Format ZIP - 303 Ko)
- [Pour La Réunion](#) (Format ZIP - 396 Ko)

■ Tables des rectangles

- [Pour la métropole](#) (Format ZIP - 53,1 Mo)

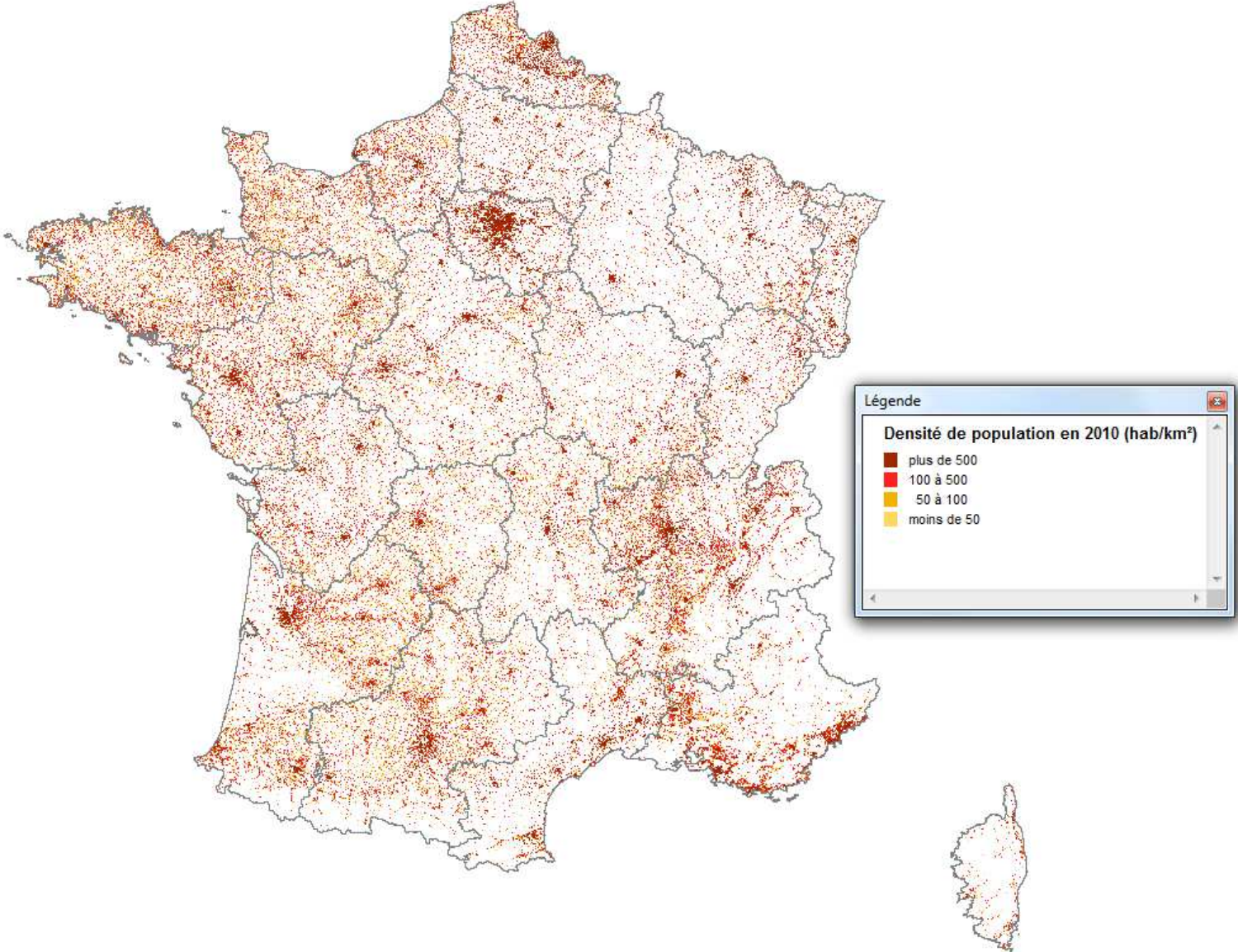
Documentation

- Documentation synthétique sur les données carroyées à 200 mètres
- Documentation complète sur les données carroyées à 200 mètres
- Modes opératoires :
 - 📄 [Sous Mapinfo, mode opératoire pour la métropole \(tables en _m\)](#) (Format PDF - 637 Ko)
 - 📄 [sous QGIS](#) (Format PDF - 494 Ko)

Pour en savoir plus

- [Documentation générale sur le carroyage](#)

Mise à disposition



La production de données carroyées à l'Insee

Merci de votre attention !

Insee

18 bd Adolphe-Pinard
75675 Paris Cedex 14

www.insee.fr  

Informations statistiques :
www.insee.fr / Contacter l'Insee
09 72 72 4000
(coût d'un appel local)
du lundi au vendredi de 9h00 à 17h00

