

Documentation générale sur le carroyage

Source

Les revenus fiscaux localisés (RFL) sont établis à partir des fichiers exhaustifs des déclarations de revenus des personnes physiques et de la taxe d'habitation fournis à l'Insee par la Direction Générale des Impôts. L'Insee procède au rapprochement de ces deux fichiers afin d'estimer le revenu fiscal à des niveaux géographiques finement localisés, tout en préservant la confidentialité des données.

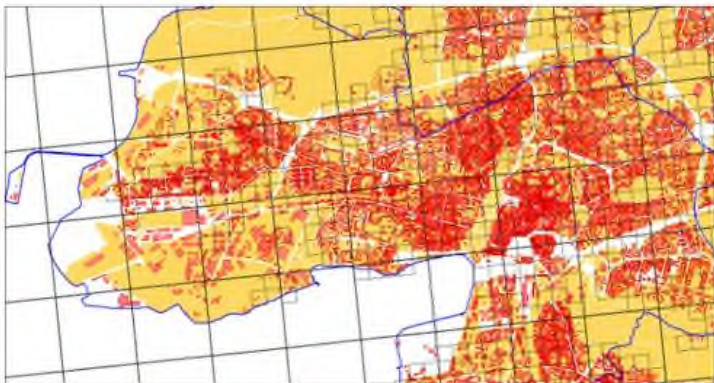
Pour plus de détails, voir [Revenus fiscaux localisés](#) dans la rubrique Définitions et méthodes.

Géographie

L'appartenance à un carreau se fait par une simple opération mathématique sur les coordonnées géographiques individuelles (partie entière). Le système de projection utilisé pour exprimer ces coordonnées dépend du territoire :

sur la métropole c'est le « Lambert Azimutal Equal Area (code EPSG 3035) » qui est conforme aux recommandations européennes dans le cadre de la directive INSPIRE, voir [INSPIRE_Specification_GGS_v3.0.1.pdf](#) ;

sur les régions d'outre mer où ces recommandations ne s'appliquent pas, il s'agit du système « UTM 20 Nord ellipsoïde de Hayford - code EPSG 2973 dans le système géodésique Fort Desaix » pour la Martinique et « UTM 40 S ellipsoïde IAG-GRS 1980 (code EPSG 2975) dans le système géodésique RGR92 » pour la Réunion. Ce système est lu comme étant WGS_1984_UTM_Zone_40S (EPSG 32740) par QGIS. Ces deux systèmes de projections sont très proches et ne diffèrent que de moins d'un centimètre. Les autres régions (Guyane, Guadeloupe et Mayotte) ne peuvent actuellement pas être l'objet des traitements décrits ici.



Commune de La Rochelle (17) : cartographie du bâti (en rouge - non inclus dans les fichiers mis à disposition) et les deux niveaux de carroyage LAEA, à 1km (sur tout le territoire) et à 200m (restreint aux carreaux habités). Les contours des carreaux sont extraits des fonds de carte livrés. Leur orientation n'est pas Nord-Sud.

Données

Les données fournies ici proviennent d'une exploitation spécifique des fichiers fiscaux visant à attribuer à chaque ménage une position géographique précise ([voir plus loin](#)). Les données carroyées de population sont cohérentes avec les données diffusées par la source « revenus fiscaux localisés », mais elles ne peuvent fournir **qu'une vision partielle de la population**, rappelons que par rapport au recensement :

les populations non référencées à la taxe d'habitation ne sont pas présentes : personnes sans domicile, collectivités ;

les étudiants sont généralement localisés au domicile de leurs parents ;

La notion de résidence principale peut être sensiblement différente.

Les données se présentent sous deux formes alternatives : un fichier en format dBase DBF et un fond de carte en format de transport MapInfo MIF/MID.

Les estimations carroyées se présentent sous forme de deux jeux de données, l'un avec une résolution de 1km, l'autre avec une résolution de 200m.

Le fond de carte MIF/MIF

Fournit les contours géométriques des carreaux. A chaque carreau sont associées les données attributaires suivantes :

Données carroyées - Comment c'est fait ?

Méthode de géolocalisation

Par rapport aux données carroyées à 1 km dont la maille géographique est suffisamment grande pour qu'un processus d'estimation donne des résultats raisonnables, une maille de 200 m de côté se révèle être bien trop sensible.

Les données publiées ici ne sont donc pas des estimations mais le résultat d'un simple comptage carreau par carreau après affectation « exacte » d'une coordonnée géographique à chaque ménage. La notion de coordonnée « exacte » correspond en fait à celle d'un point intérieur - généralement proche du centre géométrique - de la parcelle cadastrale principale où réside le ménage. Le repérage de cette parcelle provient de la taxe d'habitation, mais dans une proportion relativement faible de cas ce repérage précis n'est pas possible et une position approximative est déterminée en fonction des autres éléments géographiques disponibles.

Si l'adresse appartient à une voie numérotée sur laquelle d'autres ménages ont été localisés à la parcelle et si elle porte une numérotation alors le positionnement est déterminé par interpolation entre les adresses connues.

Sinon, si la section cadastrale et la voie sont connues, la position est celle du barycentre (centre de gravité) des ménages localisés à la parcelle sur ces deux mêmes éléments.

Sinon, si seule la section cadastrale est connue, la position est celle du centre géométrique de la section.

Sinon, si seule la rue est connue, la position est celle du barycentre des ménages localisés à la parcelle sur cette rue.

Sur l'ensemble métropole, Réunion et Martinique, le nombre de ménages 2009 non localisés à la parcelle est d'environ 120 000, soit 0,4 % de l'ensemble des ménages.

Le cumul des effectifs relatifs à l'ensemble de ces quatre cas est fourni, carreau par carreau dans l'attribut **indXYNE1**. Il est à noter qu'en cas d'échec des méthodes citées, aucune affectation de coordonnées n'est réalisée. Les effectifs non localisés sont à la fois suffisamment marginaux et peu concentrés spatialement pour que leur traitement ne soit pas indispensable pour l'exploitation ultérieure de données qui ont par construction un décalage considérablement plus important avec la réalité (localisation fiscale et ancienneté). Néanmoins ceci signifie que la correspondance comptable parfaite avec les chiffres de « revenus fiscaux localisés » n'est pas assurée.

Dans la majorité des cas la source des coordonnées géographiques est une extraction annuelle des fichiers cadastraux réalisée par la DGFIP et la localisation est donc cohérente avec la diffusion du cadastre faite sous www.cadastre.gouv.fr. Pour environ un tiers des communes, cette extraction a été complétée grâce à la BD PARCELLAIRE de l'IGN. Relativement au millésime des données de population, la plus grande fraîcheur de la première source et la moins bonne actualité de la seconde sont les deux raisons principales des défauts de localisation.



La Rochelle (autour du quartier de Port Neuf). Bâti, contour des parcelles (en vert sombre), localisation « exacte » des ménages (étoiles noires, une étoile pouvant représenter de nombreux ménages) et carroyage à 200m. L'adéquation de la localisation des ménages avec le bâti est bien moins bonne sur les grands ensembles mais cette imprécision est généralement « gommée » par la comptabilisation par carreau.

Le cadastre

La géolocalisation sur une grande échelle a été rendue possible par un accès aux fichiers cadastraux, tels qu'ils sont présentés sur le site internet du cadastre. En vertu d'un avenant à une convention d'échange de données entre l'Insee et la DGFIP, l'Insee dispose désormais au printemps de chaque année d'une copie du cadastre réalisée à ce moment, donc incluant les toutes dernières mises à jour reportées sur le plan cadastral.

Les fichiers cadastraux de la DGFIP sont constitués de trois composantes :

Pour les communes dont les planches cadastrales ont été mises en format vecteur, des fichiers en format DXF contenant les définitions vectorielles des multiples couches du cadastre : parcellaire mais aussi voirie, lieux-dits, points remarquables...

Pour les communes non vectorisées, des images de planches cadastrales en format TIF, qui ne sont pas exploitées, et des positions de parcelles exprimées en coordonnées relatives à chaque planche.

Parfois (et de plus en plus souvent), ces positions sont accompagnées d'un descriptif de la transformation mathématique (rotation, homothétie et translation) nécessaire pour passer des coordonnées locales à celles exprimées dans un système de projection standard.

Au final, les communes se répartissent en trois catégories :

Des communes vectorisées (22 000 en 2010)

Des communes non vectorisées mais avec transformation des coordonnées locales possibles (2 000 en 2010)

Des communes non vectorisées sans possibilité d'interprétation des coordonnées locales (12 000 en 2010).

Les fichiers cadastraux sont livrés dans des systèmes de projection très variés, Lambert 4 zones, Lambert 93 ou Lambert 93 9 zones, quelquefois présents simultanément sur un même département. Dans le processus de lecture à l'Insee toutes les coordonnées sont converties en système Lambert 93 à l'aide de programmes écrits spécialement en SAS à partir des spécifications disponibles sur le site de l'IGN.

À partir de 2011, une nouvelle possibilité a été offerte par l'ouverture de l'accès au Référentiel à Grande Echelle (RGE) de l'IGN. En matière de cadastre, le RGE contient la BD PARCELLAIRE qui est un sous ensemble des données fournies par la DGFIP, mais avec deux atouts importants :

La géométrie a été vérifiée.>

Le travail de transformation des coordonnées locales en coordonnées géographiques a été réalisé de façon systématique.

Compte tenu des traitements réalisés par l'IGN, l'actualité de la BD PARCELLAIRE est très nettement décalée par rapport à celle de la livraison DGFIP. Ceci est un inconvénient mais aussi un avantage car la combinaison des deux produits donne la possibilité de pouvoir encadrer les données à géolocaliser entre deux visions à des moments différents de la réalité du terrain.

Par ailleurs, sur les communes non vectorisées, la plupart des parcelles sont simultanément présentes dans les deux jeux de données : en coordonnées locales dans les fichiers de la DGFIP et en coordonnées réelles dans les fichiers de l'IGN. Ceci donne l'opportunité de retrouver les descriptifs de transformation géométrique à appliquer sur les fichiers de la DGFIP et de bénéficier de leur plus grande actualité.

Les ingrédients de la localisation infra-communale

Côté cadastre, seuls certains composants sont utilisés :

La couche parcelletex du cadastre vectorisé de la DGFIP qui donne la position où est tracé l'identifiant de la parcelle. Cette position est généralement à l'intérieur de la parcelle concernée, sauf pour des parcelles de très petite superficie ou filiformes. Ce n'est pas non plus la position du centre de gravité de la parcelle - qui aurait aisément pu être calculée - mais cette position aurait aussi pu être en dehors de la parcelle dans le cas de parcelles à trous ou de forme concave.

En 2011, deux extractions ont été faites par la DGFIP, successivement en mai-juin 2010 et en avril 2011. La première extraction est également utilisée par le processus de géolocalisation, sous forme d'une couche complémentaire parcelletex_1. Cette dernière couche donne une autre actualité du cadastre permettant de gérer certains cas de restructuration cadastrale.

Les coordonnées locales livrées par la DGFIP et convertibles en coordonnées géographiques font l'objet d'une couche complémentaire parcelletex2.

La couche LOCALISANT (localisant_rge) de la BD PARCELLAIRE.

Ces quatre couches donnent lieu à la création de trois tables de passage sous le même format entre identifiant de parcelles et coordonnées géographiques. Les identifiants de parcelle sont normalisés sous la forme : cccbbbsspppp, où :

ccc sont les trois derniers chiffres du code commune (COG) ;

bbb sont soit '000', soit les trois derniers chiffres du code commune d'une commune qui a été fusionnée à la commune ccc ;

ss deux caractères identifiant la section, commençant par un '0' si le code section est constitué d'un seul caractère ;

pppp quatre caractères - pas toujours numériques - identifiant la parcelle à l'intérieur de la section.

Une même parcelle peut être décrite dans plusieurs tables de passage. Compte tenu des différences de qualité géométrique et surtout d'actualité, le recours à une table n'a lieu qu'en cas d'échec dans une table plus prioritaire. L'ordre de priorité est le suivant :

parcelle_tex > parcelletex_1 > localisant_rge > parcelletex2

La couche décrivant les sections cadastrales du cadastre vectorisé de la DGFIP est également utilisée par le processus pour fournir dans certains cas des positions géographiques de dernier recours sous forme de positionnement au centre géométrique de la section (section_m).

La localisation des données fiscales repose sur la localisation préalable des fichiers de la taxe d'habitation. En plus du code commune ceux ci contiennent une référence cadastrale (refcad) exprimée sur 10 caractères c'est à dire avec des conventions légèrement différentes de l'identification des parcelles et sections dans les fichiers de la DGFIP : pas de '000' pour les communes fusionnées ni pour les sections d'identifiant mono caractère et codage particulier sur Paris Lyon, Marseille et Toulouse.

Dans un nombre limité d'enregistrements des fichiers de la taxe d'habitation, mais concentrés sur certains départements et certaines grandes communes, les références cadastrales ne sont pas (encore) présentes. Deux informations complémentaires sont alors mobilisées : l'identifiant RIVOLI de la voie et le numéro de l'adresse dans la voie (sans le BIS/TER/QUATTRO). Ces informations serviront à calculer une position approximative.

Le processus de localisation de la taxe d'habitation

Après recodage de l'identifiant de parcelle de la TH au format des fichiers cadastraux, le processus se déroule en plusieurs phases successives, chaque phase tentant de localiser ce qui ne l'a pas été à la phase précédente. Lorsque des coordonnées ont pu être affectées, un code fournissant l'origine de ces coordonnées est renseigné (sourceXY) pour l'enregistrement TH correspondant.

Tous les traitements se font commune par commune. Compte tenu des évolutions de la géographie communale et des millésimes différents présents dans l'ensemble des fichiers impliqués, ceci provoque de nombreux échecs de localisation. Ces cas ne sont pas traités dans la localisation des fichiers de la taxe d'habitation mais résolus ultérieurement lors de la localisation des fichiers de ménages fiscaux.

Étape 1 : Appariement sur identifiant de parcelle (sourceXY=1)

Étape 2 : Interpolations

Les enregistrements pourvus d'un numéro dans leur voie et dont la voie est présente dans les enregistrements localisés à l'étape 1 avec une numérotation numérique, font l'objet d'un positionnement par interpolation entre deux adresses positionnées à la parcelle. L'interpolation est elle même en plusieurs phases successives

Interpolation entre deux adresses connues de même parité paire/impair (SourceXY=1.1)

Positionnement à la première adresse supérieure de même parité (SourceXY=1.3)

Positionnement à la dernière adresse inférieure de même parité (SourceXY=1.5)

Interpolation entre deux adresses connus de parité opposée (SourceXY=1.2)

Positionnement à la première adresse supérieure parité opposée (SourceXY=1.4)

Positionnement à la dernière adresse inférieure de parité opposée (SourceXY=1.6)

Étape 3 : Positionnement sur l'intersection voie - section cadastrale (SourceXY=2). Les enregistrements pourvus d'un couple voie, section cadastrale présent dans les enregistrements localisés à l'étape 1 sont positionnés au barycentre de ces derniers.

Étape 4 : Positionnement à la section cadastrale (SourceXY=3). Lorsque seule la section cadastrale est connue, la position affectée est celle du centre géométrique de la section (cadastre vectorisé) ou du barycentre des positions des parcelles (cadastre non vectorisé).

Étape 5 : Positionnement à la voie (SourceXY=4). Lorsque seule la voie est connue, la position affectée est celle du barycentre des enregistrements localisés à l'étape 1.

La localisation des ménages fiscaux

Les données des ménages fiscaux ne sont pas localisées directement, mais le sont par localisation du local soumis à la taxe d'habitation auquel chaque ménage est rattaché dans le cadre de la production « revenus fiscaux localisés ». Par contre, contrairement à ce qui est fait dans cette dernière production, plutôt que de chercher à appairer données de revenu et taxe d'habitation sur des fichiers ayant sensiblement la même actualité, c'est plusieurs fichiers successifs résultant de la localisation de la taxe d'habitation qui sont utilisés. Un même local s'y trouve donc potentiellement plusieurs fois avec une qualité de localisation qui peut différer d'un millésime à un autre. Le code sourceXY décrit précédemment et dont les valeurs n'ont pas été choisies au hasard, donne un moyen simple d'arbitrer en faveur de la localisation la plus précise : la localisation qui est conservée provient du fichier de la taxe d'habitation où sourceXY est le plus faible et en cas d'égalité du fichier de millésime le plus récent. Cette façon de procéder permet de « récupérer » un certain nombre de cas de restructurations cadastrales, ainsi que les changements dans la liste des communes elles-mêmes.

Performance du processus d'appariement :

Non localisés : 461

Interpolés (ensemble des cas) : 75668

Localisés section - voie : 3179

Localisés section : 7943

Localisés voie : 24865

Total des ménages : 26615699

Localisation de RFL 2009 (métropole, Réunion et Martinique).

coordonnées géographiques du coin sud-ouest du carreau : pour la métropole le système de projection LAEA - code EPSG 3035 (attributs **x_LAEA et y_LAEA**), pour la Réunion le système de projection UTM 40 S ellipsoïde IAG-GRS 1980 de code EPSG 2975 (lu comme du WGS_1984_UTM_Zone_40S de code EPSG 32740 par QGIS) (attributs **x_PDN et y_PDN**) et pour la Martinique le système de projection UTM 20 Nord ellipsoïde de Hayford - code EPSG 2973 (attributs **x_UTM20N et y_UTM20N**).

population totale du carreau (**ind**) ;

le code carreau de la forme suivante : CRS3035RES200mN1453400E1452800 (Référence : D2.8.III.1 INSPIRE Data Specification on Statistical Units - Draft Guidelines) (**idINSPIRE**)

population du carreau qui a été comptabilisée dans l'attribut précédent mais provient d'une localisation peu précise (**indXYNE1**), c'est à dire dont la position géographique n'est pas le centre de la parcelle habitée par cette population. La position réelle de la population correspondante peut se trouver dans un carreau voisin de celui où elle comptée.

Le système de projection du fond de carte lui-même est le Lambert II étendu pour la métropole (de manière à être accepté par des SIG anciens), UTM 40 S ellipsoïde IAG-GRS 1980 de code EPSG 2975 (lu comme du WGS_1984_UTM_Zone_40S de code EPSG 32740 par QGIS) pour la Réunion et UTM 20 Nord ellipsoïde de Hayford - code EPSG 2973 (attributs x_UTM20N et y_UTM20N) pour la Martinique.

Le fichier DBF

Reprend les mêmes données que les données attributaires précédentes mais sans les contours géométriques des carreaux. Par rapport aux fichiers MIF/MID qui ne peuvent décrire que des entités géographiques existantes, le fichier DBF contient une première ligne supplémentaire sans coordonnées de carreau. Cette ligne fournit le nombre de personnes qui n'ont pas été localisées, même approximativement.

Données à 1 Km

Fin 2010, un premier jeu d'estimations carroyées de population a été mis en ligne sur insee.fr. Ce premier jeu d'estimation était basé sur une méthode de comptabilisation similaire à celle utilisée ici mais limitée à environ deux tiers des communes. Le dernier tiers avait fait l'objet d'un processus d'estimation reposant sur des sources externes sur l'occupation du sol.

Les progrès de la vectorisation du cadastre et la mise à disposition plus large du Référentiel à Grande Echelle développé par l'IGN ont complètement bouleversé la situation. Ces progrès ont rendu les estimations inutiles partout où elles étaient indispensables. Ils ont également contribué à une amélioration très sensible de la qualité de la localisation là où elle résultait d'un rapprochement avec le cadastre.

Avertissement : aucune évolution mesurée par rapport au fichier 2007 ne doit donc être considérée comme significative.

Données à 200 m

Les données régionales détaillées portent sur des carreaux de 200 m de côté, dans les mêmes systèmes de projection que les données à 1km. Les carreaux de 200m de côté correspondent à une subdivision interne en 25 éléments des carreaux de 1km de côté.

Dans les carreaux qui se trouvent à cheval sur la frontière régionale n'est comptabilisée que la population de la région. Si de la population d'une autre région est présente dans le même carreau, alors le carreau est présent dans les fichiers des deux régions.

Communes manquantes

Les communes qui suivent n'ont pu faire l'objet d'une localisation directe et sont donc comptabilisées dans les données non localisées :

Île-de-Sein (29083) : commune non cadastrée (population totale 2008 : 223 habitants) ;

Île-Molène (29084) : commune non cadastrée (population totale 2008 : 218 habitants).

L'ensemble des carreaux couvrant les communes correspondantes sont présents dans les fichiers fournis avec une population fictive (ind=-1). En cas de débordement d'un des carreaux d'une commune sans géolocalisation sur un carreau peuplé d'une commune voisine, ce carreau dont on ne connaîtrait qu'une partie de la population est « masqué » tout comme les autres carreaux de la commune défaillante.

Données carroyées - Pourquoi des carroyages

Un carroyage est un découpage de l'espace géographique en mailles régulières de forme carrée et de taille fixe.

L'espace géographique a deux dimensions

Construire des statistiques c'est élaborer des mesures relatives à des territoires donnés. Dans les premiers temps la préoccupation majeure était d'évaluer la contribution possible de chaque province à l'effort de guerre du pays. Les statistiques étaient alors réduites à une liste de noms de provinces accompagnées d'un effectif de soldats potentiels. Avec le temps les indicateurs statistiques se sont multipliés tout comme les maillages administratifs sur lesquels ces mesures étaient réalisées. Néanmoins la multiplicité des zonages administratifs a des limites et il existe obligatoirement un découpage administratif en dessous duquel aucune subdivision ne peut s'imposer au nom de la gouvernance. En France c'est le niveau communal.

Bien évidemment des questions de connaissance du territoire continuent à se poser à une échelle plus détaillée que ce niveau minimal. À ce point se trouvait une bifurcation :

on pouvait continuer à découper le territoire en nouvelles mailles géographiques plus ou moins justifiées par la réalité du terrain, en oubliant que plus aucune légitimité ne pourrait être invoquée. C'est le chemin qu'a suivi l'Insee à la fin des années 1990 en proposant son découpage en IRIS.

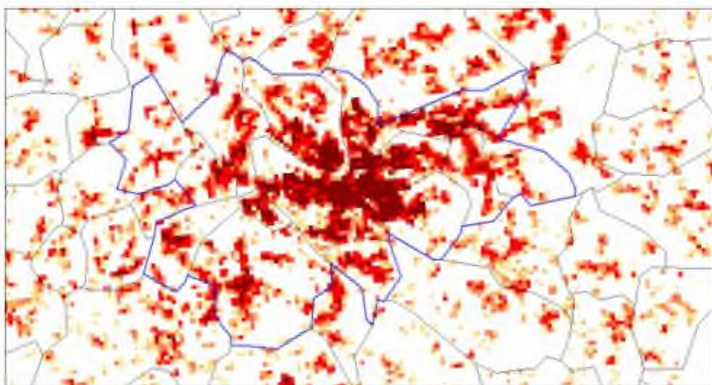
On pouvait aussi réaliser qu'en dessous du niveau de gouvernance minimal, les questions ne se posaient pas exactement dans les mêmes termes. La fonction des nouvelles mailles n'est plus tant de pouvoir restituer de l'information maille par maille, comme au niveau géographique supérieur, que de permettre de décrire le fonctionnement global de l'ensemble des mailles qui constitue ce niveau géographique supérieur.

Et c'est là que le carroyage entre en scène : l'ensemble des mailles doit rendre compte d'une réalité dans un espace à deux dimensions. Alors que les découpages classiques ne sont fondamentalement que des moyens de lister (et une liste n'a qu'une seule dimension) des entités géographiques, le besoin est ici de montrer les interactions entre parties du territoire dont c'est la position dans l'espace géographique qui est la référence la plus naturelle et non l'identifiant dans une quelconque nomenclature.

Ceci induit à penser notre maillage plus en termes de coordonnées géographiques que de contours de zones. Après, pour ce qui est de la forme des mailles, deux arguments militent pour un pavage régulier et carré. Il faut d'abord que le pavage soit suffisamment régulier pour ne pas contenir de zone de taille trop importante pour demander plus tard à être redécoupée : un pavage régulier minimise ce risque. Ensuite on peut imaginer d'utiliser des hexagones, des triangles ou des formes plus exotiques comme en a proposé Escher : le carré aligné sur les axes du système de coordonnées offre en fait l'alternative la plus simple à décrire.

Des carrés pour l'analyse

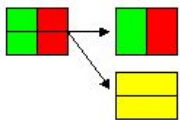
L'opposition - ou la complémentarité - entre zonages classiques et carroyages repose sur bien plus qu'une façon de référencer les zones dans l'espace. On pourra d'ailleurs objecter que les SIG permettent de donner une dimension spatiale aux zonages classiques, même si effectivement les calculs ultérieurs basés sur la proximité resteront plus délicats. En fait la véritable opposition se situe entre statistique zone par zone et ensemble de statistiques, entre vue détaillée et vue d'ensemble et finalement entre passé et futur. Le besoin de suivi couvert par les zonages classiques n'est pas à remettre en cause, ce qu'apportent les carroyages c'est un outil d'analyse, voire d'évaluation. Construit sans a priori sur ce que doivent être les zonages d'intervention ou de gestion, les carroyages permettent une vision impartiale des phénomènes et sont naturellement à même de surligner les zones à enjeu quitte à ce que, plus tard, celles-ci fassent l'objet d'une délimitation classique.



Densité de population par carreau de 200m de côté aux environs d'Angoulême (16). Le découpage communal (en noir) semble ignorer l'emprise effective de la ville alors que la délimitation en bleu (communauté d'agglomération) paraît être un bien meilleur outil d'analyse.

Pour le géographe, la question de la partialité des zonages classiques a un nom : le « Modifiable Area Unit Problem » (MAUP). Le sigle MAUP rappelle en fait que lorsqu'on utilise une statistique sur un zonage on utilise le résultat d'une agrégation de données élémentaires faite sur un mode bien précis, mode qui répondait aux besoins des inventeurs du zonage mais plus forcément aux besoins actuels de l'utilisateur de la statistique. Un autre découpage, plus fin ou fait différemment aurait peut être conduit à des chiffres dont la lecture fournirait un diagnostic totalement différent.

Deux agrégations différentes à partir de quatre unités géographiques de base (par exemple des îlots de 1 000 habitants pour constituer des IRIS de 2 000 habitants). Suivant que l'agrégation se fait dans le sens Nord - Sud ou Est Ouest, la différenciation interne est conservée ou complètement gommée.



Les carroyages sont bien entendu aussi une façon d'agréger les données et une façon brutale de le faire. Dans l'agrégation il y a donc toujours une perte de détail et on pourra argumenter à juste titre que les carroyages sont également soumis au MAUP. Par contre la brutalité même du découpage garantit l'absence de biais systématique puisqu'aucun être humain n'a pu essayer d'incorporer ses préoccupations du moment. Autrement dit, s'il y a une perte d'information, c'est la même partout.

Des carrés pour le partage

L'un des grands défauts des découpages classiques réside dans leur peu de stabilité. En France, même le niveau régional a subi des aménagements ces dernières années. Et plus le niveau de découpage est fin plus les tentations de le remettre en cause sont fortes, par le simple fait que les évolutions du terrain y sont d'autant plus visibles. Lorsqu'on sort du domaine administratif les choses sont évidemment bien pires puisque l'absence de légitimité autorise l'auteur d'un zonage à faire absolument tout ce qu'il peut désirer. À cette instabilité congénitale les carroyages répondent par une immuabilité absolue - du moins à l'échelle de temps d'une civilisation et tant que les choix techniques (système de projection, positionnement et taille) ne sont pas remis en cause.

Un autre grand défaut des découpages classiques est qu'un même niveau géographique peut recouvrir des réalités différentes. En France même le niveau communal ne signifie pas forcément partout la même chose, certains espaces ayant eu des appétits plus forts que d'autres pour la fusion de communes. Au niveau infra-communal le découpage en IRIS ne présente pas non plus l'homogénéité que laisserait supposer un nom de zonage unique car la constitution du zonage a été, à l'époque, complètement décentralisée et largement laissée aux interprétations locales de consignes assez générales. Et à l'échelle de l'Union Européennes les choses sont évidemment bien pires.

Les carroyages en apportant une définition sans possibilité d'intervention ou d'interprétation humaine ultérieure apportent, eux, un cadre unique qui est valable de la même façon sur tout l'espace où le carroyage est défini. Du moins si l'on s'entend sur ce qu'est un carré.

Des carrés pas vraiment carrés

Un carroyage nécessite trois choix : un système de coordonnées qui permettra de repérer les mailles dans l'espace géographique, une façon de positionner le premier carreau et une taille de carreau.

La question du système de coordonnées est la plus délicate car avec un globe terrestre qui n'est qu'approximativement sphérique de nombreuses alternatives ont été imaginées au cours du temps. Le couple longitude/latitude exprimé avec des angles est la première option : elle est par exemple utilisée dans la spécification harmonisée à l'échelle de l'Union Européenne des recensements agricoles. Néanmoins elle se prête mal au calcul de distances, rendant inévitable l'utilisation d'un système de projection de la surface terrestre sur une surface réglée (plan, cône ou cylindre).

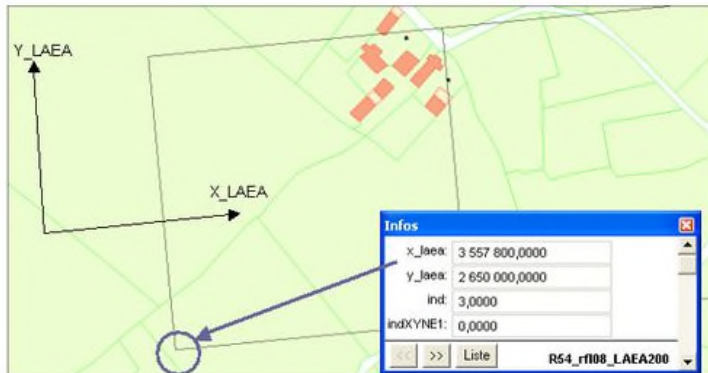
En France, sur la métropole, quatre principales alternatives existaient :

1. Le Lambert II (projection sur des cônes) encore en usage dans le dispositif de production à l'Insee.
2. Le Lambert 93 (projection sur des cônes) qui supprime depuis peu le précédent pour les échanges de données géographiques et qui permet de faire des triangulations exactes.
3. Le système UTM (projection sur des cylindres) qui, contrairement aux deux précédents, fournit un ensemble cohérent sur l'ensemble du globe terrestre, zones polaires exceptées.
4. Le système Lambert Azimutal Equal Area (LAEA) proposé par la directive INSPIRE, qui n'est pas réellement un système de projection mais qui est construit de manière à ce que les surfaces des carreaux soient strictement égales sur tout le territoire européen.

Derrière ce choix de système de projection, la question qui se pose est celle du territoire que pourra décrire le carroyage : un système de projection unique apporte évidemment plus de simplicité, même si plusieurs systèmes de projection ne constituent pas un handicap insurmontable. L'UTM est donc sacrifié au nom de ce seul argument puisqu'il faut trois zones pour décrire la métropole. Ensuite, entre Lambert 93 et le LAEA, le LAEA semble le meilleur candidat pour la seule raison qu'un système d'informations carroyées utilisant ce système a déjà commencé à se construire sous l'égide de la

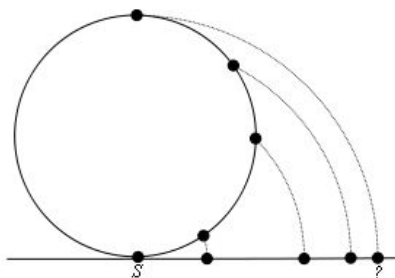
Commission Européenne.

C'est le positionnement du premier carreau qui guide celui de tous les autres. Si le parallélisme des cotés avec les axes du système de coordonnées va de soi, deux options étaient ensuite envisageables : placer le point zéro au centre du carreau ou sur l'un des points du contour. Bien que le premier choix aurait conduit à considérer les coordonnées du carreau comme une position moyenne de tous les points pouvant s'y trouver, c'est la seconde option qui a été retenue en raison de meilleures propriétés pour le calcul notamment avec des systèmes de carroyages emboîtés. En cas de besoin explicite de coordonnées pour les objets du carreau, il faudra donc tenir compte de ce léger décalage dans les deux directions.



Extrait de la base 2008, résolution 200 m. L'axe des y ne correspond pas avec la direction Sud-Nord. Le carreau au centre de l'image est identifié par la position de son coin Sud-Ouest. Il ne contient pas le ménage (étoile noire) situé le plus à droite qui a été positionné au centre de la parcelle et qui se trouve ainsi malencontreusement séparé des bâtiments d'habitation par le passage de la frontière de carreau. Le ménage le plus à gauche est compté dans le carreau mais des deux bâtiments, maison d'habitation et grange, c'est le second qui se trouve effectivement dans le carreau.

Le système de « projection » LAEA. Le point S correspond en pratique à un point arbitraire proche de Hanovre choisi comme centre géographique de l'Europe. Partant de là les points de l'ellipsoïde terrestre sont projetés sur le plan tangent à S selon des arcs de cercle. Le système conserve les directions des points vis à vis de S (c'est la signification du mot 'azimutal'), mais par contre coup ne peut conserver ni les angles ni l'orientation vers le Nord. Il faut revenir aux formules de conversion des coordonnées pour démontrer que le système conserve les surfaces.



Des petits et des grands carrés

La question de la taille des carreaux est bien plus importante que celle de leur forme réelle ou de leur orientation puisqu'elle va donner le « grain » de la photo constituée par l'ensemble des données carroyées.

En France, 1km de coté donne sur la métropole une image de 1000 x 1000 pixels correspondant à la définition moyenne d'un écran d'ordinateur, permettant donc une visualisation détaillée à l'échelle de ce territoire et a fortiori à l'échelle de l'Europe. Zoomer dans cette image est possible jusqu'à un certain point, peut être le département. Plus bas, l'image sera simplifiée à l'extrême mais notons quand même que 1km de coté c'est moins que la taille moyenne d'une commune et, en monde urbain, pas trop loin de la taille d'un IRIS, voire mieux : pour environ 14 000 IRIS d'habitat il y a environ 7 000 carreaux de plus 1 800 habitants.

S'il y a consensus autour de cette première résolution à 1km, ce n'est plus vrai en dessous : l'institut statistique autrichien utilise des subdivisions par deux : 500 m, 250 m et 125 m, tandis que d'autres des multiples de 100 m, alors que la directive INSPIRE dans ses spécifications sur les carroyages prône des tailles puissances de 10 : 100m, 10m, 1m, mais dans un contexte plus large que celui de la statistique.

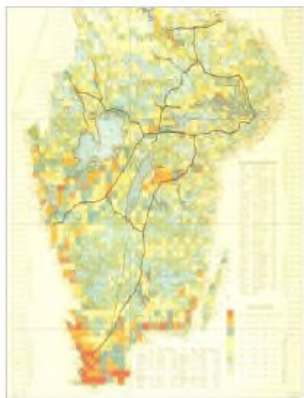
Pour ce qui est de la statistique humaine, la question sous jacente au choix d'une résolution est évidemment celle de la précision qu'on peut réellement espérer avoir pour la localisation d'un individu qui n'a rien d'un objet immobile. Même en faisant abstraction du fait qu'une personne ne reste pas dans son domicile toute la journée, une précision au mètre près est illusoire. Coté outils pour obtenir une localisation, à l'Insee le RIL ne garantit pas une précision en dessous de 100 m, tandis que l'utilisation des références cadastrales donne parfois une précision bien inférieure notamment en milieu rural où dans les grands ensembles de logements sociaux.

La deuxième livraison de données carroyées faite par l'Insee est organisée autour d'un compromis, 200 m, parce que l'objectif était une couverture complète du territoire national. Mais il reste que la question continue à mériter débat.

Les carroyages, ce n'est pas nouveau

À l'Insee, l'histoire des carroyages remonte à 2006, au moins pour ce qui est relatif à une utilisation systématique. À l'époque la problématique était l'identification de zones à enjeu infra-communales dans le contexte de la négociation débutante des Contrats Urbains de Cohésion Sociale (CUCS). Le carroyage offrait alors une réponse commode à la question de disposer d'un outil de description de la distribution spatiale des phénomènes socio démographiques à l'intérieur des territoires communaux.

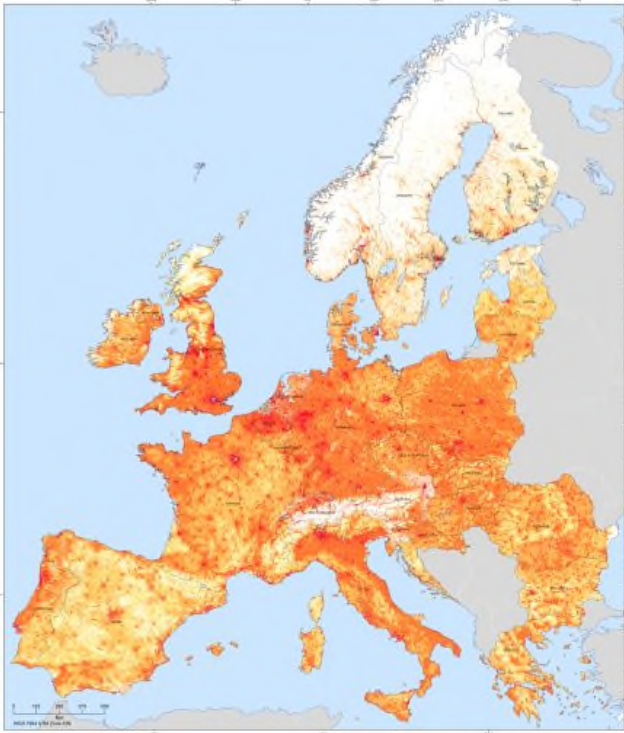
Néanmoins, l'outil n'est pas nouveau, loin s'en faut. Les pays nordiques ont sur ce sujet une « certaine » avance puisque le premier exemple recensé de statistiques au carreau remonte à 1857. Il s'agissait alors d'optimiser le tracé d'une des premières lignes de chemin de fer dans le sud de la Suède : une problématique clairement spatiale requérant plus la connaissance du mode de distribution de la population sur l'ensemble de la Suède que le dénombrement précis dans chaque sous unité géographique.



Sud de la Suède - Densité de population 1857

Depuis, plusieurs pays européens, ceux du nord, mais aussi l'Autriche, la Slovénie, la Suisse, les Pays Bas et plus ponctuellement le Royaume Uni ont mis à disposition des données carroyées. L'offre de données est généralement considérablement plus riche que de simples chiffres de population. D'un pays à l'autre ce qui fait la différence c'est bien sûr le système de projection mais surtout les conditions d'accès et les traitements appliqués pour garantir la confidentialité. Car la vraie difficulté est là puisque par construction, quelle que soit la taille du carreau on ne peut pas garantir a priori qu'il contiendra plus d'une personne. Les solutions actuellement en usage sont le blanchiment des cases à petits effectifs, le remplacement par un chiffre moyen ou un changement de taille de carreau. L'Insee étudie, lui, les possibilités offertes par le floutage qui aurait le mérite de masquer l'information individuelle tout en préservant l'image globale.

À l'échelle de l'Union Européenne, les choses s'organisent aussi petit à petit. On trouvera notamment sur le site de l'agence européenne de l'environnement (EEA) une carte de population carroyée à 1km. Une telle carte fait aujourd'hui partie de la production courante du Centre Commun de Ressources Européen (CCR ou JRC en anglais). Bien qu'elle soit majoritairement composée d'estimations basées sur la seule structure d'occupation du sol, elle commence à être utilisée systématiquement dans certains travaux statistiques : mesure de l'accessibilité des aéroports, typologie urbain/rural des maillages administratifs ou délimitation du champ de collecte de l'« audit urbain ».

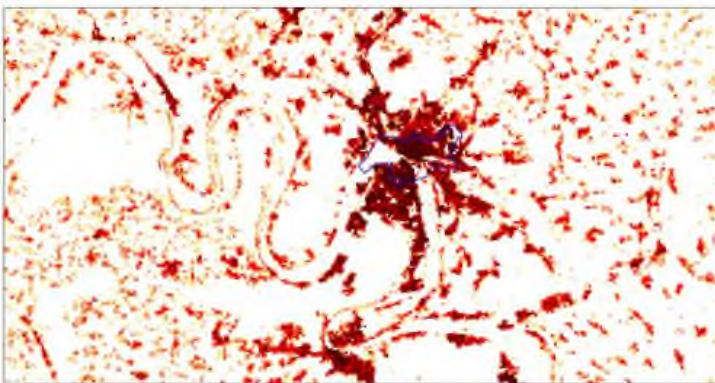


La première carte de densité de population sur carreaux de 1km² provenant des travaux du JRC. Elle combinait données « exactes » et estimations basées, à l'époque, sur le seul Corine Land Cover combiné avec les populations du niveau administratif le plus fin (LAU2). Des travaux d'amélioration sont en cours autour du processus d'estimation intégrant Soil sealing ainsi que par substitution aux estimations des données « exactes » que certains pays, comme la France, commencent à produire.

Des carrés pas si abstraits

On l'aura compris les carreaux ne tiennent absolument pas compte de la présence de l'être humain sur terre, sauf pour ce qui est de sa modélisation du globe. Les contours des carreaux ne suivent donc pas la voirie ou les éléments naturels et peuvent allègrement couper au travers de bâtiments, même de petite taille. Mais c'est bien sûr le but du jeu car tous ces éléments, bien que physiques, n'ont rien d'immuable et leur prise en compte dans un zonage entraînerait obligatoirement des remises en cause régulière de ce zonage. Ce n'est pas pour autant que les carreaux ne fournissent pas une image lisible où on peut se repérer, tout comme un écran d'ordinateur avec ses pixels est au fond aussi lisible qu'un dessin sur table traçante bien moins pratique à l'usage.

Le conflit supposé entre les carreaux et la réalité est en fait basé sur autre chose que le mode de représentation. Quelle que soit la taille de la maille géographique, il est vain d'essayer de voir des détails plus fins que la taille de cette maille.

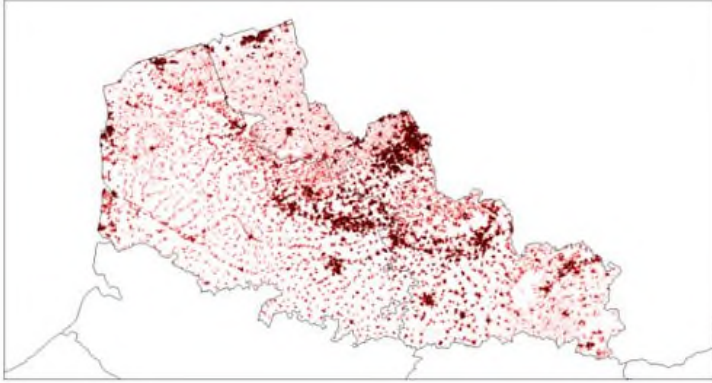


Extrait de la base 2008, résolution 200 m. Quelque part en France... une distribution de population qui en dit plus long sur la géographie physique que les zonages administratifs ne pourraient le faire (en bleu le contour de la commune principale de la zone).

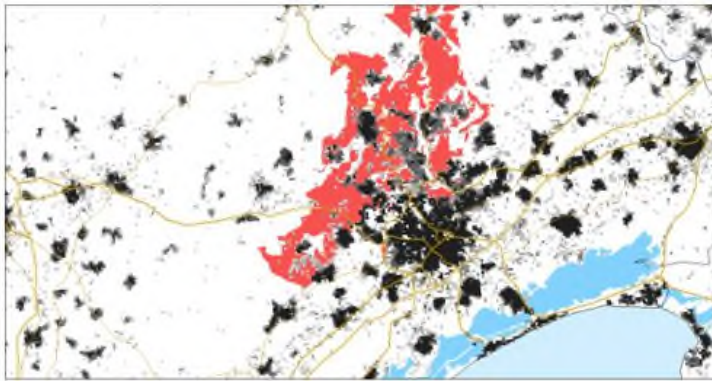
Données carroyées - Exemples

Quelques exemples cartographiques

Exemple d'exploitation : densité de population fiscale au 31/12/2008 sur la région Nord-Pas de Calais. Les données font ressortir une structure d'habitat présente nulle part ailleurs en France.



Risque incendie (en rouge, source Cartorisque, Ministère de l'Écologie, du Développement Durable, des Transports et du Logement) et distribution de population fiscale au 31/12/2008 (en gris) autour de Montpellier.



Visualiser la hiérarchie urbaine

La carte de France à base de carreaux de 1km de côté donne une image précise de la distribution de population. Néanmoins elle ne permet pas d'appréhender au premier coup d'œil la façon dont les villes sont distribuées sur le territoire. Pour passer de la donnée de base au carreau à un message sur un moins grand nombre d'entités il faut procéder à une simplification de la carte. Une façon de faire réside dans la technique du noyau, qui est en fait une moyenne mobile spatiale : en tout point on remplace la valeur connue par le résultat d'un calcul du type :

$$\hat{f}(s) = \sum_{i=1}^N \frac{1}{N\bar{h}} K\left(\frac{s - X_i}{\bar{h}}\right)$$

où :

- s est la position où est fait le calcul
- Xi les valeurs aux autres positions
- N le nombre de valeurs connues
- K une fonction dite « noyau » (« kernel » en anglais), généralement choisie pour avoir une valeur K(0) maximale et K(1) nulle.
- h une distance qui définit le voisinage de s sur lequel sera restreint le calcul, la « fenêtre ».

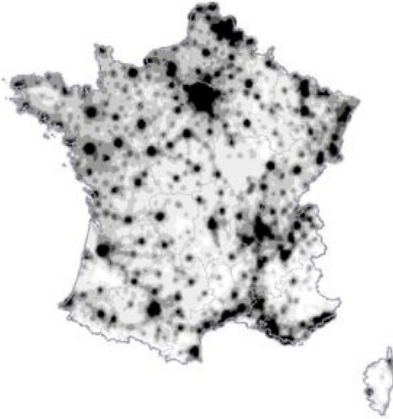
Ce type de calcul est bien connu pour donner un résultat plus « lisse » que l'original. Une meilleure interprétation qu'un simple enjolivement de la carte consiste plutôt à présenter cela comme un dé-bruitage d'un signal. D'ailleurs, dans les exemples qui suivent, la fonction noyau qui a été utilisée est une fonction gaussienne :

$$\frac{1}{2\pi h_x h_y} \exp\left(-\frac{\left(\frac{x}{h_x}\right)^2 + \left(\frac{y}{h_y}\right)^2}{2}\right)$$

qui ne s'annule pas complètement pour K(1) mais présente l'intérêt d'autoriser une optimisation de calculs potentiellement lourds par transposée de Fourier. De telles techniques ne sont pas étrangères au monde du traitement du signal, notamment radio.

Il faut noter ici que les données carroyées sont ainsi traitées comme des données ponctuelles. À l'échelle des images présentées, la surface des carreaux est négligeable et leur uniformité garantit l'absence d'artefacts. Cette ambiguïté des données carroyées entre données de surface et données ponctuelles est loin d'être exceptionnelle.

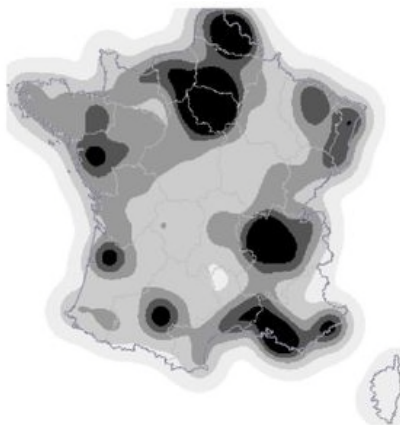
Retraitement des données de population avec une « fenêtre » de 3 km. La restitution de la carte (les pixels affichés) s'appuie sur un carroyage d'égalément 3 km. La structure du peuplement est simplifiée mais reste complexe :



Retraitement des données de population avec une « fenêtre » de 10 km. Les petites villes disparaissent de la carte et une ébauche de structure urbaine apparaît :



Retraitement des données de population avec une « fenêtre » de 20 km. Les villes moyennes disparaissent de la carte. La structure urbaine est simplifiée à l'extrême en une image qui pourrait être la représentation de la France que se ferait un aménageur travaillant à l'échelle européenne :



Zones à enjeu

L'étape suivante consiste à mobiliser d'autres données pour l'analyse. Les exemples qui suivent s'appuient sur des

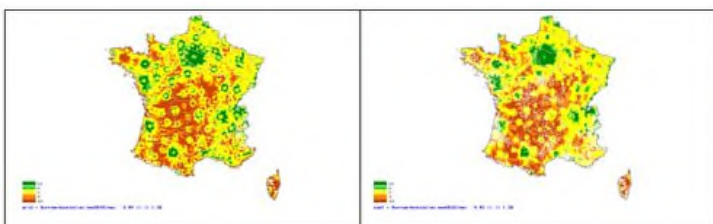
données qui ne sont pas encore sur www.insee.fr, elles servent surtout à éclairer la démarche.

Repérer des zones à enjeu c'est repérer des zones où un certain phénomène est plus particulièrement prégnant, ou, pour le statisticien, une zone à probabilité élevée d'observer, pour un individu de cette zone, une certaine caractéristique. Or les calculs présentés précédemment dans l'annexe I ne sont pas étrangers à cette notion de probabilité, en fait ils consistent très exactement en une recherche de la loi de probabilité de présence en un point. La réponse à la question qui consiste à déterminer la probabilité d'observation d'une caractéristique s'obtient par simple combinaison des lois de probabilité de présence de l'ensemble de la population et de celle de la population ayant la caractéristique donnée. Un peu de calcul sur les probabilités conditionnelles achève de guider vers l'opération à réaliser : à un terme constant près il s'agit d'un simple rapport.

$$prob(A|s) = prob(A) \frac{\hat{f}(A)}{\hat{f}(T)}$$

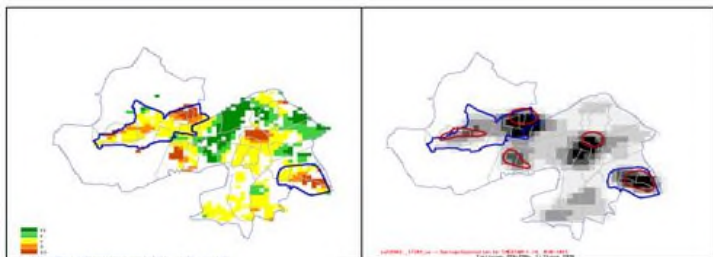
Un premier exemple utilise conjointement des dénombrements de ménages et de ménages dont le référent fiscal a plus de 65 ans. Les deux cartes qui suivent ont été fabriquées avec le même paramétrage : fenêtre à 5 km. Celle de gauche utilise des données carroyées à 1km, celle de droite des données communales classiques « spatialisées » par positionnement de l'ensemble des ménages au centre géométrique de chaque commune. Les taches blanches correspondent à des points (des carreaux de 5 km) où le calcul décrit précédemment n'a pas pu être fait en raison d'une population nulle à cet endroit.

En rouge les zones à forte concentration de ménages âgés, en vert les zones à faible concentration.



L'image de gauche, à base de données carroyées plus détaillées, décrit bien mieux un phénomène en forme de halos périurbains de ménages plus jeunes autour de villes mixtes. L'amélioration de l'image vient d'un effet combiné d'un nombre d'observations plus important et d'une maille plus régulière découpant mieux l'espace.

Le même principe peut être appliqué sur des zones plus réduites pour peu de disposer de données suffisamment détaillées géographiquement. Les deux cartes qui suivent proviennent d'un traitement de données sur les bas revenus élaborées en collaboration avec une Caisse d'Allocations Familiales (CAF), elles représentent les zones à forte concentration de bas revenus en rouge et orange sur la carte de gauche, sous forme de délimitations de zones (en rouge) sur la carte de droite qui contient, en fond, une image de la densité de population sur le territoire. Les contours bleus sont ceux des zones urbaines sensibles (ZUS).



Le résultat est une forme d'évaluation de ces territoires qui, apparemment, ne sont pas les seules zones de pauvreté.