

LCSQA

Laboratoire Central de Surveillance de la Qualité de l'Air



Programme 2014

Méthodologie de répartition spatiale de la population

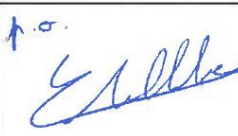
Laboratoire Central de Surveillance de la Qualité de l’Air

METHODOLOGIE DE REPARTITION SPATIALE DE LA POPULATION

MISE A JOUR 2014

Laurent Létinois

Avril 2015

	Rédaction	Vérification	Approbation
NOM	Laurent LÉTINOIS	Bertrand BESSAGNET	Laurence ROUIL
Qualité	Ingénieur de l’Unité MOCA Direction des Risques Chroniques	Responsable de l’Unité MOCA Direction des Risques Chroniques	Responsable du Pôle DECI Direction des Risques Chroniques
Visa			



LE LABORATOIRE CENTRAL DE SURVEILLANCE DE LA QUALITE DE L'AIR

Le Laboratoire Central de Surveillance de la Qualité de l'Air est constitué des laboratoires de Mines Douai, de l'INERIS et du LNE. Il mène depuis 1991 des études et des recherches à la demande du Ministère chargé de l'environnement, et en concertation avec les Associations Agréées de Surveillance de la Qualité de l'Air (AASQA). Ces travaux en matière de pollution atmosphérique ont été financés par la Direction Générale de l'Énergie et du Climat (bureau de la qualité de l'air) du Ministère de l'Écologie, du Développement durable et de l'Énergie (MEDDE). Ils sont réalisés avec le souci constant d'améliorer le dispositif de surveillance de la qualité de l'air en France en apportant un appui scientifique et technique au MEDDE et aux AASQA.

L'objectif principal du LCSQA est de participer à l'amélioration de la qualité des mesures effectuées dans l'air ambiant, depuis le prélèvement des échantillons jusqu'au traitement des données issues des mesures. Cette action est menée dans le cadre des réglementations nationales et européennes mais aussi dans un cadre plus prospectif destiné à fournir aux AASQA de nouveaux outils permettant d'anticiper les évolutions futures.

TABLE DES MATIERES

RESUME	7
REMERCIEMENTS ET COLLABORATIONS.....	8
LISTE DES ABREVIATIONS ET ACRONYMES	9
1. INTRODUCTION	10
2. DEFINITIONS.....	10
2.1 LA POPULATION EN FRANCE	11
2.1.1 LE RECENSEMENT DE LA POPULATION EN FRANCE	11
2.1.2 LE DÉCOUPAGE ADMINISTRATIF INSEE	13
2.1.3 DIFFUSION DES RÉSULTATS DU RECENSEMENT.....	14
2.2 LES BASES DE DONNÉES SPATIALES	14
2.2.1 LE RÉFÉRENTIEL À GRANDE ÉCHELLE (RGE)	15
2.2.1.1 LA BD TOPO	15
2.2.1.2 LA BD PARCELLAIRE	16
2.2.1.3 LA BD ADRESSE, POINT ADRESSE	17
2.2.2 LA BD CONTOURS...IRIS	18
2.3 TRAITEMENT SPATIAL.....	19
2.3.1 CALCUL D'INTERSECTION	19
2.3.2 AGRÉGATION SPATIALE	19
2.3.3 UTILISATION D'UN BUFFER	20
2.3.4 GEOCODAGE D'ADRESSE	20
3. LA METHODOLOGIE MAJIC.....	21
3.1 RECOMMANDATIONS SUR LES TRAITEMENTS ET VOCABULAIRE.....	22
3.2 Correction de la BD Contour...IRIS	23
3.2.1 Correction des codes IRIS	23
3.2.2 Correction de la superposition des IRIS	24
3.2.3 Correction des contours communaux des IRIS	25
3.3 Méthodologie MAJIC.....	29
3.3.1 Les données MAJIC.....	29
3.3.2 Rappel méthodologique.....	31
3.3.2.1 La méthodologie MAJIC ancienne version	31
3.3.2.2 Calcul de la surface habitable d'un bâtiment	32

3.3.2.3	Détermination du nombre de niveaux d'un bâtiment	32
3.3.3	Les limites de l'ancienne méthodologie	32
3.3.4	La méthodologie MAJIC finale	34
3.4	Application de la méthodologie et validation des données	42
3.4.1	Logiciels et outils utilisés.....	43
3.4.2	Données utilisées	43
3.4.3	Données élaborées	44
3.4.4	Validation des données.....	44
3.5	Périmètre d'utilisation des données de population	45
4.	CONCLUSIONS ET PERSPECTIVES.....	45
5.	REFERENCES.....	47

La législation européenne sur la surveillance de la qualité de l'air requiert la cartographie des zones géographiques de dépassement d'une valeur limite et l'estimation du nombre d'habitants exposés au dépassement. De nombreuses cartographies sont élaborées au niveau local et national pour répondre à cette exigence. Les cartographies des populations exposées à la pollution de l'air ambiant nécessitent deux variables : les concentrations de polluant d'une part et la population d'autre part, ainsi qu'une méthodologie permettant de croiser ces deux informations. Le LCSQA a été chargé de travailler sur cette problématique afin d'harmoniser les méthodes employées en France dans le domaine de la surveillance de la qualité de l'air. Le présent rapport s'attache spécifiquement au calcul de la répartition spatiale des populations. Complétant des travaux réalisés en 2013, il développe une approche adaptée à toutes les résolutions spatiales rencontrées pour une étude de la qualité de l'air.

La méthode de spatialisation nommée « MAJIC » permet une description très fine de la population à une échelle locale. Elle exploite plus particulièrement les données des locaux d'habitation de la base MAJIC foncière délivrée par la DGFIP. Ces données sont croisées avec des bases de données spatiales de l'IGN (BD PARCELLAIRE et BD TOPO) et les statistiques de population de l'INSEE pour estimer un nombre d'habitants dans chaque bâtiment d'un département. Les limites de la méthodologie identifiées lors de la précédente étude ont pu être levées, ce qui a permis d'appliquer cette méthode en tout point du territoire. Cette méthodologie garantit ainsi une homogénéité des données de population spatialisées utilisées dans le cadre de la surveillance de la qualité de l'air, que ce soit au niveau local ou au niveau national.

Les travaux menés par le LCSQA dans la présente étude ont permis d'automatiser les traitements pour limiter au maximum les futurs travaux nécessaires à la mise à jour de la base de données de population. Du fait de contraintes sur l'accès aux données MAJIC et de la complexité de la méthodologie MAJIC, le LCSQA assure la mise en œuvre de cette approche et met à disposition des AASQA les données spatiales de la population qui en sont issues. Ces données nécessitant une validation locale, des échanges sont prévus entre les AASQA et le LCSQA pour corriger si nécessaire les données élaborées par le LCSQA.

REMERCIEMENTS ET COLLABORATIONS

Le LCSQA remercie le CEREMA (ancien CETE de LYON) pour les nombreux échanges effectués sur la méthodologie MAJIC.

Le LCSQA remercie également la DGALN qui, par l'intermédiaire de François SALGE, a permis l'obtention d'une convention de mise à disposition des données MAJIC pour l'INERIS.

Le LCSQA remercie le CETE Nord-Picardie qui a délivré l'ensemble des bases de données MAJIC et fourni de précieux conseils sur le contenu et l'utilisation de ces données.

Le LCSQA remercie enfin les AASQA pour leur participation active aux tests des données de population spatialisée selon la méthodologie MAJIC.

LISTE DES ABREVIATIONS ET ACRONYMES

API	Application Programming Interface : un ensemble normalisé de classes, de méthodes ou de fonctions qui sert de façade par laquelle un logiciel offre des services à d'autres logiciels
CEREMA	Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement
CERTU	Centre d'Etudes sur les Réseaux, les Transports, l'Urbanisme et les constructions publiques
CETE	Centres d'Etudes Techniques de l'Equipement
Code EPSG	code unique définissant un système de projection géographique
DGFIP	Direction Générale des Finances Publiques
Fichiers MAJIC	Mise A Jour de l'Information Cadastreale
IGN	Institut national de l'information géographique et forestière
INSEE	Institut National de la Statistique et des Etudes Economiques
INSPIRE	Infrastructure for Spatial Information in the European Community
IRIS	Ilots Regroupés pour l'Information Statistique - Découpage infra-communal INSEE pour toutes communes d'au moins 10 000 habitants et une forte proportion de communes entre 5000 et 10 000 habitants
Parcelle cadastrale	portion de terrain d'un seul tenant, située dans un même lieu-dit (lui-même situé dans une section cadastrale, elle-même située dans le territoire communal), appartenant à un même propriétaire ou à une même indivision et constituant une unité foncière indépendante.
PostGIS	composante spatiale du système de base de données PostgreSQL
PostgreSQL	système de base de données
Raster	dans le domaine des SIG, ce sont des images pixellisées géoréférencées
RGE	Référentiel à Grande Echelle (http://professionnels.ign.fr/rge)
RGP	Recensement Général de la Population
RPCU	Représentation Parcellaire Cadastreale Unique
TRIRIS	Un TRIRIS est un regroupement d'IRIS (en général 3 IRIS).
SETRA	Service d'Etudes sur les Transports, les Routes et leurs Aménagements
SIG	Système d'Information Géographique

1. INTRODUCTION

La législation européenne sur l'évaluation de la qualité de l'air impose, lors du dépassement d'un objectif environnemental (valeur limite, valeur cible, objectif de long terme), de déterminer la zone de couverture du dépassement et la population impactée. Les textes précisent que les populations concernées sont les personnes résidant sur cette zone de dépassement. En France, l'évaluation des critères de déclenchement des procédures préfectorales d'information et d'alerte nécessite également le calcul de la surface et de la population potentiellement exposées à des dépassements de seuil.

Afin de répondre au mieux à ces exigences réglementaires, des travaux sont menés par le LCSQA depuis 2012 sur l'estimation de la répartition spatiale de la population. Une première étude¹ du LCSQA a permis de recenser les méthodologies employées en France et d'établir deux méthodologies de spatialisation applicables pour l'une à un niveau départemental ou national et pour une seconde à un niveau local. Des limites dans les données utilisées pour spatialiser la population ne permettaient pas d'établir une méthodologie unique quelle que soit l'échelle géographique étudiée mais des pistes d'amélioration avaient pu être relevées.

Le travail réalisé en 2014 a permis de lever ces verrous afin de constituer une méthodologie unique applicable sur l'ensemble du territoire français.

Dans une première partie, le présent rapport rappelle différentes définitions utiles pour mieux appréhender la répartition spatiale de la population : le recensement de la population en France et les bases de données spatiales disponibles sur le territoire français. La seconde partie présente la nouvelle mouture de la méthodologie de la spatialisation de la population, l'application de la méthodologie sur l'ensemble des départements français et les incertitudes restantes sur les données de population élaborées.

2. DEFINITIONS

Avant d'étudier les méthodes d'évaluation de la répartition spatiale des populations, il est utile de poser quelques définitions pour faciliter la compréhension de la problématique et la lecture du présent document.

¹ Létinois L., 2013. Méthodologie de répartition spatiale de la population. Rapport LCSQA, http://www.lcsqa.org/system/files/drc-12-126716-10615a_methodologiepopulation_0.pdf

2.1 LA POPULATION EN FRANCE

2.1.1 LE RECENSEMENT DE LA POPULATION EN FRANCE

Le recensement de la population a été institué en France en 1801. Depuis lors, et jusqu'à la mise en application de la loi citée ci-après, de nombreux recensements ont été réalisés à intervalles de temps plus ou moins réguliers. Outre le nombre de personnes résidant par foyer, des informations complémentaires étaient recueillies telles que la nationalité, l'âge, le sexe, la situation personnelle et la profession.

La loi du 27 février 2002, dite loi de démocratie de proximité, a renouvelé le recensement général : celui-ci est remplacé par un système de recensement permanent dit recensement renouvelé. Le comptage traditionnel est remplacé par des enquêtes de recensement annuelles. La méthode de recensement est adaptée à la taille de la commune étudiée :

- Les communes de moins de 10 000 habitants sont recensées exhaustivement une fois tous les 5 ans.
- Les communes de 10 000 habitants ou plus font l'objet d'une enquête annuelle auprès d'un échantillon de 8 % de la population, dispersé sur l'ensemble de leur territoire. Au bout de 5 ans, tout le territoire de ces communes est pris en compte et les résultats du recensement sont calculés à partir de l'échantillon de 40 % de la population ainsi constitué.

Ce sont les communes qui ont la responsabilité de préparer et de réaliser les enquêtes de recensement. L'INSEE, quant à lui, organise et contrôle la collecte des informations. Il exploite les questionnaires, établit et diffuse les chiffres de populations légales de chaque collectivité territoriale et de chaque circonscription administrative.

Ce nouveau système a été mis en place en 2004. Les premières statistiques qui en résultent ont été disponibles en 2009 pour l'année de référence 2006. Depuis cette date, chaque nouvelle année 'n' voit apparaître une nouvelle version se rapportant à l'année 'n-3'.

Le recensement de la population a pour objet : le dénombrement de la population de la France, la description des caractéristiques démographiques et sociales de la population et le dénombrement et la description des caractéristiques des logements. Le recensement recueille de nombreuses statistiques relatives aux habitants telles que le sexe, l'âge, l'activité, les professions exercées, les caractéristiques des ménages, la taille et le type de logement, les modes de transport utilisés, les déplacements quotidiens... Ces informations sont fournies par l'INSEE à différentes échelles géographiques variant du niveau national à un niveau infra-communal appelé IRIS.

Le décret du 5 juin 2003 précise le périmètre des populations recensées et définit les trois catégories de population recensée :

1. la *population municipale* regroupe :
 - a. Les personnes ayant leur résidence habituelle sur le territoire de la commune. Quelques cas particuliers sont traités :
 - i. Une personne mineure résidant ailleurs du fait de ses études reste attachée à la commune de la résidence familiale.

- ii. Une personne résidant dans une communauté est rattachée à la commune de la communauté.
 - iii. Un étudiant majeur vivant en internat est rattaché à la commune du lieu de ses études.
 - iv. Une personne majeure résidant du fait de ses études hors de la résidence familiale et hors communauté est recensée dans la commune de son logement.
 - v. Un conjoint, concubin ou personne liée par un pacte civil de solidarité résidant pour des raisons professionnelles hors de la résidence familiale et hors communauté, est recensé sur sa résidence familiale.
 - vi. Une personne ne se trouvant dans aucune des situations décrites précédemment est recensée dans la commune de la résidence dans laquelle elle réside le plus longtemps.
 - b. Les personnes détenues dans les établissements pénitentiaires dont le siège est situé sur le territoire de la commune ;
 - c. Les personnes sans abri recensées sur le territoire de la commune ;
 - d. Les personnes résidant habituellement dans des habitations mobiles, recensées sur le territoire de la commune.
2. la *population comptée à part* regroupe :
- a. Les personnes qui résident du fait de leurs études sur le territoire de la commune et qui ont leur résidence habituelle située dans une autre commune (situation décrite au i.a.1);
 - b. Les personnes dont la résidence familiale est située sur le territoire de la commune et qui ont leur résidence habituelle située dans une autre commune (situation décrite au ii.a.1);
 - c. Les personnes majeures âgées de moins de vingt-cinq ans dont la résidence familiale se trouve sur le territoire de la commune et qui ont leur résidence habituelle située dans une autre commune (situation décrite au iii.a.1);
 - d. Les personnes majeures âgées de moins de vingt-cinq ans dont la résidence de la famille se trouve sur le territoire de la commune et qui ont leur résidence habituelle située dans une autre commune (situation décrite au iv.a.1);
 - e. Les personnes sans domicile fixe rattachées, au sens de la loi du 3 janvier 1969, à la commune et non recensées sur le territoire de la commune.
3. et la *population totale* qui est la somme des deux catégories précédentes de population.

La population municipale ne comporte pas de doubles comptes et, de ce fait, est retenue pour les traitements statistiques.

Dans le présent document, nous nous intéresserons uniquement aux données de population municipale.

Le graphique Figure 1 présente les différentes catégories de populations légales dans le recensement. Le terme générique de « populations légales » regroupe pour chaque commune sa population municipale, sa population comptée à part et sa population totale qui est la somme des deux précédentes.

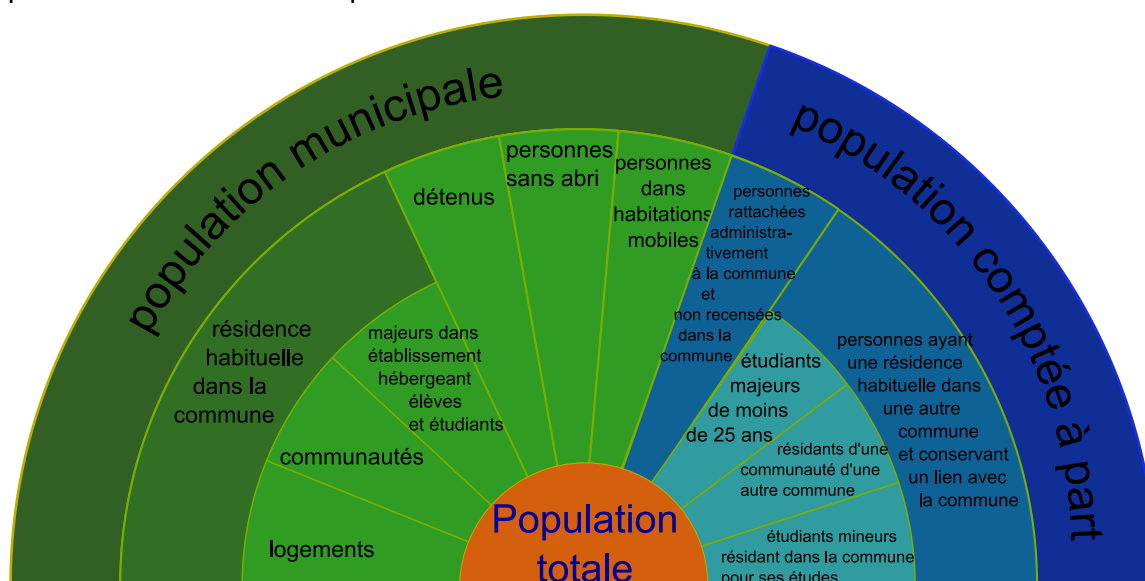


Figure 1 : Les catégories de populations légales

2.1.2 LE DÉCOUPAGE ADMINISTRATIF INSEE

Les découpages administratifs officiels sont les régions, les départements, les communes et les arrondissements. Lors du recensement de la population de 1999, l'INSEE a développé un découpage du territoire en mailles de taille homogène appelées IRIS 2000. Le sigle IRIS signifie « Ilots Regroupés pour l'Information Statistique » et fait référence à la taille visée de 2 000 habitants par maille élémentaire. Depuis la réforme du recensement de 2002, ces découpages sont simplement appelés IRIS.

Les IRIS respectent des critères géographiques et démographiques et ont des contours identifiables sans ambiguïté et stables dans le temps². Les communes d'au moins 10 000 habitants et la plupart des communes de 5 000 à 10 000 habitants sont découpées en IRIS. Les autres communes sont assimilées à un IRIS.

Trois types d'IRIS sont ainsi définis :

- Les IRIS d'habitat : leur population se situe en général entre 1 800 et 5 000 habitants. Ils sont homogènes quant au type d'habitat et leurs limites s'appuient sur les grandes coupures du tissu urbain (voies principales, voies ferrées, cours d'eau...).
- Les IRIS d'activité : ils regroupent plus de 1 000 salariés et comptent au moins deux fois plus d'emplois salariés que de population résidente.

² Si le contour géographique d'un IRIS était amené à évoluer du fait par exemple d'échanges de parcelles entre deux communes limitrophes. La référence de cet IRIS serait supprimée et une nouvelle serait créée avec les nouveaux contours.

- Les IRIS divers : il s'agit de grandes zones spécifiques peu habitées et ayant une superficie importante (parcs de loisirs, zones portuaires, forêts...).

2.1.3 DIFFUSION DES RÉSULTATS DU RECENSEMENT

Les différentes statistiques du recensement de la population sont disponibles sur le site internet de l'INSEE selon des résolutions différentes :

- Informations communales : ces informations sont généralement disponibles au début de l'année $n+3$ pour l'année de recensement n : les chiffres de l'année 2011 sont publiés au début de l'année 2014. Les chiffres sont présentés selon 12 thématiques et sont diffusés sur l'adresse suivante : <http://www.recensement.insee.fr/basesTableauxDetailles.action>.
- Informations infra-communales : ces données à l'IRIS sont disponibles la même année que les données communales mais seulement à partir du quatrième trimestre. Ces données sont diffusées selon 5 thématiques à l'adresse suivante : <http://www.insee.fr/fr/bases-de-donnees/default.asp?page=infracommunal.htm>.

Il est à noter qu'il existe les fichiers détail anonymisés. Ceux-ci visent à fournir un complément d'information et une possibilité d'analyse plus fine des données du recensement. Ils sont à destination d'utilisateurs avertis. Ces données sont disponibles à l'adresse internet suivante : <http://www.insee.fr/fr/bases-de-donnees/fichiers-detail.asp>.

2.2 LES BASES DE DONNÉES SPATIALES

La spatialisation de la population sur une zone administrative nécessite l'emploi d'un support spatial pour répartir cette population. Ce support est généralement constitué de bases de données spatiales qui répertorient, entre autres, les contours administratifs, la localisation et la typologie des bâtiments, le registre parcellaire français, les adresses ou le type d'occupation du sol.

Même si certaines de ces bases couvrent l'ensemble du territoire, celles-ci ne sont pas nécessairement à jour : la prise en compte de nouvelles constructions, la modification des parcelles cadastrales suite à un remembrement ou une fusion de communes ne sont pas immédiatement reportées. Ces bases de données sont donc en constante évolution pour essayer d'intégrer au fil de l'eau les nouveautés du territoire et corriger les imperfections remontées par les utilisateurs.

2.2.1 LE RÉFÉRENTIEL À GRANDE ÉCHELLE (RGE)

Le RGE décrit le territoire national et l'occupation de son sol de façon précise, complète et homogène. Cette infrastructure de données géographiques, partagée par un grand nombre d'acteurs publics et privés, est une référence indispensable pour comprendre, décider et communiquer, notamment dans les domaines de la protection de l'environnement, de l'aménagement du territoire, des transports, de l'agriculture et de la prévention des risques.

L'IGN a pour mission de constituer et mettre à jour ce référentiel. Depuis le 1er janvier 2011, le RGE est diffusé par l'IGN au seul coût de reproduction et de diffusion des données concernées, dès lors que celles-ci sont utilisées pour l'exercice d'une mission de service public ne revêtant pas un caractère industriel ou commercial. Pour ce cas, une convention doit être établie entre l'IGN et l'organisme utilisateur de ces données.

Le RGE se décompose en quatre grandes composantes qui fournissent des informations géographiques superposables entre elles : ortho photographique, topographique, parcellaire, adresse et altimétrie (Figure 2).

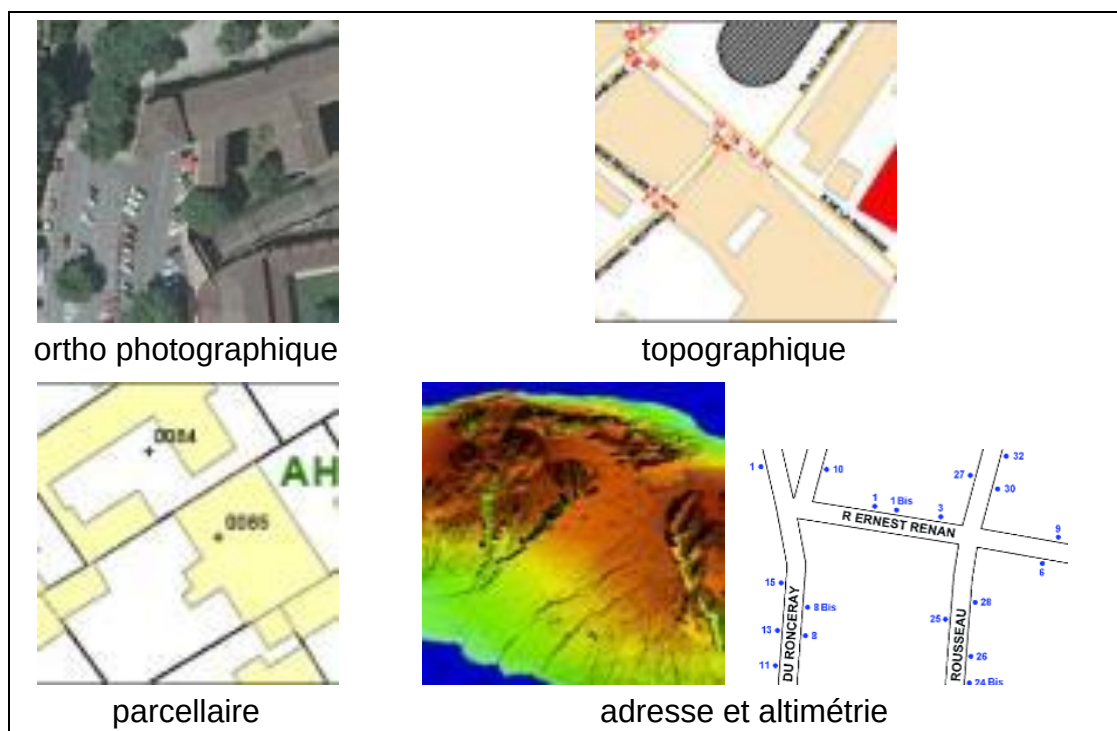


Figure 2 : Les quatre composantes du RGE

Nous nous attachons à présenter dans ce document trois de ces composantes utiles pour la spatialisation des populations : la composante topographique représentée par la BD TOPO, la composante parcellaire représentée par la BD PARCELLAIRE et la composante adresse représentée par la BD ADRESSE, POINT ADRESSE.

2.2.1.1 LA BD TOPO

La base de données spatiale BD TOPO est la base de données de référence en 3 dimensions. Elle décrit les éléments du paysage avec une précision métrique (Figure 3). Elle est fournie dans un format vectoriel. Cette base couvre la totalité du territoire français. Elle se décompose en différentes thématiques :

- réseau routier,
- voies ferrées et autres,
- transports énergie,
- hydrographie,
- bâti,
- végétation,
- orographie (lignes de rupture des pentes),
- administratif,
- zones d'activité,
- et toponymes.

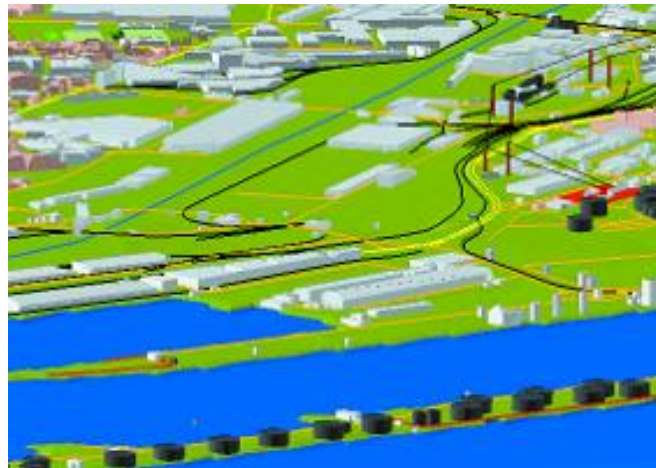


Figure 3 : Représentation en 3D du paysage grâce à la BD TOPO

2.2.1.2 LA BD PARCELLAIRE

Cette base de données fournit l'information cadastrale numérique, géoréférencée. Elle est construite à partir de l'assemblage du plan cadastral dématérialisé. Elle se compose de deux parties :

- Une couche de dalles "raster" de 1km*1km : les rasters sont des images pixélisées géoréférencées. Elles apportent uniquement une information visuelle, elles ne peuvent donc être utilisées pour produire des traitements SIG.



Figure 4 : plan cadastral présenté dans la BD PARCELLAIRE Image de l'IGN

- Une couche d'objets vecteurs décrivant entre autre :
 - chaque parcelle cadastrale (couche Parcelle)
 - chaque bâtiment (couche Bâtiment) en distinguant les bâtiments en dur (bâtiment industriel ou bâtiment possédant des fondations et fermé sur les 4 côtés) et les constructions légères.



Figure 5 : Plan cadastral présenté dans la BD PARCELLAIRE Vecteur de l'IGN superposé avec la BD Ortho

La BD PARCELLAIRE Vecteur possède une couverture non homogène sur le territoire français et très variable d'un département à un autre. Les travaux de mise à jour de cette base se poursuivent afin d'améliorer cette couverture et dureront encore quelques années avant d'obtenir une couverture totale.

2.2.1.3 LA BD ADRESSE, POINT ADRESSE

Dérivée de la BD Adresse, la base de données Point Adresse répertorie environ 26 millions d'adresses postales géolocalisées sur le territoire national (en urbain, inter-urbain et rural). Chaque adresse est localisée par un point.

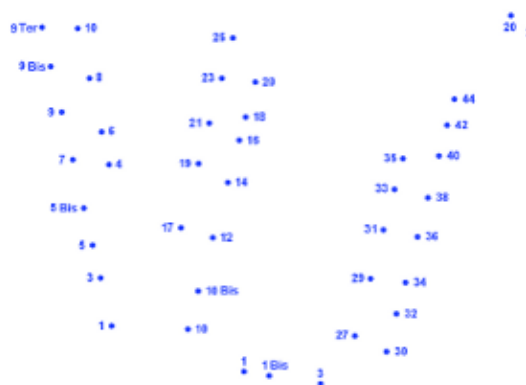


Figure 6 : Adresses localisées sur des points dans la BD Point Adresse

2.2.2 LA BD CONTOURS...IRIS

Cette base de données, coproduite par l'INSEE et l'IGN, représente les contours des IRIS sur le référentiel de l'année 2008. Elle permet, en théorie, de créer un lien direct entre les données de population et les contours géographiques associés. En pratique, le lien entre ces deux informations nécessite quelques ajustements car certains IRIS ont changé de référence depuis 2008.

Cette base de données comporte une précision de localisation des IRIS entre 10 et 100 m. La Figure 7 présente, pour des communes de moins de 5000 habitants, la précision de cette base en comparaison de la base BD TOPO. Cette « faible » qualité oblige l'utilisateur à prendre des précautions s'il souhaite travailler à une échelle fine. L'IGN ne recommande pas l'utilisation de cette base de données à une échelle communale.

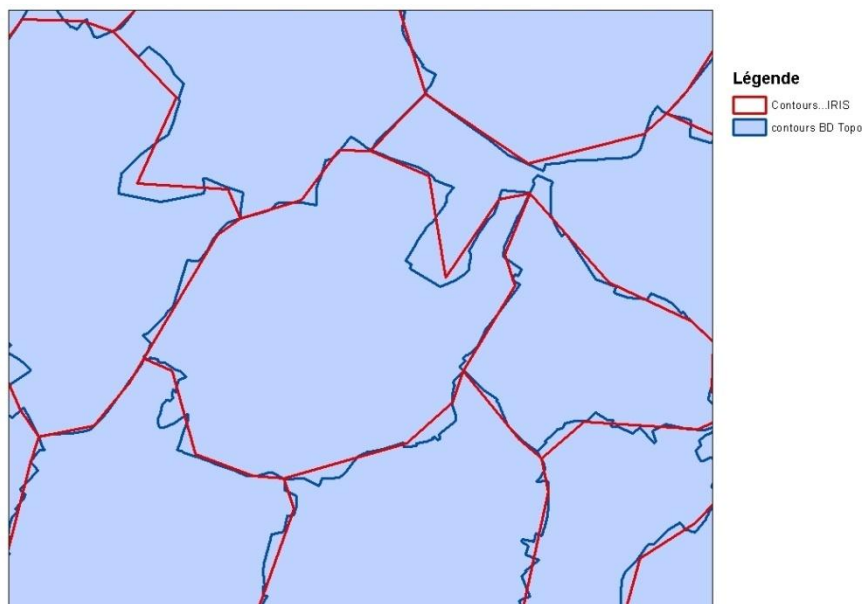


Figure 7 : Comparaison des contours IRIS de la BD Contours...IRIS avec les contours communaux de la BD TOPO

L'IGN produit actuellement la base de données spatiale Topo...Iris. Elle exploite les éléments de la BD TOPO pour corriger les contours des IRIS. Cette base ne couvre pas l'ensemble du territoire et possède un coût élevé de l'ordre de 10 000€ en moyenne par région.

2.3 TRAITEMENT SPATIAL

L'utilisation des SIG facilite le croisement spatial des informations. Certains traitements spatiaux utilisés dans ce document sont présentés dans cette partie.

2.3.1 CALCUL D'INTERSECTION

Il est souvent utile de rechercher les zones de recouvrement ou l'intersection géométrique entre deux couches géographiques. Les résultats de ce calcul sont des objets géographiques distincts représentant chacun les zones d'intersection entre objets. Chaque objet résultant contient les données attributaires des deux objets intersectés. Ce traitement est utilisé à de nombreuses reprises dans ce document notamment pour des calculs de ratio de population : part de la surface d'un bâtiment sur la totalité d'un IRIS, croisement de bâtiments avec un carroyage.

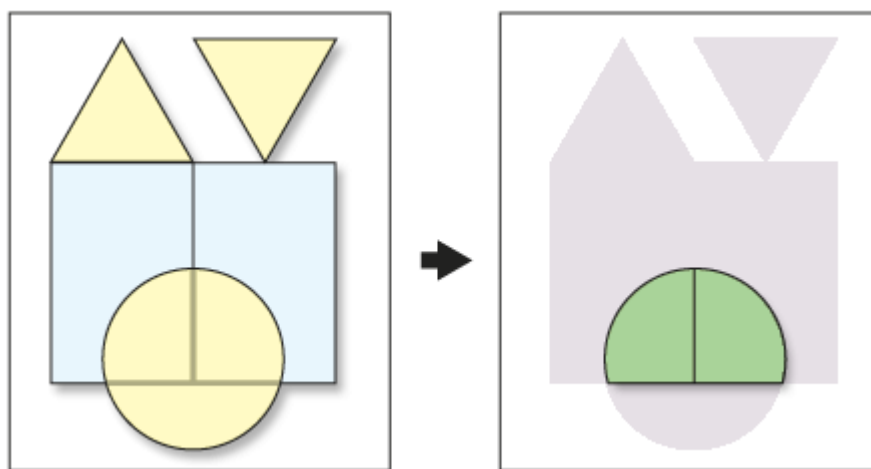


Figure 8: A droite : résultat de l'intersection des couches géographiques de la partie gauche (une en bleu et l'autre en jaune)

2.3.2 AGRÉGATION SPATIALE

Il est souvent utile suite à une intersection entre deux couches géographiques d'agréger les morceaux d'intersection en fonction d'un attribut spécifié commun entre les objets, par exemple le code IRIS, code commune ou l'identifiant d'une maille. L'agrégation spatiale fusionne les objets géographiques entre eux pour obtenir des régions uniques en fonction de chaque valeur unique de l'attribut.

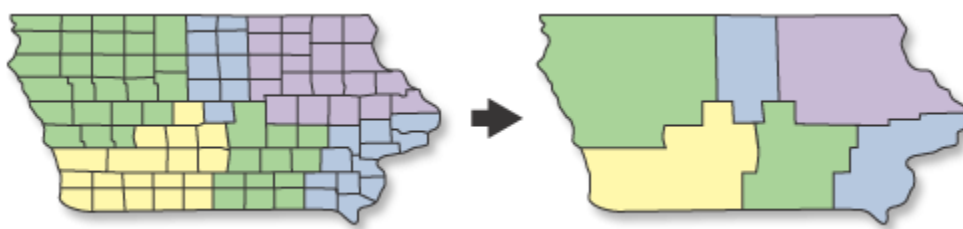


Figure 9 : Fusion d'entités géographiques en fonction d'un attribut commun

2.3.3 UTILISATION D'UN BUFFER

L'application d'un zone tampon ou buffer est souvent employée pour rechercher des entités géographiques à proximité d'une entité de référence : par exemple, quels sont les restaurants dans un rayon de 1 km autour d'une rue, quelles sont les émissions d'un polluant autour d'une station de mesure, quelle population habite dans un rayon de 10 km autour d'une usine... Les applications liées aux zones tampons sont très variées. Dans le présent document, elles sont utilisées notamment pour rechercher des entités géographiques voisines (même frontière) d'une entité cible. Pour cela, le rayon de recherche utilisé est de l'ordre du mètre. Elles sont également utilisées pour rechercher des bâtiments à proximité d'autres bâtiments ou de centres de parcelles selon un certain un rayon de recherche fixé.

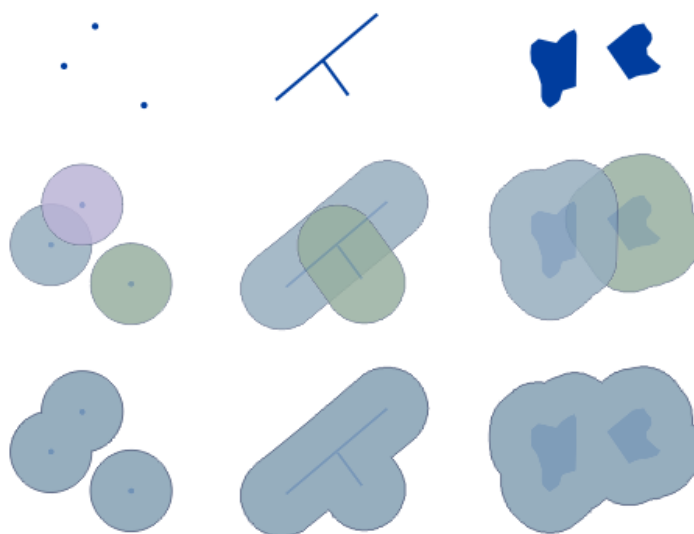


Figure 10 : Application de zones tampon sur des entités ponctuelles, linéiques et surfaciques

2.3.4 GEOCODAGE D'ADRESSE

Le géocodage est un procédé qui consiste à transformer la description d'un emplacement, telle qu'une paire de coordonnées, une adresse ou un nom de lieu, en un emplacement à la surface du globe. Les emplacements obtenus en sortie du géocodage sont des entités géographiques (généralement ponctuelles) avec des attributs (définition de l'adresse par exemple).

Différents outils de géocodage sont disponibles. Certains sont des outils « clés en main » et permettent de trouver directement la localisation d'un lieu à partir de son adresse. Ils font généralement appel à des API de géocodage (Google Map, Yahoo, Bing ou autres) et se présentent sous la forme d'applications web ou de librairies de fonction utilisables dans différents environnements de développement. Le paramétrage de ces outils reste limité.

Des outils SIG proposent également des outils de géocodage qui doivent nécessairement être associés à une base de données spatiale d'adresses (par exemple la BD adresse de l'IGN). La qualité du géocodage est dépendante de la qualité de la base de données d'adresses de référence utilisée mais également de la qualité de l'adresse à géocoder. Ces outils SIG sont généralement difficiles à mettre en œuvre car ils nécessitent un formatage et une traduction correcte de l'adresse avant le géocodage.

Prenons l'exemple de l'adresse '11 av. du Général de Gaulle, 60100 CREIL'. Celle-ci doit être traduite et formatée avant une utilisation dans un géocodeur. Selon la base de données d'adresses utilisée, il sera probablement nécessaire de découper l'adresse en plusieurs parties : le numéro de la voie, le nom de la rue, le code postal ou INSEE de la commune et le nom de la commune. Dans l'exemple présenté le type de la rue est une avenue, le terme 'av.' devra peut-être être traduit en 'avenue'. De même, cette voie est peut-être référencée dans la base de données comme étant 'avenue Général de Gaulle', le terme 'du' devrait peut-être être retiré de l'adresse. Cet exemple démontre la complexité du géocodage d'une adresse.

3. LA METHODOLOGIE MAJIC

Les travaux conduits par le LCSQA en 2012 sur la répartition spatiale de la population avaient abouti à une proposition de méthodologie fine de spatialisation : la méthodologie MAJIC. Elle était proposée comme méthodologie de référence pour l'échelle locale. Cette méthodologie était limitée par la qualité des données MAJIC qui n'étaient pas vectorisées sur l'ensemble du territoire. Cette limitation imposait l'utilisation de la méthodologie BD TOPO, moins précise, sur les zones non couvertes et ne permettait pas d'obtenir une méthodologie unique applicable sur l'ensemble du territoire français.

A l'échelle régionale ou nationale, le LCSQA avait proposé la méthodologie carroyage. Cette méthodologie corrigeait les données de populations carroyées de l'INSEE³ à partir des données de population RGP spatialisées à l'IRIS.

Le présent document propose d'appliquer la méthodologie MAJIC quelle que soit la qualité de spatialisation des données MAJIC, permettant ainsi d'employer une méthodologie unique sur l'ensemble de la Métropole et des DROM. En remplacement de la méthodologie carroyage, l'agrégation de ces populations sur des mailles kilométriques garantit une cohérence des données quel que soit le niveau géographique exploité.

³ http://www.insee.fr/fr/themes/detail.asp?reg_id=0&ref_id=donnees-carroyees&page=donnees-detaillees/donnees-carroyees/donnees-carroyees-200m.htm

3.1 RECOMMANDATIONS SUR LES TRAITEMENTS ET VOCABULAIRE

La mise en œuvre de la méthodologie nécessite l'utilisation de logiciels SIG et l'emploi de bases de données ayant une volumétrie conséquente.

Avant tout croisement d'informations, qu'elles soient spatiales ou tabulaires, il est essentiel d'évaluer la qualité des données employées et leur cohérence. Il est également utile, lorsqu'une méthodologie comprend de nombreuses étapes de traitement, de vérifier au fil de ces étapes la conservation des surfaces ou des populations et de valider les résultats finaux avec des bilans communaux, régionaux ou nationaux.

Parmi les vérifications à réaliser figurent notamment les contrôles suivants :

- Le référentiel de la population est délivré par l'INSEE (IRIS ou commune), la mise en relation des données du recensement avec les contours administratifs s'effectue par jointure entre les codes INSEE des deux sources de données. Ces codes INSEE pouvant évoluer d'une année sur l'autre suite à une fusion de communes ou un déplacement de frontières communales, il est nécessaire de vérifier que la jointure n'exclut pas certaines zones géographiques. Le nombre d'entités administratives d'origine doit être égal au nombre d'entités administratives finales. Dans le cas d'un changement de code INSEE, il est nécessaire d'adapter les fichiers fournis par l'INSEE ou de créer une table de passage d'une année à une autre.
- Quelle que soit l'origine des bases de données spatiales (commerciale ou non), la qualité de la donnée spatiale n'est pas nécessairement celle attendue. Différents types d'erreur peuvent être relevés :
 - Deux zones géométriques se superposent : selon le traitement effectué, la superposition de ces zones peut biaiser les résultats. Prenons l'exemple d'un calcul de pondération effectuée sur des IRIS à l'échelle d'une commune et supposons que certains de ces IRIS se recouvrent partiellement. Si on souhaite affecter un poids à chacun des IRIS selon leur surface, la somme des surfaces de ces IRIS sera supérieure à la surface réelle de la commune et nécessairement la somme des poids sera supérieure à 1. De ce fait, le poids des IRIS concernés par cette superposition sera surestimé par rapport aux autres IRIS.
 - Il existe un vide (trou) entre deux zones géométriques : à l'inverse de l'exemple précédent, les poids des IRIS seront sous-estimés et la somme des poids sera inférieure à 1.
 - La géométrie d'un polygone comporte des auto-intersections sur ses propres contours. Ce type d'erreur peut engendrer des erreurs lors de calcul d'intersection entre deux couches. Des outils sont généralement fournis avec les logiciels SIG pour détecter et corriger ces erreurs.

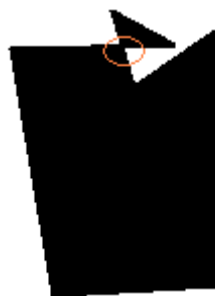


Figure 11 : Auto-intersection sur un polygone (zone orange)

- Il est préférable de travailler sur des couches SIG comportant un même système de coordonnées géographiques. En effet, lorsque deux couches sont croisées ensemble (intersection ou union), le logiciel SIG ne gère pas nécessairement automatiquement et correctement la projection d'un système de coordonnées vers un autre. Des écarts de plusieurs dizaines de mètres peuvent alors apparaître sur ces traitements. Il est donc préférable de projeter dans un premier temps les couches dans un référentiel unique avant tout croisement de ces données.

3.2 Correction de la BD Contour...IRIS

Les données de population à l'IRIS sont une des sources de données pour répartir la population. Il est donc nécessaire de disposer des contours géographiques de ces IRIS. Les bases à disposition sont la BD Contours...Iris et la BD TOPO...Iris (cf. 2.2.2). La BD TOPO...Iris, la plus fine en résolution, est la base recommandée par l'IGN pour une utilisation à une échelle locale. Celle-ci étant onéreuse, le LCSQA a choisi d'utiliser la BD Contours...Iris. Le LCSQA propose une méthode pour améliorer les imperfections de cette base : une résolution « grossière » et des superpositions d'IRIS.

3.2.1 Correction des codes IRIS

Les codes IRIS utilisés dans cette base font référence aux codes IRIS de l'année de recensement 2008. Depuis cette date, certains codes IRIS ont évolué. L'INSEE fournit année par année un suivi des modifications apportées sur les IRIS. Ces changements sont disponibles à l'adresse suivante : <http://www.insee.fr/fr/methodes/default.asp?page=zonages/iris.htm> dans le fichier `table_passage_IRIS_geo2013_geo2008.xls` (pour l'année 2013). Il n'est pas toujours aisé d'apprécier les changements effectués dans la codification de certains IRIS. En effet, par exemple, deux IRIS échangent quelques parcelles entre eux : le nombre d'habitants pour ces deux IRIS reste quasiment identique par contre le contour de ceux-ci est nécessairement modifié. Faute d'informations complémentaires, les contours existants sont conservés.

3.2.2 Correction de la superposition des IRIS

Lors de la première utilisation de Contours..., le LCSQA a constaté l'existence dans cette base de chevauchements de frontières entre des IRIS voisins (Figure 12). Ces superpositions ne représentent pas la réalité du terrain et faussent les calculs de ratio lors du croisement des IRIS avec d'autres couches géographiques.



Figure 12 : Chevauchement d'IRIS sur l'agglomération de LILLE

En l'absence d'autres informations pour déterminer où se trouvent les frontières réelles des IRIS, il a été choisi simplement de retenir le premier IRIS couvrant cette intersection.

La méthode employée est la suivante :

- On effectue une intersection de la BD Contours...Iris avec elle-même.
- On retient les intersections qui existent entre deux IRIS différents.

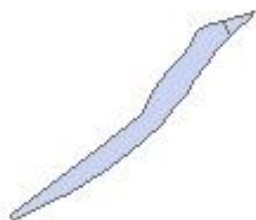


Figure 13 : Extraction des intersections entre deux IRIS

- On sélectionne le premier IRIS participant à l'intersection.

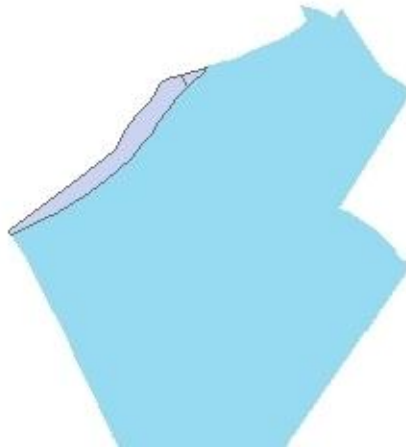


Figure 14 : Association du morceau d'intersection avec le premier IRIS rencontré par l'algorithme

- On agrège géographiquement les morceaux d'IRIS ensemble pour obtenir un polygone unique décrivant le contour de l'IRIS.

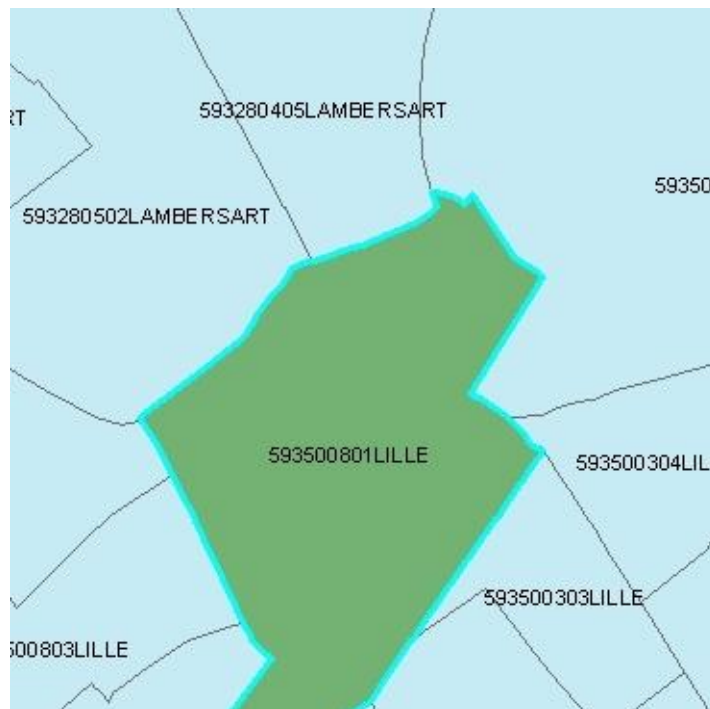


Figure 15 : Agrégation des morceaux d'IRIS pour obtenir le contour final de l'IRIS

3.2.3 Correction des contours communaux des IRIS

Lors du croisement des contours IRIS avec des données spatiales fines du type BD TOPO, la résolution de la BD Contours...Iris crée un biais sur la sélection des bâtiments inclus dans l'IRIS. Certains bâtiments ne sont pas nécessairement placés dans la commune attendue (cf. Figure 7). La solution pour affiner la définition des contours IRIS est de remplacer les contours communaux des IRIS par les contours communaux de la BD TOPO. Deux cas sont traités séparément :

- La commune ne comporte qu'un seul IRIS (mono-IRIS). Le contour de l'IRIS est remplacé directement par le contour BD TOPO.
- La commune comporte plusieurs IRIS (multi-IRIS). Les contours extérieurs communaux des IRIS sont corrigés par les contours communaux de la BD TOPO.

La Figure 18 présente les principales étapes de la correction. A l'étape 6, une zone tampon (1 m) est définie autour de chacun des IRIS pour localiser ses IRIS voisins. Tous les IRIS qui croisent cette zone tampon sont retenus.

Lors de l'intersection, certains morceaux d'IRIS (A) de l'étape 6 sont imbriqués au milieu d'autres morceaux d'IRIS (B) à corriger. Tant que les morceaux B ne sont pas corrigés, les morceaux A ne peuvent l'être. Il est donc nécessaire de dérouler les étapes 2 à 8 deux fois de suite pour corriger dans un premier temps la majorité des IRIS (A) pour ensuite corriger les quelques morceaux restants (B).

Après que les contours ont été corrigés et agrégés (étape 8) en fonction du code IRIS, il apparaît des « îlots » d'IRIS ; ceux-ci sont des morceaux d'IRIS détachés de l'IRIS principal. Le code IRIS de ces morceaux est corrigé en recherchant l'IRIS voisin ayant le même code communal et la plus grande frontière commune (étapes 9 à 11). La Figure 17 présente un exemple d'îlot (entouré en orange) séparé de la partie principale de l'IRIS colorée en rose. Cet îlot est rattaché à l'îlot voisin (en bleu) appartenant à la même commune.

Cette méthodologie corrige pour partie les erreurs situées aux frontières communales mais conserve l'incertitude sur les frontières entre IRIS à l'intérieur des communes multi-IRIS.

La Figure 16 présente un exemple de rendu de la correction des contours IRIS.

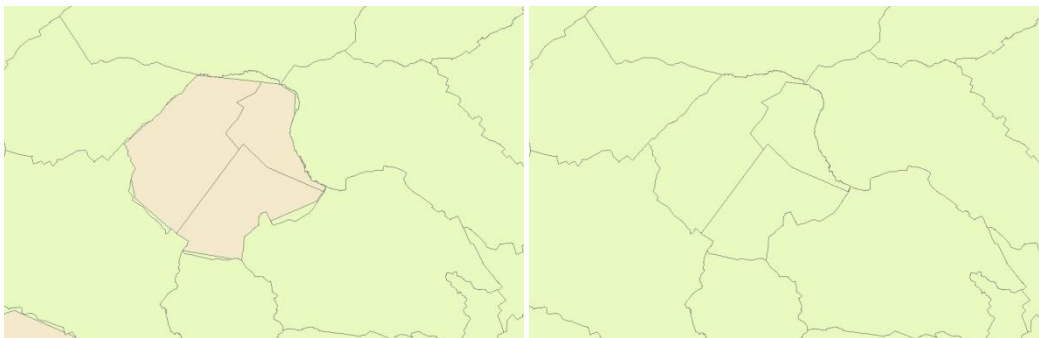


Figure 16 : A gauche, en rose, contour IRIS d'une commune avant correction ; à droite contours corrigés avec les contours communaux de la BD TOPO

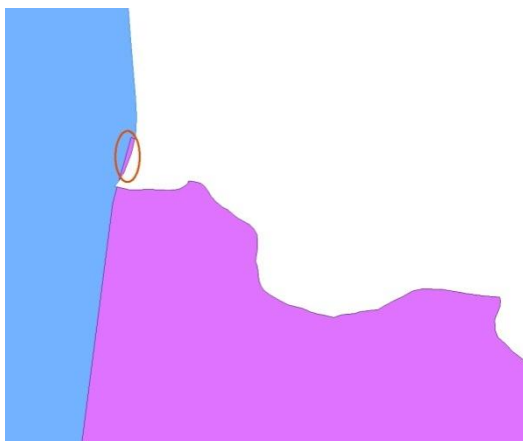


Figure 17 : Un îlot (entouré en orange) est séparé de l'IRIS principal coloré en rose.

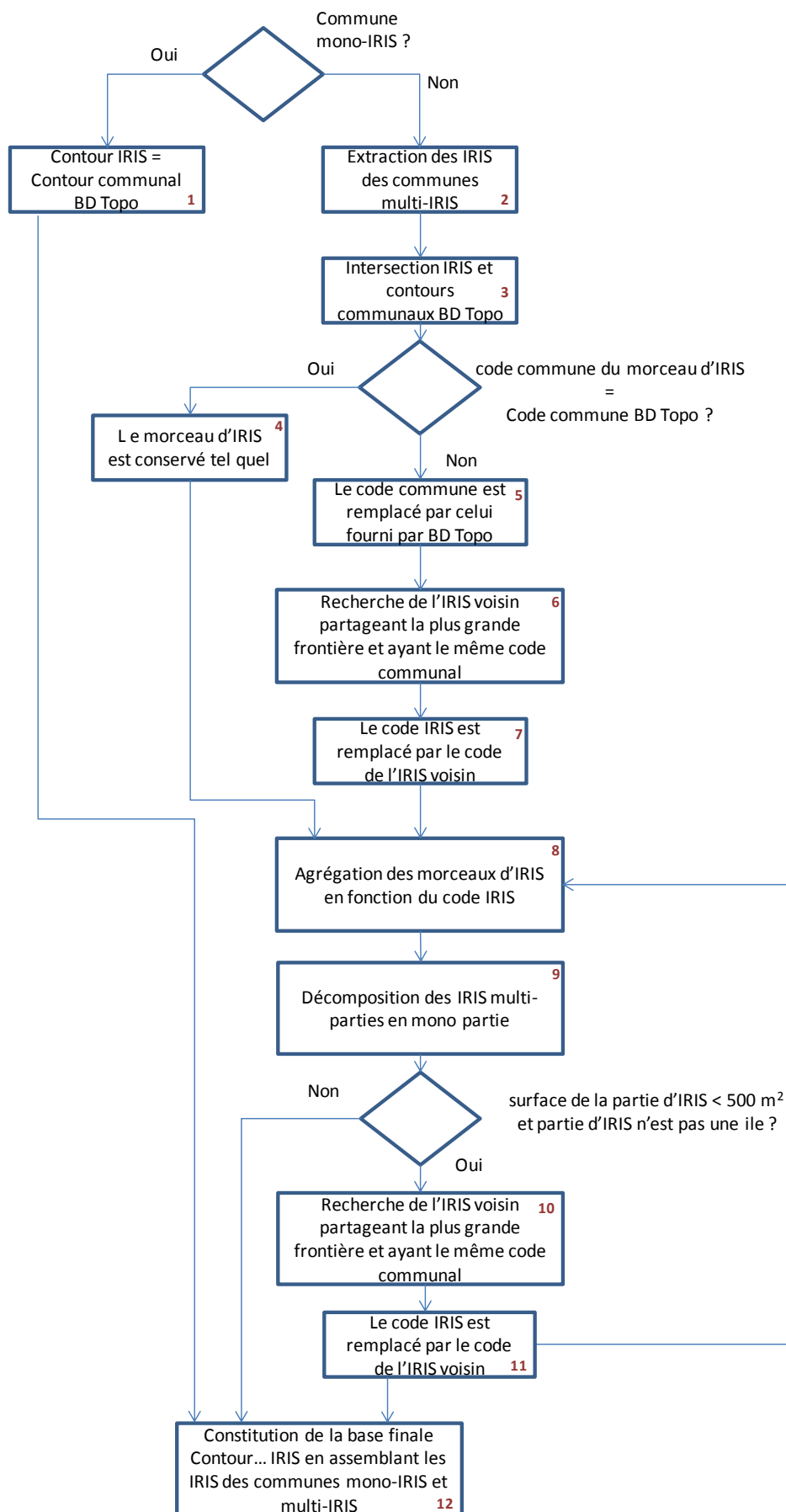


Figure 18 : Algorithme de correction des contours IRIS par les contours communaux de la BD TOPO

3.3 Méthodologie MAJIC

La précédente version de la méthodologie proposait uniquement une spatialisation des populations sur les parcelles dont le contour géographique était connu. Le présent document apporte des améliorations à cette méthodologie pour la rendre applicable sur l'ensemble du territoire. Cette partie du document présente les données à disposition dans la base MAJIC du CEREMA, un rappel sur la méthodologie proposée en 2013, les évolutions de la méthodologie réalisées sur l'année 2014 pour résoudre certains écarts constatés et la méthodologie finale appliquée sur l'année de référence INSEE 2011.

3.3.1 Les données MAJIC

Les fichiers MAJIC regroupent de nombreuses informations associées à tous les locaux et parcelles cadastrales du territoire français:

- Informations sur les propriétaires
- Usage et caractéristiques du local : habitation ou commercial, surfaces, nombre de pièces, année de construction, numéro de parcelle cadastrale à laquelle il appartient...
- Informations sur les copropriétés
- Information sur les parcelles : commune, adresse de la parcelle si elle existe.

Les services des impôts utilisent ces informations pour définir les taxes foncières.

Le terme « local », utilisé dans les fichiers MAJIC, désigne la partie habitation d'une maison ou d'un appartement (l'ensemble des pièces qui les constituent) ou la partie professionnelle. Dans ce document, le terme « local » MAJIC désignera un logement (maison ou appartement) à caractère d'habitation. Une même maison peut contenir plusieurs locaux MAJIC si celle-ci est destinée à la location et est composée de logements distincts.

Le CEREMA (Direction territoriale Nord-Picardie, anciennement CETE Nord-Picardie) met à disposition les données foncières sous la forme de bases de données spatiales (PostgreSQL/PostGIS). Une base est délivrée par département. Dans la suite du document, cette base de données sera appelée BD MAJIC CETE.

Les données qu'elle contient sont distribuées dans 18 tables. Seules deux de ces tables sont utilisées pour la répartition spatiale de la population :

- Une première table fournit des informations sur les parcelles. Les champs suivants sont exploités dans la méthodologie :
 - Le champ idpar fournit l'identifiant unique de la parcelle.
 - Le champ idpropte donne l'identifiant du propriétaire.
 - Le champ idparef donne l'identifiant de la parcelle de référence.
 - Le champ nbat indique le nombre de locaux référencés sur la parcelle.
 - Le champ cmp indique si la parcelle fait partie d'une copropriété multi-parcelles (valeur 1).
 - Le champ dcntpa indique la surface de la parcelle.
 - Le champ vecteur indique si la parcelle est vectorisée ou non (valeur 'V', 'I' ou 'N').

- Le champ `source_geo` indique la source : BD PARCELLAIRE + année ou Cadastre.gouv.fr + année.
- Le champ `geompar` contient la géométrie du contour de la parcelle s'il existe.
- Le champ `geomloc` contient la géométrie du centre de la parcelle s'il existe.
- Une seconde table fournit des informations sur les locaux. Les champs suivants sont exploités dans la méthodologie :
 - Le champ `idpar` fournit l'identifiant de la parcelle sur laquelle se trouve le local.
 - Le champ `dteloc` fournit le type du local : si '1' local de type habitation maison – si '2' local de type habitation appartement.
 - Le champ `stoth` indique la surface totale des pièces d'habitation.
 - Le champ `idcom` donne le code INSEE de la commune accueillant le local.
 - Le champ `idcomtxt` donne le nom de la commune accueillant le local.
 - Le champ `dnvoiri` fournit le numéro du bâtiment dans la rue accueillant le local.
 - Le champ `dindic` présente l'indice de répétition du numéro du bâtiment (bis, ter, ...).
 - Le champ `dvoilib` donne le nom de la rue accueillant le local.
 - Le champ `dniv` fournit le numéro d'étage du local.

Avant l'exploitation de la BD MAJIC CETE, il est utile d'analyser la constitution de cette base et les cas particuliers qui peuvent être rencontrés :

- Un local est attribué uniquement à une seule parcelle,
- mais une même parcelle peut contenir plusieurs locaux.
- Certaines copropriétés sont réparties sur plusieurs parcelles. Celles-ci sont appelées copropriétés multi-parcelles et sont repérées par le champ '`cmp`' qui est fixé à 1. Ces copropriétés ont la particularité de posséder une parcelle de référence unique (champ `idparef`) sur laquelle sont référencés tous les locaux attachés à cette copropriété. La Figure 19 présente un exemple de copropriétés multi-parcelles sur la ville d'Amiens. Les 101 locaux de type habitation de cette zone sont référencés sur la parcelle de référence numéroté situé en bas. Cette particularité sera prise en compte dans la méthodologie en répartissant les locaux référencés dans la parcelle de référence sur l'ensemble des parcelles constituant la copropriété.



Figure 19 : Exemple d'une copropriété multi-parcelles

- Un cas similaire apparaît sur certains terrains où une parcelle a subi une division cadastrale. Ces parcelles appartiennent au même propriétaire dont l'identifiant est repéré dans le champ 'idprocte'. Le logement n'est alors pas nécessairement associé à la parcelle accueillant physiquement le bâtiment d'habitation. La recherche des bâtiments existants sur les parcelles voisines du même propriétaire est nécessaire.
- Lors de la création de nouvelles habitations sur une parcelle, les données foncières sont nécessairement mises à jour par la DGFIP mais l'information spatiale associée (contour du bâtiment dans la BD PARCELLAIRE) n'est pas mise à jour dans le même temps. Un bâtiment ne sera donc pas trouvé nécessairement sur chaque parcelle d'habitation. Ce cas est traité dans la méthodologie.

3.3.2 Rappel méthodologique

3.3.2.1 La méthodologie MAJIC ancienne version

La méthodologie MAJIC s'appuie principalement sur les données MAJIC pour répartir les populations municipales de l'INSEE sur les bâtiments d'habitation. Cette méthodologie se décompose en plusieurs étapes présentées de manière détaillée dans le précédent guide méthodologique.

Pour résumé, la méthodologie se décomposait de la façon suivante :

- On répartit les locaux de type habitation (maison ou appartement) recensés dans la base MAJIC sur chacun des bâtiments des parcelles cadastrales qui contiennent ces locaux pour établir un nombre de locaux par bâtiment.

- Ces données sont ensuite croisées avec les données de recensement de l'INSEE pour déterminer un nombre moyen d'habitants par type de local MAJIC.
- Un nombre d'habitants est enfin calculé sur chacun des bâtiments.

3.3.2.2 Calcul de la surface habitable d'un bâtiment

La couche « bâti indifférencié » de la BD TOPO comporte un champ numérique 'Hauteur' caractérisant la hauteur maximale de chaque bâtiment. Cette hauteur est utilisée pour établir le nombre de niveaux d'un bâtiment pour ensuite estimer une surface habitable qui sera le produit du nombre de niveaux par la surface de l'emprise au sol du bâtiment.

3.3.2.3 Détermination du nombre de niveaux d'un bâtiment

L'application d'une hauteur « type » par niveau est délicate, cette hauteur est dépendante de nombreux paramètres (région, zone rurale ou urbaine, âge de l'habitat,..) difficiles à caractériser.

A défaut d'informations précises, les hauteurs suivantes seront utilisées :

- Un bâtiment d'une hauteur inférieure à 5 mètres ne possède qu'un niveau.
- La hauteur des niveaux supérieurs se situe entre 2.5 et 3 m. Le LCSQA a fixé une hauteur de 2,7 m par niveau pour la méthodologie MAJIC.

Du fait d'une information incomplète dans la BD TOPO, certains bâtiments possèdent une hauteur nulle, en général moins de 1 % des bâtiments. La hauteur des bâtiments de faible surface est fixée à un étage ; pour ceux présentant une surface plus importante, une observation humaine est conseillée.

3.3.3 Les limites de l'ancienne méthodologie

La méthodologie MAJIC (2013) a été appliquée pour essai sur plusieurs départements français. Des limites dans les données de population élaborées ont été relevées et ont restreint ainsi l'usage potentiel de cette méthodologie.

Une couverture géographique imparfaite de la BD PARCELLAIRE

Du fait d'une couverture partielle de la BD PARCELLAIRE Vecteur (cf 2.2.1.2), les travaux menés en 2013 exploitaient uniquement les parcelles dont le contour géographique était connu. La méthodologie n'était donc pas applicable sur l'ensemble du territoire.

Des bâtiments Bd Parcellaire inexistants

La couche 'bâtiment' de la BD PARCELLAIRE recense les bâtiments situés sur les parcelles vectorisées. L'utilisation de cette couche fait apparaître deux problèmes : la hauteur des bâtiments n'est pas connue et certains bâtiments récents n'y sont pas répertoriés.

La hauteur des bâtiments est utilisée pour évaluer la surface habitable d'un bâtiment qui est le produit entre la surface au sol et le nombre de niveaux de la construction. La hauteur est établie en croisant spatialement cette couche avec celle de la BD TOPO. Il a été constaté que les couches BD TOPO et BD PARCELLAIRE ne se superposaient pas parfaitement : un bâtiment BD PARCELLAIRE peut croiser deux bâtiments BD TOPO et il est alors impossible à l'algorithme de calcul de déterminer la hauteur correcte à appliquer.

De plus, la BD TOPO est mise à jour plus rapidement que la BD PARCELLAIRE et paraît recenser plus fidèlement les nombreuses zones nouvellement construites. L'utilisation exclusive des bâtiments de la BD PARCELLAIRE empêche sur ces zones une bonne prise en compte du bâti.

Une densité d'habitants trop élevée dans certains bâtiments

Comme expliqué dans le paragraphe 3.3.1, certains locaux sont en réalité situés sur plusieurs parcelles mais sont référencés dans la base MAJIC sur une seule parcelle dite parcelle de référence. Dans le meilleur des cas, la base MAJIC référence le lien entre cette parcelle de référence et les autres parcelles dites filles. Mais il a été constaté que ce lien n'est pas toujours établi dans cette base : l'application de la méthodologie sur certaines parcelles crée une surpopulation artificielle sur certains bâtiments.

Une incohérence entre les statistiques des locaux MAJIC et les données de population

La méthodologie applique un nombre moyen d'habitants par IRIS et par type de local (maison ou appartement). Cette valeur moyenne est établie par le croisement du nombre de locaux MAJIC agrégés à l'IRIS et les données de statistiques de population INSEE. Il a été relevé sur certains IRIS que le nombre moyen d'habitants calculé par local MAJIC peut être très important (supérieur à 10 habitants). Ces valeurs non réalistes s'expliquent soit par une incertitude trop grande du recensement de la population pour ces IRIS, soit par une délimitation géographique de l'IRIS ne correspondant pas à la zone géographique réelle utilisée par l'INSEE ou soit, peut-être, par un manque d'information dans les données MAJIC. Un exemple est donné sur la commune de Saint-Memmie située dans la Marne. L'INSEE a découpée celle-ci en 2 IRIS (515060101 et 515060102).

Les statistiques de population et de locaux MAJIC associées à ces IRIS sont présentées dans le Tableau 1. Les résultats montrent une incohérence entre les données MAJIC et les données de l'INSEE : le nombre d'habitants par local (1180) dans le deuxième IRIS est bien trop important et à l'inverse, pour le premier IRIS, la valeur paraît trop faible (1.14). Par contre, la même statistique calculée sur l'ensemble de la commune paraît convenable.

Tableau 1 : Comparaison des locaux MAJIC avec les données de population INSEE sur les IRIS de Saint-Memmie (51)

En superposant les contours de ces IRIS à une photographie aérienne (Figure 20), la délimitation telle qu'elle est proposée par la base Contour...Iris montre que le deuxième IRIS couvre une zone agricole comprenant seulement deux exploitations agricoles. Pour ce cas de figure, les données MAJIC sont cohérentes avec les contours des IRIS (2 locaux recensés pour le deuxième IRIS). L'écart se trouve donc sur la définition du périmètre géographique de ces deux IRIS qui est probablement faux dans la base Contour...Iris.

Figure 20 : Contours géographiques des IRIS (515060101 et 515060102) de la commune de Saint-Memmie

Les limites relevées dans la partie 3.3.3 montrent que la méthodologie MAJIC élaborée en 2013 n'est pas applicable en l'état. Le LCSQA s'est attaché à améliorer la méthodologie pour la généraliser sur l'ensemble du territoire français. Des solutions ont été apportées pour lever les différents blocages ; il reste néanmoins ponctuellement quelques écarts qui nécessitent une expertise humaine pour valider les données obtenues.

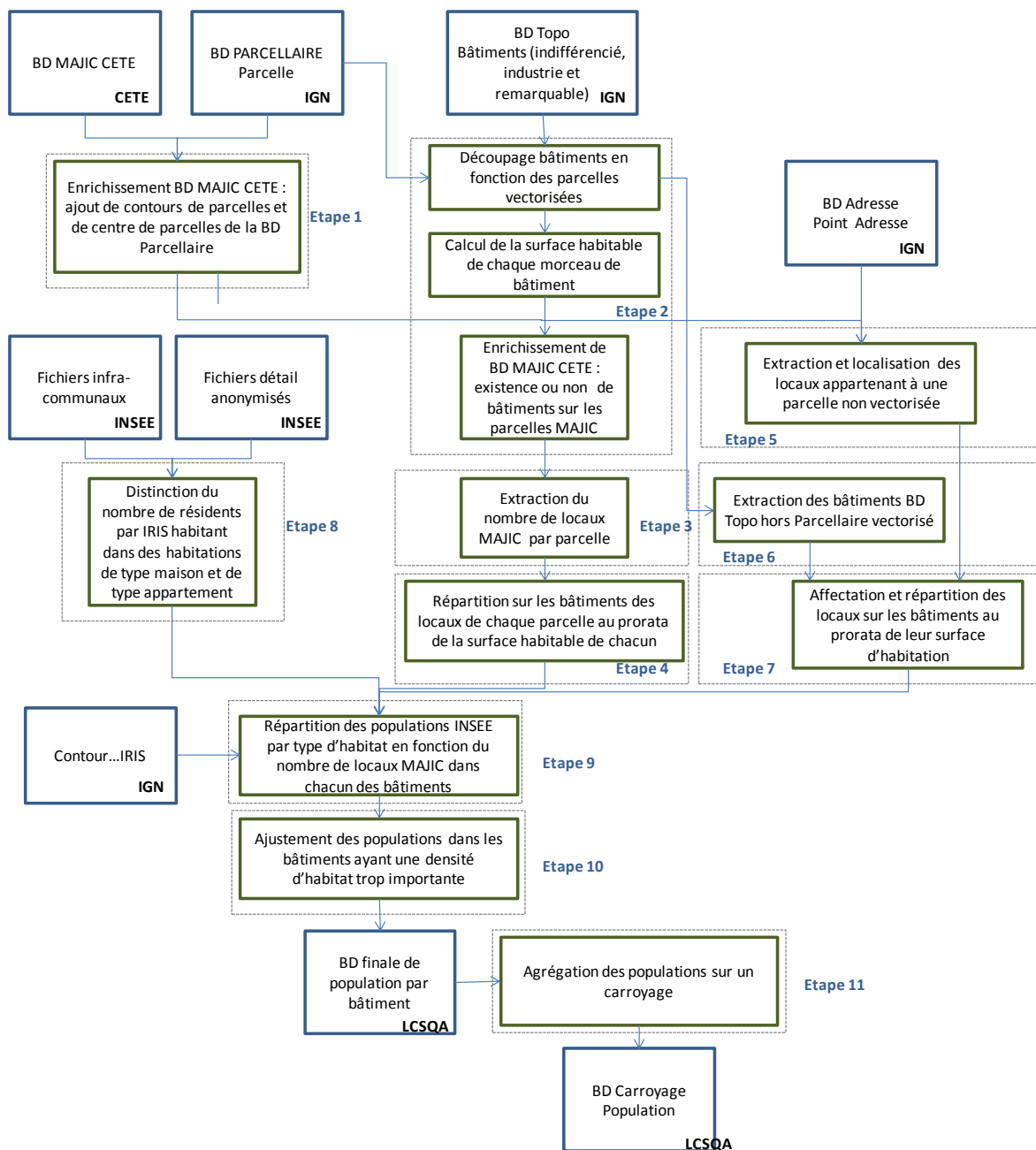


Figure 21 : Présentation des différentes étapes de la méthodologie MAJIC 2014

- **Etape 1** : Lors de la constitution de la BD MAJIC CETE, le CETE a exploité les fichiers fonciers DGFIP, la BD PARCELLAIRE disponible en 2011 et d'autres sources de données. Lors de l'application de la méthodologie MAJIC de spatialisation de la population, il est fort probable que la BD PARCELLAIRE disponible soit plus récente que celle employée par le CETE et qu'elle possède donc des parcelles vectorisées ou centres de parcelles complémentaires. Afin d'enrichir la BD MAJIC CETE, il est utile de confronter les deux bases de parcelles (BD MAJIC CETE et BD PARCELLAIRE la plus récente). Pour cela, on extrait les identifiants des parcelles (BD MAJIC CETE) qui ne sont pas vectorisées (champ 'vecteur' <> 'V') pour ensuite vérifier si ces parcelles sont vectorisées dans la BD PARCELLAIRE actuelle. Si certaines sont vectorisées, les géométries de celles-ci sont intégrées dans la BD MAJIC CETE. De même, pour les parcelles dont les coordonnées du centre ne seraient pas connues (champ 'vecteur' = 'N'), celles-ci sont recherchées dans la BD PARCELLAIRE puis introduites dans la BD MAJIC CETE si elles sont trouvées.
- **Etape 2** : Cette étape concerne uniquement les parcelles dont le contour géographique est défini. Nous avons vu dans les cas particuliers du 3.3.1 qu'il serait judicieux de recenser les parcelles possédant des bâtiments.
Les bases BD PARCELLAIRE et BD TOPO sont croisées pour caractériser les bâtiments et déterminer les parcelles contenant ces bâtiments :
 - Les bâtiments de la BD TOPO ('indifférencié', 'industriel' et 'remarquable') sont découpés (intersection) selon le contour géographique des parcelles qui les croisent. Un bâtiment BD TOPO peut se retrouver ainsi découper en plusieurs morceaux si celui-ci traverse plusieurs parcelles. Un identifiant de parcelle est ajouté à chacun des bâtiments.
 - Le nombre d'étages et la surface habitable sont calculés pour chacun des bâtiments.
 - Les bâtiments sont ensuite croisés avec la couche 'Parcelles' de la BD MAJIC CETE pour lui insérer des informations complémentaires : existence ou non de bâtiments sur la parcelle, surface habitable (cf. 3.3.2.2) cumulée des bâtiments sur la parcelle.
- **Etape 3** : Cette étape concerne uniquement les parcelles dont le contour géographique est connu. Les deux précédentes étapes ont enrichi la BD MAJIC CETE de nouvelles informations. Cette base est exploitée pour positionner les locaux à caractère d'habitation sur chacune des parcelles. La Figure 22 présente l'algorithme qui affecte les locaux sur les parcelles. Les cas particuliers présentés en 3.3.1 sont traités dans l'algorithme. A l'issue des traitements 1, 2 et 4, les locaux sont affectés sur des bâtiments. Les traitements 3 et 5 ne permettent pas de trouver de bâtiment dans lequel placer les locaux, un bâtiment fictif devra être dans ce cas. Les traitements 2 et 4 utilisent les informations récoltées à l'étape 2 pour connaître la surface habitable des bâtiments dans chaque parcelle. Lors de cette étape, une distinction est faite sur le type des locaux habités : les habitations de type maison et de type appartement sont comptabilisés séparément. Cette distinction est conservée dans les étapes suivantes.

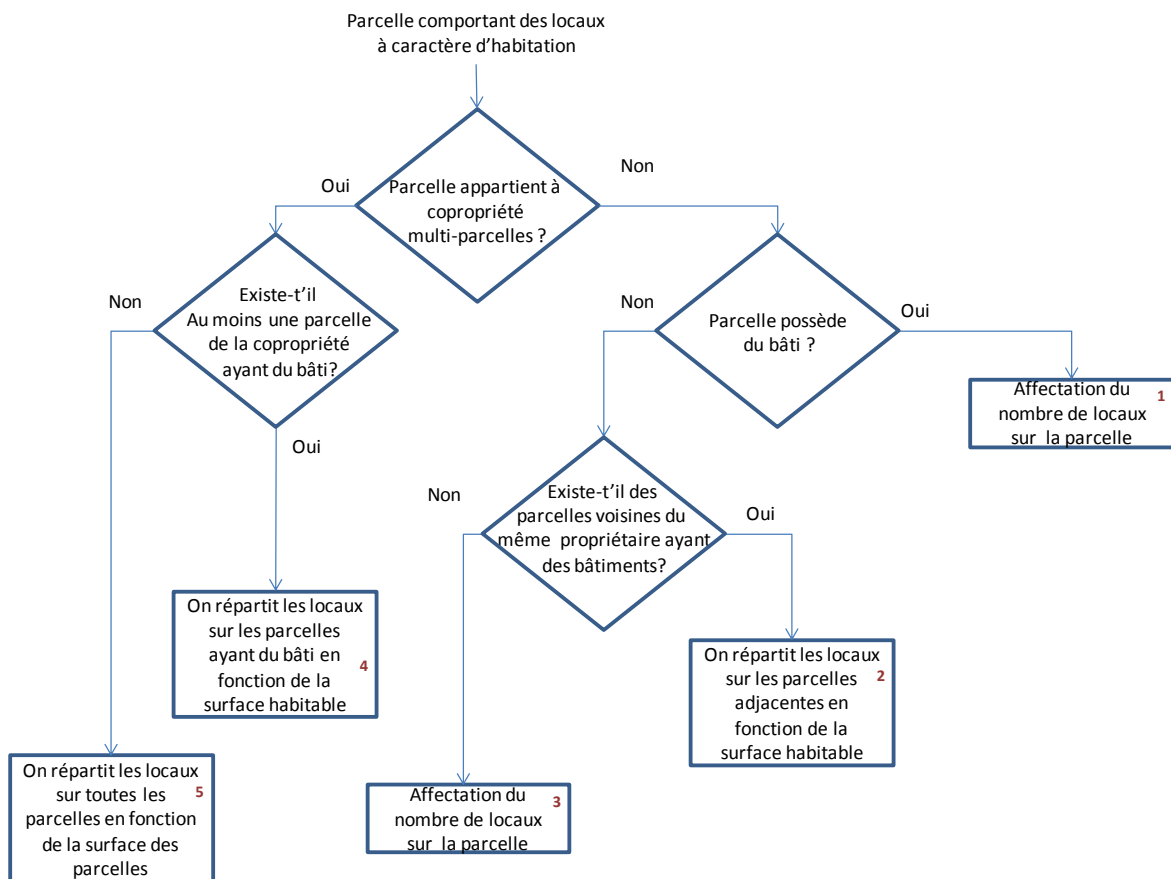


Figure 22 : Algorithme d'affectation des locaux dans les parcelles.

- **Etape 4 :** La présente étape concerne uniquement les parcelles dont le contour géographique est connu. Suite à l'étape précédente, les locaux sont ventilés sur les bâtiments des parcelles retenues. Les locaux ne sont répartis sur des bâtiments de type industriel ou commercial (BD TOPO) que dans l'une des conditions suivantes :
 - Le nombre de locaux est inférieur ou égal à 5 ;
 - Le nombre de locaux est supérieur à 5 et aucun autre bâtiment de type 'indifférencié' n'est recensé pour cette parcelle.

Si une parcelle ne contient pas de bâtiment mais possède des locaux MAJIC, un bâtiment fictif est créé. Celui-ci reprend les caractéristiques des locaux situés sur la parcelle : surface habitable et nombre de niveaux. Lorsque qu'une parcelle contient plusieurs bâtiments, les locaux sont répartis sur les différents bâtiments au prorata de leur surface habitable. La distinction par type de local est conservée lors de la répartition des locaux. Si une parcelle comporte des habitations de type maison et de type appartement, chacun des bâtiments conservera une part de ces deux types de locaux. Il est en effet difficile de déterminer sur une même parcelle quels bâtiments seraient susceptibles d'être des maisons ou d'accueillir des appartements. Le résultat final de cette étape est une base de données spatiale de bâtiments contenant un nombre de locaux. Nous appellerons cette base BD Locaux.

- **Etape 5 :** Les étapes 5 ,6 et 7 concernent uniquement les parcelles dont le contour géographique n'est pas connu. Pour ces parcelles, soit dans le meilleur des cas les coordonnées du centre de la parcelle sont connues, soit l'adresse des locaux est connue.

Les coordonnées du centre de la parcelle sont connues

La table des parcelles contenue dans la BD MAJIC comporte un champ géographique (geomloc) qui contient la géométrie du centre de chaque parcelle lorsqu'il est connu. Le contenu de ce champ est utilisé pour localiser les locaux situés sur la parcelle.

Les coordonnées du centre la parcelle ne sont pas connues

La table des locaux contient plusieurs champs définissant l'adresse de chaque local : idcom, idcomtxt, dnvoiri, dindic, dvoilib (cf 3.3.1). L'application d'un outil de géocodage sur cette adresse permet de situer le local géographiquement. Certaines adresses ne seront probablement pas reconnues par le géocodeur car elles ne seront pas référencées dans la base d'adresse ou seront incorrectement traduites pour le géocodeur (cf 2.3.4).

- **Etape 6 :** Une fois les centres des parcelles positionnés lors de l'étape 5, il est nécessaire de repérer les bâtiments à proximité de ces parcelles pour leur affecter des locaux d'habitation. Ces bâtiments doivent nécessairement se trouver sur des zones non couvertes par la BD PARCELLAIRE vecteur. La présente étape retient uniquement les bâtiments de la BD TOPO ne se trouvant pas sur une parcelle vectorisée et calcule la surface habitable de chacun de ceux-ci.
- **Etape 7 :** L'étape 5 a permis de placer des locaux sur un localisant ponctuel. Ces locaux doivent ensuite être distribués sur les bâtiments recensés à l'étape 6. Deux informations sont exploitées : la surface d'habitation cumulée des locaux sur le localisant et la surface habitable des bâtiments situé à proximité. La recherche des bâtiments autour du localisant est restreinte à l'intérieur d'un rayon. Différents essais ont permis de définir un rayon maximal de recherche égal à la racine carré de la surface de la parcelle. Le champ dcntpa de la table MAJIC parcelle fournit la surface déclarée de la parcelle.

La Figure 23 présente un exemple de recherche de bâtiment sur une zone située dans le département de l'Ain. L'algorithme de recherche des habitations et d'affectation des locaux est présenté par la Figure 24.

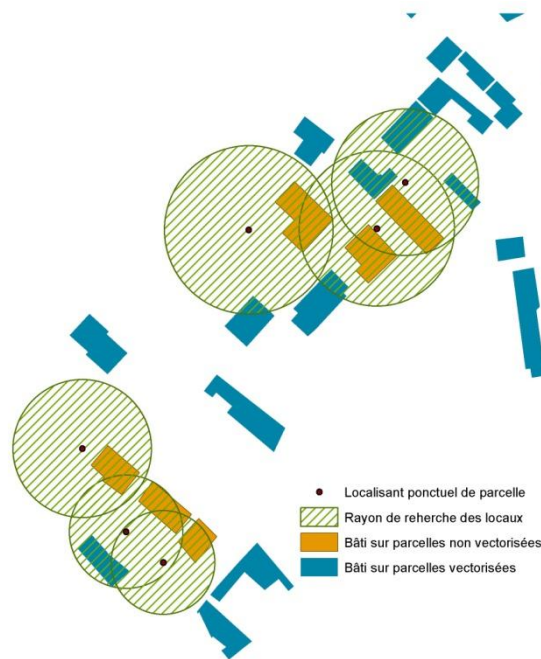


Figure 23 : Recherche de bâtiments d'habitation à proximité de localisants ponctuels.

Cet algorithme se décompose en plusieurs phases :

- Une première phase (1) extrait les bâtiments potentiels : hors parcelle vectorisée et à proximité des localisants ponctuels à une distance inférieure au carré de la surface de la parcelle. Cette étape détermine la distance pour chaque couple localisant/bâtiment et trie ces couples en fonction de cette distance.
- Une seconde phase (2) répertorie les bâtiments dans lesquels appliquer les locaux d'un localisant et répartit ensuite les locaux dans ceux-ci. Cette phase itère sur chaque localisant et se décompose en deux étapes :
 - L'algorithme (2-1) itère sur chacun des bâtiments associé à un même localisant (étape 1) et retient uniquement les bâtiments utiles à la répartition des locaux : l'algorithme s'arrête lorsqu'il a trouvé suffisamment de bâtiments pour accueillir la surface habitable de la parcelle. Il est à noter qu'un même bâtiment peut accueillir les locaux de plusieurs localisants si sa surface habitable le permet.
 - Les locaux d'un même localisant sont ensuite répartis sur les bâtiments retenus au prorata de leur surface d'habitation. Si aucun bâtiment n'est trouvé, un bâtiment fictif est créé avec une emprise au sol égale à la surface d'habitation des locaux divisée par le nombre de niveaux recensé dans la base MAJIC CETE.

Les bâtiments associés à des locaux d'habitation sont insérés dans la base BD Locaux de l'étape 4.

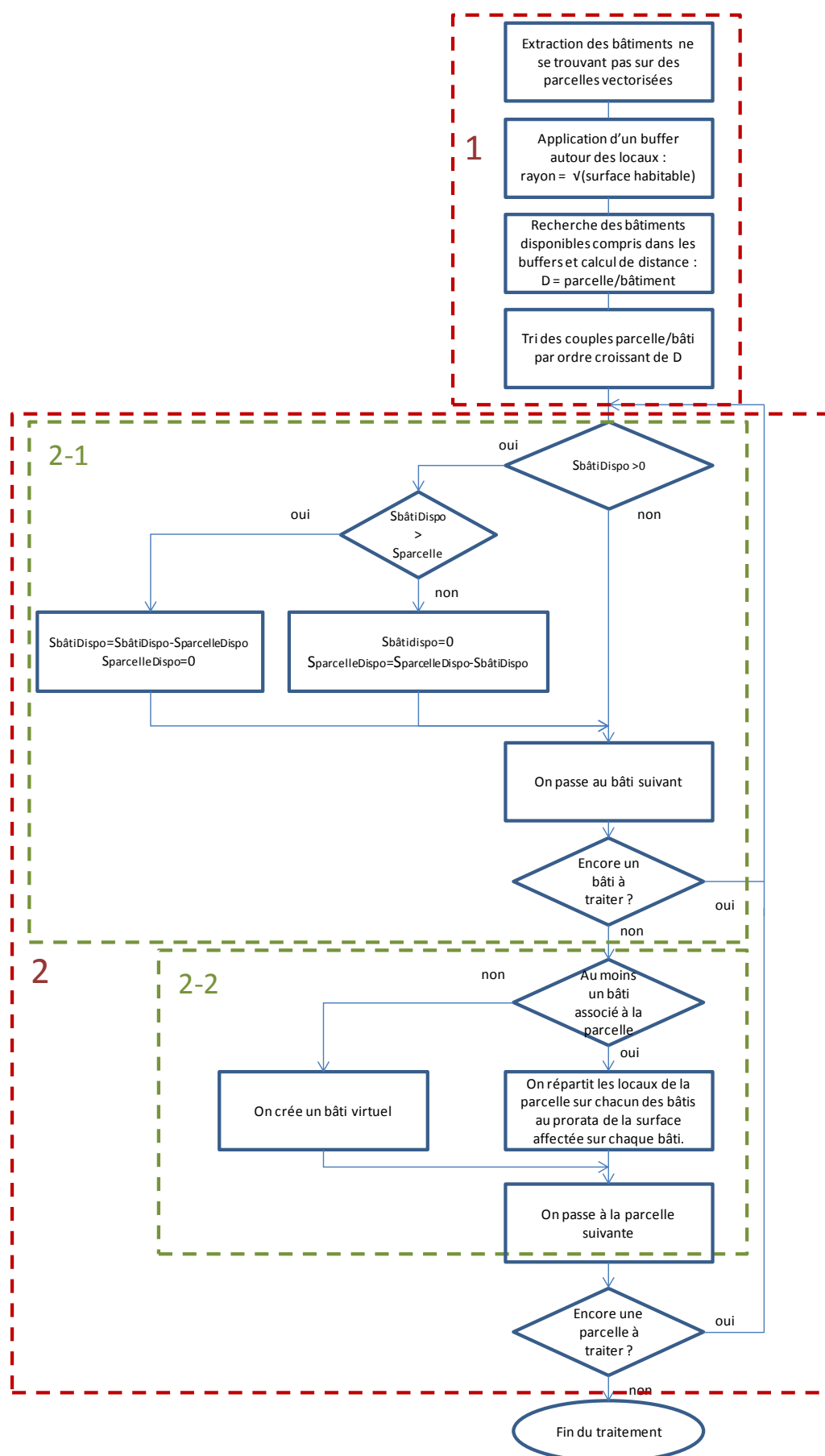


Figure 24 : Algorithme d'affectation des locaux sur des bâtiments uniquement pour les parcelles dont on ne connaît pas le contour.

- **Etape 8** : Avant de répartir les populations INSEE dans les bâtiments, il est nécessaire de comptabiliser par IRIS le nombre d'habitants résidant dans des habitations de type maison ou de type appartement. L'utilisation croisée des fichiers infra-communaux et des fichiers détail anonymisés de l'INSEE permet d'obtenir ces statistiques :
 - Les fichiers détails INSEE délivrent des statistiques par IRIS et par classe de logements. Ces classes de logements sont caractérisées par de nombreuses variables : nombre de pièces, surface du logement, type du logement, ... Ces fichiers comptabilisent uniquement les populations des ménages. Les personnes vivant dans des habitations mobiles, les marins, les sans-abris et les personnes vivant en communauté n'y sont pas recensées. Le cumul des populations par IRIS et par type de logement (maison ou appartement) permet d'établir le nombre d'habitants associés à chaque type de logement par IRIS.
 - La population hors-ménage de l'INSEE est comptabilisée dans la population municipale (cf. 2.1). Afin de tenir compte de cette catégorie de population, cette population est extraite des fichiers infra-communaux. Ces habitants seront localisés sur un habitat de type appartement. Ces populations sont cumulées aux populations résidant en appartement (fichiers détails).
- **Etape 9** : Cette étape distribue les populations (par type d'habitat de l'étape 8) recensées par IRIS dans les bâtiments d'habitation.

Pour cela, la couche BD Locaux est intersectée avec la couche Contours...IRIS. Cette intersection a pour effet dans certains cas de découper en deux parties certains bâtiments situés sur la frontière de deux IRIS. Les locaux situés dans ces bâtiments sont répartis sur chaque morceau au prorata de leur surface respective. Le nombre de locaux MAJIC contenus dans les bâtiments est cumulé par IRIS pour calculer le nombre total de locaux MAJIC par IRIS.

Ce nombre des locaux par IRIS est associé au nombre d'habitants de l'IRIS (établi lors de l'étape 8). Le paragraphe 3.3.3 montre qu'une incohérence importante entre les locaux d'habitation recensés dans la base MAJIC et ceux de l'INSEE apparaît sur certains IRIS : ceux-ci reçoivent une densité d'habitants trop importante (> à 5 habitants par local). Si l'algorithme de calcul détecte une densité d'habitants moyenne trop importante, il tente d'obtenir des statistiques plus adaptées en utilisant des données de population INSEE plus agrégées : soit les statistiques par type de local sur un regroupement d'IRIS (TRIRIS au sens INSEE), soit les statistiques par type de local sur la commune qui contient les IRIS ou à défaut les statistiques générales de population sur la commune.

Finalement, un nombre moyen d'habitants par type de local est établi dans chacun des IRIS. Ce nombre moyen est appliqué sur chacun des bâtiments de l'IRIS pour estimer par bâtiment le nombre de personnes résidentes. L'Équation 1 résume ce calcul :

$$P_{b,i,t} = P_{i,t} * \frac{L_{b,i,t}}{\sum_a L_{a,i,t}} \quad \text{Équation 1}$$

$P_{b,i,t}$: Nombre d'habitants situés dans le bâtiment b de l'IRIS i pour le type de logement t ,

$P_{i,t}$: Nombre d'habitants recensés dans l'IRIS i résidant dans le type de logement t ,

$L_{b,i,t}$: Nombre de locaux MAJIC situés dans le bâtiment b de l'IRIS i pour le type de logement t .

$L_{a,i,t}$: Nombre de locaux MAJIC situés dans un bâtiment a de l'IRIS i pour le type de logement t .

- **Etape 10** : L'application des populations sur les bâtiments d'habitation fait ressortir pour certains bâtiments une densité d'habitants trop importante (valeur retenue : au-delà de 1 habitant pour 10 m² d'habitation). Comme présenté dans la partie 3.3.3, certains locaux sont en réalité situés sur plusieurs parcelles mais sont référencés dans la base MAJIC sur une parcelle unique. Dans ce cas, les habitants associés sont placés dans cette parcelle au lieu d'être distribués sur l'ensemble des parcelles. Ceci provoque une surpopulation artificielle dans ces bâtiments. L'étape 10 corrige ces écarts. L'algorithme tente de distribuer la population en surplus dans d'autres bâtiments. Ces bâtiments doivent être proches (dans un rayon de 100 m) et susceptibles d'accueillir une population complémentaire (densité d'habitants inférieure à 0.1 hab/m²). Si des bâtiments sont trouvés, la population est répartie sur l'ensemble des bâtiments au prorata des surfaces habitables de chacun. Sinon, la population du bâtiment est laissée telle quelle.
- **Etape 11** : Une dernière étape est ajoutée pour répartir les populations dans un carroyage régulier. Pour des besoins opérationnels, il est parfois utile d'estimer les populations selon un maillage régulier dans le but, par exemple, de croiser directement des données de concentrations de polluants avec des données de population. Cette étape consiste simplement à réaffecter les populations des bâtiments dans des carreaux. Deux cas sont rencontrés :
 - Le bâtiment est totalement inclus dans la maille. La totalité de ses habitants est comptabilisée dans celle-ci.
 - Le bâtiment se trouve à la frontière de 2, 3 ou 4 mailles. Sa population est répartie dans chacune des mailles proportionnellement à la surface de chevauchement sur chacune d'elles.

3.4 Application de la méthodologie et validation des données

La méthodologie a été appliquée et automatisée en décembre 2014 sur l'ensemble des départements métropolitains et sur quatre départements d'outre-mer : la Guadeloupe, la Guyane, la Martinique et La Réunion.

3.4.1 Logiciels et outils utilisés

Les travaux ont été menés sous l'environnement Windows en utilisant les applications ArcMap, PostgreSQL et sa composante spatiale PostGIS. Le langage de programmation python a été utilisé pour automatiser les traitements et faciliter les échanges de données entre les différentes applications. Outre l'automatisation, l'utilisation de ce langage a permis de paralléliser les traitements et traiter ainsi plus rapidement les données.

Le LCSQA a utilisé deux outils pour géocoder les adresses des parcelles :

- Le premier outil est intégré directement dans ArcMap. ESRI France, fournisseur de ce logiciel, a élaboré un guide et paramétré cet outil pour faciliter son utilisation avec la BD Adresse de l'IGN. Le tutoriel datant de quelques années, il a été nécessaire d'adapter celui-ci pour utiliser le géocodeur avec la version actuelle de la BD Adresse. Celui-ci utilise une reconnaissance syntaxique pour déchiffrer les adresses et leur associer un localisant ponctuel. Cet outil est utilisé prioritairement pour géolocaliser les adresses.
- Mais si certaines adresses sont rejetées car jugées incorrectement orthographiées pour l'outil ou bien non enregistrées dans la BD Adresse, un outil complémentaire est utilisé pour traiter ces adresses. Un script python utilise l'API de géocodage de Google. Les adresses non reconnues par ces deux outils sont placées dans un fichier texte.

Cette automatisation facilitera le travail de mise à jour de ces données pour les années à venir.

3.4.2 Données utilisées

Le Tableau 2 présente les données utilisées. Il est à noter que, pour la BD PARCELLAIRE, il n'y a pas d'année de référence unique sur l'ensemble des départements. Les versions les plus récentes ont été utilisées.

Tableau 2 : Données exploitées pour la mise en œuvre de la méthodologie MAJIC sur l'ensemble de la France.

base de données	données utilisées	année de référence
MAJIC CETE	Tables 'parcelle' et 'local'	2011
INSEE	RGP IRIS et fichiers détail logement	2011
BD TOPO	Bâti : 'bâti indifférencié', 'bâti industriel' et 'bâti remarquable' Commune	Avril 2014
BD PARCELLAIRE	parcelle	Version la plus récente
BD Adresse	Point Adresse	Avril 2014
Contour...IRIS		Avril 2012

3.4.3 Données élaborées

Les calculs de population ont été effectués par département. Une archive compressée est mise à disposition pour chacun d'eux et contient différentes données utiles. Une note technique⁴ décrit le contenu de cette archive.

Deux types de données sont fournis : des données de population spatialisée au bâtiment et des données de population agrégée sur des maillages réguliers.

- La méthodologie MAJIC permet de localiser les populations dans les bâtiments de la BD TOPO découpés selon les parcelles. Ces données sont fournies dans l'archive sous la forme de données spatialisées qui référencent pour chaque bâtiment les populations associées et divers indicateurs complémentaires tels que la zone administrative de référence (IRIS, TRIRIS ou communes), la densité d'habitants dans l'habitation, le nombre d'étages, ... Ce type de données sera plutôt utilisé pour évaluer les populations exposées à l'échelle d'un quartier ou d'une agglomération.
- Ces données aux bâtiments sont agrégées sur des mailles régulières d'une définition de 200m et 1km selon une règle de proportionnalité : la population d'un bâtiment croisant plusieurs mailles est répartie entre celles-ci proportionnellement à la surface d'intersection de ces mailles avec le bâtiment (étape 11 de la méthodologie). Ce mode de répartition garantit la cohérence des données de population entre un niveau local ou urbain et un niveau régional ou national. La définition géographique de ces deux carroyages est identique à celle utilisée par l'INSEE pour ses carroyages⁵.

3.4.4 Validation des données

Plusieurs sources d'incertitude ont été mises en évidence dans ce rapport : le recensement de la population, la localisation des locaux MAJIC et la qualité des données géographiques utilisées.

Certaines incertitudes ne sont pas détectables aisément par un algorithme de correction automatique. Deux sources principales d'incertitudes dans la méthodologie ont été ainsi répertoriées :

- Une première source est la localisation dans la base MAJIC de certains locaux sur une parcelle unique alors que ceux-ci sont répartis en réalité sur plusieurs parcelles.
- La seconde est le manque de couverture de la BD PARCELLAIRE vecteur. Pour un certain nombre de parcelles, uniquement le centre de parcelle est référencé et empêche ainsi de connaître précisément les bâtiments situés sur ces parcelles.

⁴ Létinois L., 2015. Fourniture et validation des données de population spatialisées selon la méthodologie nationale (méthodologie MAJIC) - Données de population livrées au 09/01/2015 pour l'année de référence INSEE 2011

⁵ http://www.insee.fr/fr/themes/detail.asp?reg_id=0&ref_id=donnees-carroyees

Le volume important des données et leur vaste couverture géographique ne permettent ni de réaliser une vérification exhaustive à une échelle nationale et ni de définir aisément une méthode de validation couvrant la totalité des situations. Une expertise humaine et locale est donc nécessaire.

Ce travail de validation doit tenir compte des incertitudes précédemment mentionnées mais aussi de l'utilisation qui sera faite des données. Une analyse très fine de ces dernières requiert un temps important et doit être proportionnée en fonction des applications envisagées et des autres sources possibles d'incertitude (par exemple la délimitation géographique des zones de dépassement dans le calcul de l'exposition).

La note technique d'utilisation de ces données propose des recommandations pour les analyser et les valider. Elle définit également les modalités d'échange entre l'AASQA et le LCSQA pour effectuer, le cas échéant, les corrections nécessaires et produire un nouveau jeu de données qui sera réintégré dans la base nationale.

A l'issue de cette étape de validation, le LCSQA sera en mesure d'agréger les carroyages de population départementaux dans un carroyage national. Celui-ci sera utilisé notamment dans le cadre de PREV'AIR Urgence et garantira ainsi une homogénéité entre les données locales et nationales.

3.5 Périmètre d'utilisation des données de population

Au moment de la rédaction de ce document, aucune indication spécifique n'a été fournie sur le périmètre d'utilisation de ces données. A défaut de directive, ces données seront utilisées uniquement dans le cadre d'une mission de service public.

Les données de population spatialisée sur les bâtiments selon la méthodologie MAJIC sont élaborées en croisant diverses bases de données dont celles de l'IGN (RGE). L'utilisation de ces données nécessite d'avoir au préalable signé une convention avec l'IGN pour l'exploitation des données du RGE⁶.

4. CONCLUSIONS ET PERSPECTIVES

La législation européenne et nationale imposent un suivi de la pollution dans l'air ambiant et de l'exposition des populations à cette pollution. Les textes de loi précisent que les zones géographiques de dépassement d'une valeur seuil doivent être délimitées ; ils requièrent en outre une estimation de la population touchée par ce dépassement. De nombreuses cartographies sont élaborées pour répondre à ces exigences. Deux variables sont nécessaires pour cartographier les populations exposées : les concentrations de polluant d'une part et la population spatialisée d'autre part. La précision du résultat est grandement liée à la qualité de ces informations ce qui suppose la mise en œuvre de méthodes appropriées. Dans ce contexte, et en s'appuyant sur les travaux conduits en France, le LCSQA a été chargé d'élaborer un cadre méthodologique harmonisé, adapté aux différentes échelles d'étude et applicable par les AASQA et le LCSQA. Le présent rapport porte spécifiquement sur le calcul de la répartition spatiale des populations.

⁶ NB : Cette obligation s'applique séparément au LCSQA et à chacune des AASQA.

Les précédents travaux réalisés sur ce sujet n'avaient pas permis d'établir une méthodologie unique applicable sur l'ensemble du territoire quelle que soit l'échelle considérée mais des voies possibles d'amélioration pour atteindre cet objectif avaient été mises en évidence. Les travaux menés en 2014 ont ainsi permis d'améliorer la méthodologie MAJIC et garantir une homogénéité des données de population spatialisée utilisées au niveau local et national dans le cadre de la surveillance de la qualité de l'air.

Des échanges ont eu lieu avec le CEREMA dans le cadre de l'Observatoire RhôneAlpin des Nuisances Environnementales (ORHANE). Les données de population élaborées par le LCSQA pour cette région y seront intégrées.

Les travaux menés par le LCSQA en 2014 ont permis d'automatiser les traitements pour limiter au maximum les travaux nécessaires à la mise à jour de la base de données de population. Du fait de contraintes sur l'accès aux données MAJIC et de la complexité de la méthodologie MAJIC, le LCSQA assure la mise en œuvre de cette approche et met à disposition des AASQA les données spatiales de la population qui en sont issues. Une première mise à jour est prévue au dernier trimestre de l'année 2015 qui utilisera le nouveau recensement de l'INSEE (IRIS 2012) et les données MAJIC 2013. Pour les années suivantes, il est envisagé d'ajuster chaque année les données de population en fonction du nouveau recensement de la population et d'utiliser tous les deux ans une mise à jour des données MAJIC.

Une note technique d'utilisation et de validation de ces données a été mise à disposition des AASQA. Celle-ci sera amenée à évoluer selon le retour d'expérience sur la validation des données. Ces dernières nécessitent une validation locale avant exploitation.

Il est à noter qu'un grand chantier est en cours associant la DGFIP et l'IGN : la RPCU. La Représentation Parcellaire Cadastre Unique a pour objectif de constituer un référentiel géographique unique (DGFIP/IGN) qui permettra :

- Une continuité des limites des parcelles, des feuilles cadastrales et des limites communales ;
- Des parcelles vectorisées sur l'ensemble du territoire ;
- Une cohérence entre le référentiel cadastral et les autres couches du RGE, notamment la BD TOPO.

La mise à jour cadastrale d'une parcelle devrait ainsi être appliquée plus rapidement et intégrée directement dans les données mises à disposition par l'IGN. La production a commencé cette année et le produit final devrait être disponible après 2020. En attendant la fin de ce chantier, l'IGN continue à vectoriser la BD PARCELLAIRE diminuant ainsi d'année en année l'incertitude associée à cette source de données.

5. REFERENCES

CERTU, 2005, Méthodes d'estimation des populations - http://www.observation-urbaine.certu.equipement.gouv.fr/IMG/pdf/Rapport_d_etude_cle28e914-1.pdf .

SETRA, 2007, Production des cartes de bruit stratégiques des grands axes routiers et ferroviaire, guide méthodologique - <http://www.setra.developpement-durable.gouv.fr/Production-des-cartes-de-bruit.html> .

CETE Normandie, 2001, Densité de population et morphologie du bâti - <http://www.certu-catalogue.fr/densite-de-population-et-morphologie-du-bati.html>

CETE de l'Ouest, 2011, Projet d'observatoire national des services de communication électronique.

CETE de Lyon, 2012, cartes de bruit 2012, affectation des populations

CETE Nord-Picardie, 2012, Fichiers fonciers 2009 et 2011 - Principes de traitement, outils choisis et descriptif de livraison.

CETE Nord-Picardie, 2012, Fichiers fonciers 2009 et 2011 - La géolocalisation des fichiers fonciers - Repères méthodologiques et résultats.

CETE Nord-Picardie, 2012, Fichiers fonciers 2009 et 2011 - Mise en lumière des fichiers fonciers bruts.

Létinois L., 2013, Méthodologie de répartition spatiale de la population. Rapport LCSQA, http://www.lcsqa.org/system/files/drc-12-126716-10615a_methodologiepopulation_0.pdf

Létinois L., 2015, Fourniture et validation des données de population spatialisée selon la méthodologie nationale (méthodologie MAJIC) - Données de population livrées au 09/01/2015 pour l'année de référence INSEE 2011



direction et secrétariat du LCSQA

INERIS - parc technologique Alata - BP 2 - F60550 Verneuil-en-Halatte
tél. 03 44 55 64 04 - www.lcsqa.org