

Note technique

FOURNITURE ET VALIDATION DES DONNEES DE POPULATION SPATIALISEES SELON LA METHODOLOGIE NATIONALE (METHODOLOGIE MAJIC)

Données de population livrées au
09/01/2015 pour l'année de référence
INSEE 2011

Laurent LÉTINOIS (INERIS)

SYNTHESE

Les données de population spatialisées sont couramment utilisées :

- pour évaluer l'exposition des populations telle que rapportée à l'Europe en application des directives,
- comme critère de choix dans l'implantation et la caractérisation des sites de mesure (sites fixes et campagnes de mesure),
- et comme variable auxiliaire dans des travaux de cartographie.

Des travaux LCSQA menés depuis 2012 ont permis d'élaborer une méthodologie nationale harmonisée de distribution géographique (ou « spatialisation ») de la population : cette méthodologie a été nommée méthodologie MAJIC en référence aux données MAJIC utilisées. Celle-ci est applicable à une échelle locale et nationale. Lors du dernier trimestre 2014, des jeux de données test ont été fournis à des AASQA volontaires et au CEREMA¹ pour validation. Début 2015, les données finales (année de référence INSEE 2011) ont été fournies à l'ensemble des AASQA pour une utilisation dans le cadre du rapportage de l'année 2013.

¹ Le CEREMA sera susceptible d'utiliser ces données de population spatialisée dans le cadre des observatoires du bruit.

Ces données de population peuvent nécessiter localement quelques ajustements selon l'expérience de chacune des AASQA. Des échanges entre le LCSQA et les AASQA permettront de corriger, si nécessaire, ces données pour obtenir une base de données spatialisée des populations homogène sur l'ensemble du territoire. La présente note décrit les vérifications locales qu'il est recommandé d'effectuer et définit les modalités d'échange entre le LCSQA et les AASQA.

1. INTRODUCTION

Les données de population spatialisées sont couramment utilisées dans un cadre réglementaire pour évaluer et rapporter à la Commission Européenne l'exposition des populations à la pollution atmosphérique, comme critère de choix dans l'implantation et la caractérisation des sites de mesure (sites fixes et campagnes de mesure) et pour des travaux de cartographie.

Des travaux LCSQA menés depuis 2012 ont abouti à une méthodologie nationale harmonisée de spatialisation de la population appelée méthodologie MAJIC. Lors du dernier trimestre 2014, des jeux de données test ont été fournis à certaines AASQA et au CEREMA pour validation. Début 2015, les données finales (année de référence INSEE 2011) ont été fournies à l'ensemble des AASQA pour une utilisation pour le reportage de l'année 2013.

Cette note a pour objet de présenter succinctement la méthodologie de spatialisation, de décrire les fichiers délivrés par le LCSQA pour les AASQA, de proposer des pistes pour valider les données de population et définir les échanges nécessaires entre le LCSQA et les AASQA pour corriger, s'il est besoin, les données transmises.

2. RAPPEL SUR LA METHODOLOGIE MAJIC

Cette partie présente succinctement la méthode employée pour répartir les populations INSEE sur les bâtiments de la BD Topo. Cette méthode est décrite dans le guide méthodologique de 2013². Elle sera plus amplement détaillée dans une mise à jour du guide qui paraîtra en 2015. Les fichiers MAJIC fournis par la DGFIP³ référencent toutes les parcelles cadastrales françaises (DROM compris) et les locaux associés à celles-ci. De nombreuses informations y sont disponibles : notamment l'usage de ces locaux (habitations, commerce), la surface de chacun de ceux-ci, le nombre d'étages, le nombre de pièces, le type d'habitat, ...

² Létinois L., 2013. Méthodologie de répartition spatiale de la population. Rapport LCSQA, http://www.lcsqa.org/system/files/drc-15-144366-01026a_methodologiepopulation_vf.pdf

³ Direction Générale des Finances Publiques

Le CEREMA (Direction territoriale Nord-Picardie, anciennement CETE Nord-Picardie) a pour mission de mettre en forme ces fichiers pour l'ensemble de la sphère Ecologie et de les spatialiser en utilisant en priorité la couche 'Parcelles' de la BD Parcellaire de l'IGN. Le produit (BD MAJIC CETE) fourni au LCSQA est une base de données Postgres/PostGIS par département comportant plusieurs tables. Il est à noter que toutes les parcelles n'y sont pas nécessairement spatialisées du fait que la BD Parcellaire Vecteur ne couvre pas actuellement l'ensemble du territoire français. Lors de l'élaboration de la base, lorsque le contour de la parcelle n'était pas connu, le CEREMA a utilisé d'autres sources d'information pour essayer de localiser à minima le centre des parcelles. La base fournie comporte donc une qualité de localisation différente selon les parcelles.

Différentes sources sont utilisées par le LCSQA pour répartir la population:

- La BD Parcellaire :
 - la couche parcelle : les travaux du CETE Nord-Picardie datant de début 2012, il est possible d'obtenir plus de parcelles vectorisées en utilisant une version plus récente de la BD Parcellaire.
- La BD Topo :
 - Les couches 'bâti indifférencié', 'bâti industriel' et 'bâti remarquable' référencent les bâtiments sur l'ensemble du territoire. Ces bâtiments sont utilisés pour y placer les locaux d'habitation.
 - La couche commune de la BD Topo définit les contours géographiques des communes françaises.
- La BD Contours IRIS référence les contours géographiques des IRIS. De qualité médiocre, les limites communales sont corrigées par la couche commune de la BD Topo.
- Le fichier Détail Logement (2011) de l'INSEE rapporte par IRIS la répartition des populations résidant dans des maisons et des appartements sur l'année de référence 2011.
- Le recensement général de la population de l'INSEE 2011 fournit le nombre de personnes habitant dans chacun des IRIS.

La méthodologie de répartition des populations selon la méthodologie MAJIC se décompose en plusieurs étapes :

- Une première étape consiste à extraire par parcelle un nombre de locaux MAJIC d'habitation de type 'maison' et 'appartement'.
- Une seconde étape affecte les locaux aux bâtiments de la BD Topo susceptible de les accueillir. Les bâtiments sont au préalable découpés en fonction des contours des parcelles. Comme indiqué précédemment, les parcelles MAJIC ont une qualité de géolocalisation variable. Trois qualités de localisation existent pour ces parcelles : soit le contour géographique est connu, soit le centre de la parcelle est connu ou soit l'adresse postale des locaux associés à celle-ci est connue. Les traitements pour spatialiser les locaux d'habitation associés à ces parcelles diffèrent selon la qualité du référencement géographique de celles-ci :

- o le contour géographique de la parcelle est connu : les locaux sont distribués sur chacun des bâtiments se trouvant sur la parcelle au prorata de la surface habitable de chacun.
- o le centre de la parcelle est connu : parmi les bâtiments les plus proches du centre de parcelle, l'algorithme recherche ceux qui ont une surface habitable suffisamment importante pour accueillir l'ensemble de ces locaux. Un ou plusieurs bâtiments peuvent être retenus. Le rayon de la recherche des bâtiments proches est égal à la racine carrée de la surface de la parcelle. Si aucun bâtiment n'est trouvé, un bâtiment fictif est créé selon la surface habitable recensée dans la BD MAJIC, l'emprise au sol de ce bâtiment prend en compte le nombre d'étages recensés dans MAJIC.
- o seule l'adresse des locaux situés sur la parcelle est connue : la parcelle est lors d'une première étape localisée à partir de son adresse (géocodage par la BD Adresse) pour ensuite, dans une seconde étape, rechercher les bâtiments les plus proches de la parcelle (à l'identique des centres de parcelle connus). Le géocodage ne permet pas de localiser à 100% les adresses : les adresses rejetées sont stockées dans une table avec les caractéristiques des locaux associés et fournis à l'AASQA.

Le produit final de cette étape est une base de données spatiale contenant par bâtiment le nombre de locaux de type habitation (maison ou appartement). La qualité de localisation apparaîtra dans la BD finale de population dans le champ 'TypeBati' qui contiendra les valeurs suivantes : 'Bat' : contour parcelle BD Parcellaire ayant des bâtiments recensés sur celle-ci, 'Parc' : contour parcelle BD Parcellaire n'ayant aucun bâtiment recensé sur celle-ci (création d'un bâtiment fictif), 'CP' : centre parcelle BD Parcellaire, 'NV' : géocodage des adresses des locaux MAJIC.

- Une troisième étape confronte, pour chaque IRIS, le nombre de locaux MAJIC avec les statistiques de population INSEE pour établir un nombre d'habitants par type de local MAJIC : division du nombre d'habitants INSEE par type d'habitat (source Détails Logements et recensement de la population) par la somme des locaux MAJIC de l'IRIS par type d'habitat (maison/appartement). Le croisement de ces deux informations permet d'établir un nombre moyen d'habitants par local MAJIC selon le type d'habitat. L'analyse de ce nombre d'habitants par local fait ressortir pour certains IRIS des valeurs extrêmes, par exemple une dizaine d'habitants par local de type appartement. Ces chiffres extrêmes 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 par un manque d'information dans les données MAJIC. Dans le cas où des valeurs aberrantes apparaissent (> à 5 habitants par local), l'algorithme de calcul tente de trouver une valeur plus cohérente dans des données de population plus agrégées : statistiques par type de local sur un regroupement d'IRIS (TRIIRIS au sens INSEE), statistiques par type de local sur la commune qui contient les IRIS ou statistiques générales de population sur la commune. Le type de statistique de l'INSEE utilisé est reporté dans la BD finale de population via le champ 'TypPop2011' qui prendra 4 valeurs possibles : 'I' : population estimée à partir de la population de l'IRIS en distinguant les populations par type d'habitat, 'T' : population estimée à partir de la population du TRIIRIS en distinguant les populations par type d'habitat, 'C' : population estimée à partir de la population de la commune en distinguant les

populations par type d'habitat, 'M' : population estimée à partir de la population de la commune sans distinction des populations par type d'habitat

- Une quatrième étape répartit les populations sur chacun des bâtiments en utilisant le nombre moyen d'habitants par type de local.
- Une dernière étape corrige les populations dans les bâtiments où apparaît une trop forte densité d'habitants : au-delà de 1 habitant pour 10 m² d'habitation. Pour ces bâtiments, l'algorithme tente de distribuer la population en surplus sur des bâtiments proches dans un rayon de 100 m susceptible d'accueillir de la population complémentaire. Si des bâtiments sont trouvés, la population est répartie sur les habitations selon la surface habitable disponible et l'indicateur 'TypeBati' est modifié en ajoutant le préfixe 'A_'. On aura dans le cas des bâtiments modifiés un indicateur de la forme suivante : 'A_Bat', 'A_NR', 'A_Parc'. Si aucun bâtiment complémentaire n'est trouvé, la population du bâtiment est laissée telle quelle.

3. DESCRIPTIONS DES LIVRABLES

Les données de population sont mises à disposition sur un espace ftp du LCSQA. L'url du serveur ftp est la suivante : <ftp.lcsqa.org>. La connexion à cet espace nécessite un nom d'utilisateur et un mot de passe qui sont respectivement : public et public. Les données se situent dans le répertoire donnees_population.

Une archive zip est fournie par département contenant plusieurs fichiers de type csv ou shapefile.

La chaîne de caractères 'XX' utilisée dans la suite du document représente le numéro du département. Ce chapitre détaille le contenu des différents fichiers fournis.

Les données suivantes sont fournies dans l'archive :

- Le fichier '**Descriptifs_2011_DXX.txt**' fournit un descriptif de l'ensemble des données fournies et présente également les références des données utilisées en entrée du calcul des populations.
- Le shapefile '**PopulationBati_2011_DXX.shp**' répertorie tous les bâtiments dans lesquels des populations ont été placées. Les champs utilisés dans ce shapefile sont les suivants :
 - DcomIris : code de l'IRIS 2011 contenant le bâtiment
 - TypeBati : Qualité de la localisation des locaux MAJIC : Bat : contour parcelle BD Parcellaire qui contient au moins un bâtiment, Parc : contour parcelle BD Parcellaire n'ayant pas de bâtiment recensé sur celle-ci (création d'un bâtiment fictif), CP : centre parcelle BD Parcellaire, NV : géocodage des adresses des locaux MAJIC. Un préfixe 'A_' peut être ajouté pour des bâtiments ayant un ajustement de population pour corriger une densité d'habitants trop importante. (Cf. chapitre précédent)
 - NbEtages : Nombre de niveaux calculé du bâtiment
 - Surf3d : Surface développée habitable du bâtiment : surface au sol multiplié par le nombre de niveaux
 - PopM2011 : Population 2011 résidant dans un habitat de type maison
 - PopA2011 : Population 2011 résidant dans un habitat de type appartement (Ménages + Hors Ménages)
 - PopT2011 : Population 2011 totale (maison + appartement)

- DenP2011 : Densité de population (PopT2011/Surf3d)
- TypPop2011 : indique la qualité des données INSEE utilisées pour répartir les populations (cf. chapitre précédent) : 'I' : population estimée à partir de la population de l'IRIS 2011 en distinguant les populations par type d'habitat, 'T' : population estimée à partir de la population du TRIIRIS 2011 en distinguant les populations par type d'habitat, 'C' : population estimée à partir de la population de la commune 2011 en distinguant les populations par type d'habitat, 'M' : population estimée à partir de la population de la commune 2011 sans distinction des populations par type d'habitat
- IdRef : Identifiant de la zone administrative de référence : code IRIS 2011, code TRIIRIS 2011 ou code commune 2011 selon la valeur du champ TypePop2011
- Perclris : pourcentage de la population de la zone administrative de référence contenue dans le bâtiment
- IDBATI : Identifiant unique du bâtiment
- ID : identifiant du bâtiment Bd Topo. Lorsque le bâtiment est un bâtiment fictif, ce champ est renseigné avec la valeur suivante : 'BatiFictif'.
- TypTopo : Indicateur de la couche BD Topo d'origine du bâtiment : 0 : 'bâti indifférencié', 1 : 'bâti industriel', 2 : 'bâti remarquable'.
- Le shapefile '**BatiNonUtilise_2011_DXX.shp**' contient les bâtiments ne comportant aucune population. Les champs utilisés dans ce shapefile sont les suivants :
 - Surf3d : Surface développée habitable du bâtiment : surface au sol multiplié par le nombre de niveaux
 - NbEtages : Nombre de niveaux calculé du bâtiment
 - TypeBati : Source du Bâti : 'V' bâti sur zone BD Parcellaire vectorisé, 'NV' : bâti sur zone en dehors BD Parcellaire vectorisé
 - DcomIris : code de l'IRIS 2011 contenant le bâtiment
 - IDBATI : Identifiant unique du bâtiment
 - ID : identifiant du bâtiment Bd topo.
 - TypTopo : Indicateur de la couche BD Topo d'origine du bâtiment : 0 : 'bâti indifférencié', 1 : 'bâti industriel', 2 : 'bâti remarquable'.
- Le shapefile '**IrisStat_XX.shp**' contient la géométrie des IRIS 2011 du département. Différents champs apparaissent :
 - DcomIris : code de l'IRIS 2011
 - InseeCom : code de la commune 2011
 - TRIIRIS : code du TRIIRIS 2011 s'il existe, sinon est fixé à 'ZZZZZZ'
 - PopT : Population 2011 municipale (Ménages + Hors Ménages)
 - PopM : Population 2011 ménages
 - PopHM : Population 2011 hors ménages
 - PopMM : Population 2011 résidant dans un habitat de type maison
 - PopMA : Population 2011 résidant dans un habitat de type appartement
 - TypPop: indique la qualité des données INSEE 2011 utilisées pour répartir les populations (cf. chapitre précédent) : 'I' : population estimée à partir de la population de l'IRIS 2011 en distinguant les populations par type d'habitat, 'T' : population estimée à partir de la population du TRIIRIS 2011 en distinguant les populations par type d'habitat, 'C' : population estimée

à partir de la population de la commune 2011 en distinguant les populations par type d'habitat, 'M' : population estimée à partir de la population de la commune 2011 sans distinction des populations par type d'habitat.

- IdRef : Identifiant de la zone administrative de référence : code IRIS 2011, code TRIIRIS ou code commune selon la valeur du champ TypPop
- PopAdmin : population municipale 2011 de la zone administrative de référence
- Le fichier **'PopulationIris_2011_DXX.txt'** présente les mêmes champs que le shapefile précédent.
- Le fichier **'AdresseNonGeocodees_2011_DXX.txt'** présente les adresses des locaux pour lesquels le script de géocodage n'a pu trouver de localisation. Les champs utilisés sont les suivants :
 - IdCom : code INSEE 2011 de la commune
 - NomCom : nom de la commune
 - dnvoiri : numéro de la voirie
 - dvoilib : libellé de la voirie
 - NbLocM : Nombre de locaux d'habitations MAJIC de type maison
 - NbLocA : Nombre de locaux d'habitations MAJIC de type appartement
- Le shapefile **'MaillagePop200m_2011_DXX.shp'** contient des données de population (couche PopulationBati_2011_DXX.shp) agrégées sur un maillage régulier d'une définition de 200 m. La définition de ce maillage est identique au carroyage utilisé par l'INSEE. Les mailles contiennent uniquement les populations résidentes dans le département XX. Les populations des autres départements ne sont pas prises en compte. Les champs utilisés sont les suivants :
 - IdM200m : identifiant unique de la maille 200m
 - IdM1Km : identifiant de la maille 1 km contenant la maille 200 m
 - PopT2011 : Population municipale 2011 (Ménages + Hors Ménages)
- Le shapefile **'MaillagePop1km_2011_DXX.shp'** contient des données de population (couche PopulationBati_2011_DXX.shp) agrégées sur un maillage régulier d'une définition de 1 km. La définition de ce maillage est identique au carroyage utilisé par l'INSEE. Les mailles contiennent uniquement les populations résidentes dans le département XX. Les populations des autres départements ne sont pas prises en compte. Les champs utilisés sont les suivants :
 - IdM1Km : identifiant de maille 1 km contenant la maille 1 km
 - PopT2011 : Population municipale 2011 (Ménages + Hors Ménages)
- Le fichier csv **'TabMaillagePop200m_2011_DXX.csv'** contient le détail des populations utilisées par maille de 200 m. Si une maille croise plusieurs IRIS, ce fichier présente pour chaque IRIS intersecté le ratio de population de l'IRIS dans la maille. Par exemple, si une maille croise 3 IRIS, trois lignes seront présentes dans le fichier pour donner les ratios de population des 3 IRIS. Les champs utilisés sont les suivants :
 - IdM200m : identifiant unique de la maille 200m
 - IdM1Km : identifiant de la maille 1 km contenant la maille 200m
 - PopT2011 : Population municipale 2011 (Ménages + Hors Ménages)
 - TypPop2011 : indique la qualité des données INSEE pour répartir les populations (cf. chapitre précédent) : 'I' : population estimée à partir de la population de l'IRIS 2011 en distinguant les populations par type d'habitat,

- 'T': population estimée à partir de la population du TRIIRIS 2011 en distinguant les populations par type d'habitat, 'C' : population estimée à partir de la population de la commune 2011 en distinguant les populations par type d'habitat, 'M' : population estimée à partir de la population de la commune 2011 sans distinction des populations par type d'habitat.
- IdRef : Identifiant de la zone administrative de référence : code IRIS 2011, code TRIIRIS 2011 ou code commune 2011 selon la valeur du champ TypPop2011
- Perclris : pourcentage de population de la zone administrative de référence contenue dans la maille
- Le fichier csv **'TabMaillagePop1km_2011_DXX.csv'** contient le détail des populations utilisées par maille de 1 km. Si une maille croise plusieurs IRIS, ce fichier présente pour chaque IRIS intersecté le ratio de population de l'IRIS dans la maille. Par exemple, si une maille croise 3 IRIS, trois lignes seront présentes dans le fichier pour donner les ratios de population des 3 IRIS. Les champs utilisés sont les suivants :
 - IdM1Km : identifiant de la maille 1 km
 - PopT2011 : Population municipale 2011 (Ménages + Hors Ménages)
 - TypPop2011 : indique la qualité des données INSEE pour répartir les populations (cf. chapitre précédent) : 'I' : population estimée à partir de la population de l'IRIS 2011 en distinguant les populations par type d'habitat, 'T': population estimée à partir de la population du TRIIRIS 2011 en distinguant les populations par type d'habitat, 'C' : population estimée à partir de la population de la commune 2011 en distinguant les populations par type d'habitat, 'M' : population estimée à partir de la population de la commune 2011 sans distinction des populations par type d'habitat.
 - IdRef : Identifiant de la zone administrative de référence : code IRIS 2011, code TRIIRIS 2011 ou code commune 2011 selon la valeur du champ TypPop2011
 - Perclris : pourcentage de population de la zone administrative de référence contenue dans la maille

L'archive contient deux shapefile complémentaires destinés au CEREMA. En effet, dans le cadre des observatoires du bruit, le CEREMA utilise actuellement la BD Topo pour évaluer les populations exposées aux nuisances sonores. Pour cela, la méthode utilise les façades des bâtiments BD Topo en bordure de voie pour repérer les populations. La méthodologie MAJIC actuelle découpe les bâtiments BD Topo en fonction des contours des parcelles. Ceci a pour effet de ne pas nécessairement affecter les populations du bon côté du bâtiment. Les données de population sont donc ré-agrégées en fonction de l'identifiant BD Topo pour fournir des populations par bâtiment BD Topo au CEREMA. Une perte de finesse de l'information sur la répartition spatiale peut être constatée localement à l'échelle d'un îlot urbain.

Les deux shapefile élaborés pour le CEREMA sont les suivants :

- **'BdTopoAvecPop_2011_DXX.shp'** : ce shapefile répertorie tous les bâtiments BD Topo dans lesquels des populations ont été placées. Les champs utilisés dans ce shapefile sont les suivants :
 - PopM2011 : Population 2011 résidant dans un habitat de type maison

- PopA2011 : Population 2011 résidant dans un habitat de type appartement (Ménages + Hors Ménages)
- PopT2011 : Population 2011 municipale (Ménages + Hors Ménages)
- ID : identifiant du bâtiment Bd topo. Lorsque le bâtiment est un bâtiment fictif, ce champ est renseigné avec la valeur suivante : 'BatiFictif'.
- TypTopo : Indicateur de la couche BD Topo d'origine du bâtiment : 0 : 'bâti indifférencié', 1 : 'bâti industriel', 2 : 'bâti remarquable'.
- Hauteur : hauteur référencée dans la BD Topo,
- NbEtages : Nombre de niveaux calculé du bâtiment
- **'BdTopoSansPop_2011_DXX.shp'** : ce shapefile répertorie tous les bâtiments BD Topo sans population. Les champs utilisés dans ce shapefile sont les suivants :
 - ID : identifiant du bâtiment Bd Topo. Lorsque le bâtiment est un bâtiment fictif, ce champ est renseigné avec la valeur suivante : 'BatiFictif'.
 - TypTopo : Indicateur de la couche BD Topo d'origine du bâtiment : 0 : 'bâti indifférencié', 1 : 'bâti industriel', 2 : 'bâti remarquable'.
 - Hauteur : hauteur référencée dans la BD Topo,
 - NbEtages : Nombre de niveaux calculé du bâtiment

Lors de l'import des données spatiales dans un logiciel SIG, il est nécessaire que le système de projection soit correctement reconnu dans le logiciel.

Les projections utilisées sont les suivantes et correspondent à celles utilisées par l'IGN pour la fourniture des données du RGE :

- Métropole : Lambert 93 (code EPSG 2154)
- Guadeloupe : UTM Zone 20 N (code EPSG : 32620)
- Martinique : UTM Zone 20 N (code EPSG : 32620)
- Guyane : RGFG 1995 UTM Zone 22N (code EPSG : 2972)
- La Réunion : RGR 1992 UTM Zone 40S (code EPSG : 2975)

Pour le logiciel MapInfo, il est conseillé d'utiliser l'outil 'Traducteur universel' pour transformer les shapefile en tables MapInfo standard en spécifiant la projection du fichier de destination.

L'exploitation des données spatiales impose un agrément d'utilisation de ces données par l'IGN. Cet agrément se traduit par la signature d'une convention entre l'AASQA et l'IGN.

4. VALIDATION DES DONNEES ET ECHANGES SUR LES DONNEES

La grande quantité de données et leur vaste couverture géographique ne permettent pas d'en faire une vérification exhaustive à l'échelle nationale ni de définir aisément une méthode de validation couvrant la totalité des situations. Il est toutefois possible d'effectuer localement des contrôles. Ce chapitre fournit aux AASQA quelques recommandations sur le processus de validation.

Une observation trop fine des données n'est pas utile et nécessiterait un temps trop important d'analyse. L'oubli de quelques maisons dans un lotissement n'a a priori pas d'incidence importante sur la répartition spatiale de la population. A l'inverse, l'oubli d'une barre d'immeuble dans une zone fortement urbanisée peut être problématique selon le contexte local. De manière générale, les populations des bâtiments « oubliés » seront nécessairement réparties sur l'ensemble des bâtiments de la zone administrative de rattachement.

Une première étape recommandée est d'établir des statistiques générales par département sur la couche 'PopulationBati_2011_DXX.shp' :

- L'analyse des valeurs minimales et maximales de certains champs (hauteur, Surf3d, DenP2011 ...) peut permettre de repérer certaines valeurs aberrantes.
- Il convient de vérifier que les populations cumulées de tous les bâtiments par département, ou par zone administrative de référence (IRIS, TRIRIS ou communes définies selon les champs TypPop et IdRef) sont identiques aux populations recensées du département ou des zones administratives.

L'objectif premier de ces données étant le calcul des populations exposées à des dépassements, il peut être judicieux de focaliser la validation sur les zones où des dépassements pourraient être relevés : sur ces zones, une analyse des bâtiments ayant une densité d'habitants trop importante ($>0,1$ habitants/m²) et ayant un nombre d'habitants considéré comme élevé (> 10 ou 15) sera conduite.

Certaines incertitudes existent dans la BD MAJIC CETE mais ne sont pas détectables aisément par un algorithme de correction automatique. Une expertise humaine est donc nécessaire. Des éléments sont présentés dans la suite de ce chapitre.

4.1 Locaux sur parcelle de référence

Certains locaux sont en réalité situés sur plusieurs parcelles mais sont référencés dans la base MAJIC sur une seule parcelle dite parcelle de référence. De ce fait, lors de l'extraction des données MAJIC, ces locaux sont placés uniquement sur la parcelle de référence. Dans certains cas, le nombre de locaux peut atteindre plusieurs centaines d'habitations. La densité d'habitants est alors très importante sur le ou les bâtiments associés à la parcelle. L'algorithme essaye de corriger l'écart constaté en recherchant des bâtiments dans un rayon de 100 mètres susceptibles d'accueillir de la population. Dans la majorité des cas, cette correction suffit mais pour certaines zones, certains bâtiments importants qui devraient avoir des habitants ne sont pas retenus.

Un exemple est présenté sur la ville de Rennes dans la carte Figure 1 où des locaux MAJIC sont situés uniquement sur une passerelle de référence. Différentes informations sont présentées sur cette vue :

- les bâtiments contenant des habitants sont présentés en surface pleine jaune, le nombre d'habitants associé à ces bâtiments est écrit en rouge sur ceux-ci et l'identifiant du bâtiment est écrit en bleu foncé.
- Les bâtiments sans habitant sont représentés en surface transparente avec un contour bleu clair. L'identifiant de chacun de ces bâtiments est écrit en utilisant une couleur bleu foncé.

- Une photographie aérienne (BD Ortho) est placée en fond de carte.

Il apparaît sur cette vue que certains immeubles d'habitation ne comportent pas d'habitants après traitement des données MAJIC. Au vu de cette carte et des informations sur la densité d'habitants dans ces bâtiments, on peut décomposer les bâtiments en deux groupes (délimités en rouge). On remarque pour le groupe 1 que les locaux MAJIC étaient situés initialement sur le premier bâtiment de gauche. L'algorithme de spatialisation a repéré une densité trop forte sur ce bâtiment et a donc cherché des bâtiments proches de celui-ci pour leur affecter une partie des habitants. L'algorithme a trouvé les deux bâtiments situés directement sur sa droite et a transféré une partie des habitants dans ceux-ci. La densité d'habitants devenant acceptable, l'algorithme s'est arrêté à ces bâtiments. Si l'opérateur de validation juge que les habitants devraient être transférés dans les autres bâtiments du groupe 1, une correction est alors nécessaire et doit passer par un échange avec le LCSQA.



Figure 1 : locaux d'habitation placés uniquement sur une parcelle de référence (Rennes)

La solution proposée par le LCSQA pour corriger cette erreur est de répertorier tous les bâtiments d'un même groupe et de répartir ensuite les populations sur ces bâtiments au prorata de leur surface habitable. On propose donc de créer un fichier csv (séparateur ',') dont le modèle (ModeleCorrectionGroupeBati.txt) est disponible sur l'espace ftp du LCSQA où sont mises à disposition les données de population. Ce fichier contiendra sous forme de tableau le regroupement des bâtiments. Il sera nommé : CorrectionGroupeBati_DXX.csv. 'XX' correspond au numéro du département, par exemple 08 ou 34.

Ce fichier se composera des colonnes suivantes :

- IdGroupe : identifiant du groupe de bâtiment

- TypeBati : indicateur de l'origine du bâtiment. Deux valeurs sont proposées : '0' pour les bâtiments n'ayant pas population recensée et '1' pour les constructions avec population
- IdBati : identifiant unique du bâtiment issu selon le cas du shapefile 'BatiBdTopoNonUtilise_2011_DXX.shp' ou 'PopulationBati_2011_DXX.shp'.

Dans l'exemple fourni Figure 1, nous aurions un contenu du fichier 'CorrectionGroupeBati_D35.csv' de la forme suivante :

IdGroupe,TypeBati,IdBati

1,1,374912

1,1,376386

1,1,375627

1,1,375465

1,0,130384

1,0,130386

1,0,130388

1,0,130389

2,1,234708

2,0,130396

4.2 Adresses non géocodées

Nous avons vu dans le chapitre 2 qu'un géocodage à l'adresse est utilisé pour les parcelles pour lesquelles ni le contour et ni le centre de la parcelle ne sont connus. L'algorithme de calcul utilise l'adresse des locaux de la parcelle pour géolocaliser cette parcelle. Si l'adresse n'est pas trouvée par l'algorithme de géocodage, celle-ci est renseignée dans le fichier '**AdresseNonGeocodees_2011_DXX.txt**' avec le nombre de locaux associés. Pour les adresses présentant un nombre de locaux importants, l'AASQA procèdera, si elle le juge utile, à une recherche de la position des locaux manquants et fournira au LCSQA un fichier au format csv (séparateur ',') contenant la position des locaux. Le fichier sera nommé : AjoutAdresseNonLocalisee_DXX.csv. 'XX' correspond au numéro du département, par exemple 08 ou 34. Le modèle du fichier (ModeleAjoutAdresseNonLocalisee.txt) est disponible sur l'espace ftp du LCSQA où sont mises à disposition les données de population.

Ce fichier se composera des colonnes déjà utilisées dans le fichier d'adresses (IdCom,NomCom,dnvoiri,dvoilib,NbLocM,NbLocA) et des deux colonnes supplémentaires suivantes :

- X : valeur entière de l'abscisse des coordonnées du centre de la parcelle exprimées en mètre
- Y : valeur entière de l'ordonnée des coordonnées du centre de la parcelle exprimées en mètre

Ces coordonnées sont exprimées dans le même référentiel que les données fournies par le LCSQA (cf. chapitre 3).

Le contenu du fichier se présentera sous la forme suivante :

IdCom,NomCom,dnvoiri,dvoilib,NbLocM,NbLocA,X,Y

35023,Bédée,5482,LA LANDE,1,0,365231,6691095

35120,Gévezé,5268,LA MAISON NEUVE,1,0,353712,6698307

35257,Saint-Brice-en-Coglès,5254,LA MAISON NEUVE,1,0,353357,6791095

Le contenu de ce fichier est identique au fichier 'AdresseNonGeocodees_2011_DXX.txt' hormis l'ajout des deux dernières colonnes X et Y.

4.3 Correction de la base de données spatialisée

Si une AASQA souhaite apporter des modifications à la spatialisation des données par l'intermédiaire des deux fichiers 'AjoutAdresseNonLocalisee_DXX.csv' et 'CorrectionGroupeBati_DXX.csv', celle-ci transmettra ces données transmis par email à Laurent Létinois (laurent.letinois@ineris.fr). Un nouveau calcul sera lancé par le LCSQA pour intégrer le contenu de ces fichiers. Le LCSQA mettra alors à disposition les données mises à jour (données spatialisées sur le bâti et données agrégées par maille de 200 m et 1 km).

L'AASQA portera une attention particulière sur le bon respect du format de ces fichiers afin d'éviter tout échec lors du traitement automatique de ces fichiers :

- le nom des colonnes ne devra en aucun cas être modifié,
- le séparateur de champs sera le ',',
- les coordonnées seront exprimées en valeur entière,
- les valeurs des coordonnées seront dans le même référentiel que les données fournies par le LCSQA,
- les valeurs des coordonnées seront situées sur la zone du département,
- le nommage des fichiers suivra les indications du LCSQA.

Si un autre problème est détecté par une AASQA, celle-ci en informera le LCSQA qui examinera avec elle la manière de le traiter.

5. PERIMETRE D'UTILISATION DES DONNEES DE POPULATION

Faute d'information actuelle sur l'utilisation possible de ces données, les AASQA se limiteront à utiliser celles-ci dans le cadre d'une mission de service public. Le périmètre d'utilisation pourra évoluer selon l'avis de la DGALN⁴.

⁴ Direction Générale de l'Aménagement, du Logement et de la Nature sous l'autorité conjointe du ministre chargé de l'Écologie et du Développement durable et du ministre chargé de l'Égalité des territoires et du Logement

6. CONCLUSION

Les populations françaises pour l'année de référence INSEE 2011 ont été spatialisées par le LCSQA pour la première fois en utilisant la méthodologie MAJIC. Ces données ont été livrées sur un espace ftp à destination des AASQA depuis le 9 janvier 2015. Les limites de la méthodologie étant connues, ces données peuvent nécessiter selon la zone géographique et selon l'appréciation de l'AASQA quelques ajustements. Malgré une validation réalisée par plusieurs AASQA en fin d'année 2014, certains cas particuliers peuvent encore apparaître. C'est pourquoi l'ensemble des AASQA est invité à réaliser un contrôle local des données. Un retour sur ces données serait souhaitable pour le 1^{er} avril 2015.

Après avoir pris en compte les erreurs signalées par les AASQA sur ces données, le LCSQA mettra à disposition une nouvelle version de la base de données.

Ainsi l'application de la méthodologie MAJIC et les échanges consécutifs entre le LCSQA et les AASQA permettront-ils aux différents utilisateurs de disposer de données de population spatialisées de manière homogène sur l'ensemble du territoire. Une mise à jour de ces données sera réalisée à la fin du quatrième trimestre 2015. Elle utilisera l'année de référence INSEE 2012.