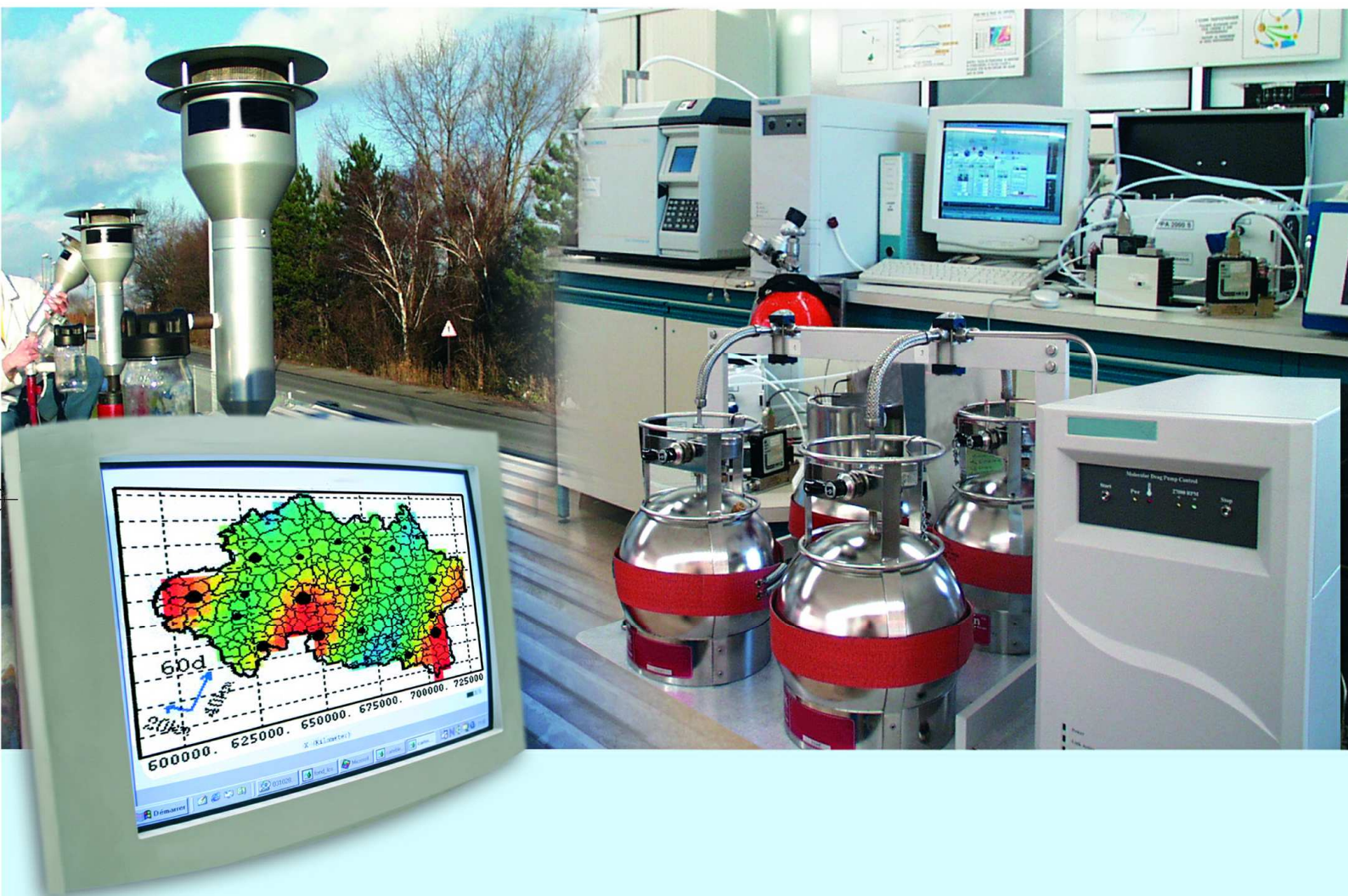




## Laboratoire Central de Surveillance de la Qualité de l'Air



Traitements numériques

**Méthodologie de répartition spatiale de la population**

Août 2013

Programme 2012

L. LÉTINOIS







## PREAMBULE

### **Le Laboratoire Central de Surveillance de la Qualité de l'Air**

**Le Laboratoire Central de Surveillance de la Qualité de l'Air est constitué de laboratoires de l'Ecole des Mines de Douai, de l'INERIS et du LNE. Il mène depuis 1991 des études et des recherches finalisées 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'Energie et du Climat (bureau de la qualité de l'air) du Ministère de l'Ecologie, du Développement durable, des Transports et du Logement. 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.**





# Méthodologie de répartition spatiale de la population

Laboratoire Central de Surveillance  
de la Qualité de l'Air

## Traitements numériques

Programme financé par la  
Direction Générale de l'Energie et du Climat (DGEC)

2012

L. LÉTINOIS

Ce document comporte 69 pages (hors couverture et annexes)

|         | Rédaction                                                                                                                                   | Vérification                                                                                                                                  | Approbation                                                                                                           |
|---------|---------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------|
| NOM     | Laurent LÉTINOIS                                                                                                                            | Bertrand BESSAGNET                                                                                                                            | Laurence ROUIL                                                                                                        |
| Qualité | Ingénieur de l'Unité<br>Modélisation Atmosphérique<br>et Cartographie<br>Environnementale (MOCA)<br><br>Direction des Risques<br>Chroniques | Responsable de l'Unité<br>Modélisation Atmosphérique<br>et Cartographie<br>Environnementale (MOCA)<br><br>Direction des Risques<br>Chroniques | Responsable du Pôle<br>Modélisation Environnementale<br>et Décision (DECI)<br><br>Direction des Risques<br>Chroniques |
| Visa    |                                                          |                                                            |                                  |



## TABLE DES MATIÈRES

|                                                                                          |           |
|------------------------------------------------------------------------------------------|-----------|
| <b>REMERCIEMENTS .....</b>                                                               | <b>9</b>  |
| <b>LISTE DES ABBREVIATIONS ET ACCRONYMES.....</b>                                        | <b>10</b> |
| <b>RESUME .....</b>                                                                      | <b>11</b> |
| <b>SYNTHESE DE LA METHODOLOGIE .....</b>                                                 | <b>12</b> |
| <b>1. INTRODUCTION .....</b>                                                             | <b>17</b> |
| <b>2. DEFINITIONS .....</b>                                                              | <b>18</b> |
| 2.1 La population en France.....                                                         | 18        |
| 2.1.1 Le recensement de la population en France.....                                     | 18        |
| 2.1.2 Le découpage administratif INSEE .....                                             | 20        |
| 2.1.3 Diffusion des résultats du recensement.....                                        | 21        |
| 2.2 Les bases de données spatiales .....                                                 | 21        |
| 2.2.1 Le référentiel à grande échelle (RGE) .....                                        | 21        |
| 2.2.1.1 LA BD Topo.....                                                                  | 22        |
| 2.2.1.2 LA BD Parcellaire .....                                                          | 23        |
| 2.2.1.3 La BD Contours...IRIS .....                                                      | 24        |
| 2.2.2 Traitement spatial .....                                                           | 25        |
| 2.2.2.1 Calcul d'intersection .....                                                      | 25        |
| 2.2.2.2 Agrégation spatiale.....                                                         | 26        |
| 2.2.2.3 utilisation d'un buffer .....                                                    | 26        |
| <b>3. ETAT DE L'ART .....</b>                                                            | <b>27</b> |
| 3.1 Données et Méthodologies employées au sein des AASQA.....                            | 28        |
| 3.1.1 Données employées .....                                                            | 28        |
| 3.1.2 Méthodologies employées .....                                                      | 28        |
| 3.1.2.1 Répartition uniforme de la population.....                                       | 29        |
| 3.1.2.2 Méthodologie BD Topo.....                                                        | 29        |
| 3.2 Autres Méthodologies.....                                                            | 32        |
| 3.2.1 Utilisation de l'occupation des sols .....                                         | 32        |
| 3.2.2 Utilisation des fichiers fonciers MAJIC.....                                       | 33        |
| 3.2.2.1 Observatoire National des Services de Communications Electroniques<br>.....      | 35        |
| 3.2.2.2 Observatoire régional des points noirs environnementaux de Rhône-<br>Alpes ..... | 35        |
| 3.2.3 Carroyage INSEE .....                                                              | 36        |
| <b>4. PROPOSITIONS METHODOLOGIQUES.....</b>                                              | <b>37</b> |
| 4.1 Recommandations sur les traitements et vocabulaire .....                             | 38        |
| 4.2 Correction de la BD contours... IRIS.....                                            | 39        |
| 4.2.1 Correction des codes IRIS.....                                                     | 40        |

|                                                                      |           |
|----------------------------------------------------------------------|-----------|
| 4.2.2 Correction de la superposition des IRIS .....                  | 40        |
| 4.2.3 Correction des contours communaux .....                        | 42        |
| 4.3 Correction des populations IRIS.....                             | 45        |
| 4.4 Echelle régionale.....                                           | 47        |
| 4.4.1 Les données utilisées .....                                    | 47        |
| 4.4.2 Méthodologie carroyage INSEE .....                             | 47        |
| 4.5 Echelle locale .....                                             | 49        |
| 4.5.1 Estimation de la surface habitable d'un bâtiment .....         | 50        |
| 4.5.2 Méthodologie MAJIC .....                                       | 50        |
| 4.5.2.1 Les données MAJIC .....                                      | 50        |
| 4.5.2.2 Etapes de la méthodologie .....                              | 52        |
| 4.5.3 Méthodologie BD topo .....                                     | 57        |
| 4.5.3.1 Les données utilisées.....                                   | 57        |
| 4.5.3.2 Etapes de la méthodologie .....                              | 58        |
| 4.6 Validation des méthodes .....                                    | 59        |
| 4.6.1 La méthodologie MAJIC .....                                    | 59        |
| 4.6.2 Les autres méthodologies.....                                  | 63        |
| <b>5. CONCLUSIONS ET PERSPECTIVES.....</b>                           | <b>68</b> |
| <b>6. REFERENCES .....</b>                                           | <b>70</b> |
| <b>7. LISTE DES ANNEXES .....</b>                                    | <b>71</b> |
| <b>CLASSES D'OCCUPATION DU SOL DE LA BASE CORINE LAND COVER ....</b> | <b>77</b> |

## REMERCIEMENTS

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

Le LCSQA remercie également la DGALN<sup>1</sup> 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 fournit de précieux conseils sur le contenu et l'utilisation de ces données.

Le LCSQA remercie enfin ATMO Champagne-Ardenne, AIR Lorraine, Air Normand, Air Rhône-Alpes, MADININAIR et LIGAIR. Ces AASQA ont testé activement la base de données de population issue de la méthodologie MAJIC sur leur territoire respectif.

---

<sup>1</sup> 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

## LISTE DES ABBREVIATIONS ET ACCRONYMES

|                |                                                                                                                                                                                                                                                                                                                    |
|----------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 Cadastre                                                                                                                                                                                                                                                                              |
| 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                                                                                                                |
| LAEA           | système de projection géographique (Lambert azimuthal equal-area)                                                                                                                                                                                                                                                  |
| 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<br>RFL : dispositif Revenus Fiscaux Localisés des ménages ( <a href="http://www.insee.fr/fr/methodes/default.asp?page=sources/sou-adm-rf-menages.htm">http://www.insee.fr/fr/methodes/default.asp?page=sources/sou-adm-rf-menages.htm</a> ) |
| RGE            | Référentiel à Grande Echelle ( <a href="http://professionnels.ign.fr/rge">http://professionnels.ign.fr/rge</a> )                                                                                                                                                                                                   |
| RGP            | Recensement Général de la Population                                                                                                                                                                                                                                                                               |
| SETRA          | Service d'Etudes sur les Transports, les Routes et leurs Aménagements                                                                                                                                                                                                                                              |
| SIG            | Système d'Information Géographique                                                                                                                                                                                                                                                                                 |

## RESUME

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 dans 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 actuellement au sein des AASQA et du LCSQA. La présente note s'attache spécifiquement au calcul de la répartition spatiale des populations pour lequel elle propose un guide méthodologique. Celui-ci développe une approche adaptée aux différentes résolutions spatiales contraintes par le contexte d'étude et aux données disponibles.

Deux nouvelles méthodes sont proposées: la méthodologie carroyage et la méthodologie MAJIC.

La méthodologie carroyage a été conçue pour exploiter les sorties de modèles régionaux des AASQA et la plateforme nationale PREV'AIR. PREV'AIR Urgence utilise déjà cette méthodologie pour évaluer des populations sujettes à des risques de dépassement.

La méthode MAJIC proposée par le CETE de Lyon et complétée par le LCSQA permet une description très fine de la population à une échelle locale. Elle emploie les données foncières MAJIC délivrées par la DGFIP. Du fait d'un manque de couverture géographique, cette méthodologie ne peut actuellement être appliquée sur l'ensemble du territoire français et doit, le cas échéant, être remplacée par la méthodologie basée sur la BD Topo. La méthodologie MAJIC nécessite une expertise locale forte des AASQA pour la valider. Des validations sont en cours avec des ASQAA volontaires.

Le LCSQA met à disposition des AASQA qui le souhaitent une extraction des données carroyées kilométriques issues de la méthodologie carroyage. Celles-ci sont mises à jour chaque année avec les nouvelles statistiques du recensement de la population.

Du fait, de contraintes sur le droit d'accès aux données MAJIC et une mise en œuvre complexe de la méthodologie, le LCSQA mettra à disposition des AASQA les données de population spatialisées issues de la méthodologie MAJIC. Des échanges seront nécessaires entre les AASQA et le LCSQA pour la validation et l'utilisation de ces données.

## SYNTHESE DE LA METHODOLOGIE

La synthèse présentée s'appuie sur du vocabulaire et des notions présentés dans la suite de ce document.

Le dispositif de surveillance de la qualité de l'air doit répondre à des problématiques différentes qui nécessitent des échelles d'étude plus ou moins fines : suivant le cycle de vie des polluants surveillés, une approche locale ou régionale est recommandée. Les méthodes de répartition spatiale de la population doivent s'adapter à ces exigences.

Les méthodes proposées dans le présent document s'appuient sur les méthodologies recensées au niveau national en conservant les meilleures techniques employées et incluent de nouveaux éléments lorsque des données complémentaires sont disponibles. Elles s'appliquent à être en cohérence avec le recensement officiel de la population en employant les statistiques de population au niveau de l'IRIS.

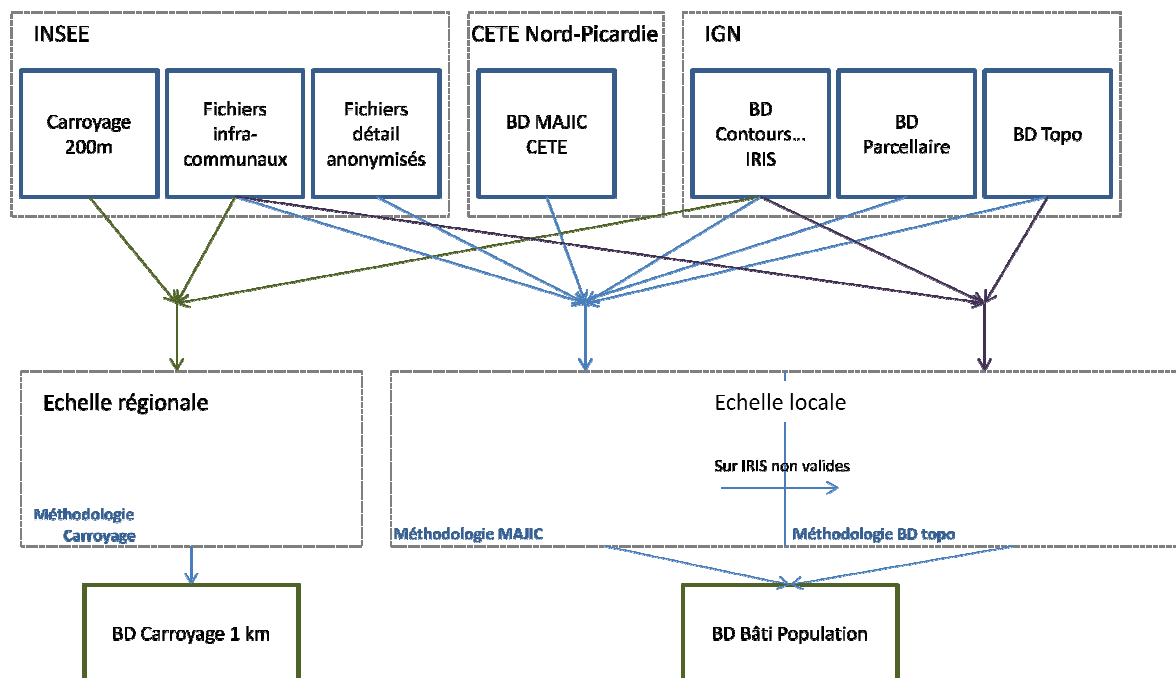


Figure 1 : Bases de données et méthodologies employées selon la résolution attendue

Toutes les méthodologies proposées emploient la BD Contours...IRIS. Celle-ci n'étant pas recommandée, selon l'IGN, pour une exploitation sur des résolutions fines, le LCSQA propose différents correctifs pour améliorer ses imperfections (cf 4.2).

## 1 - Echelle régionale : Méthodologie 'Carroyage'

Pour une évaluation à un niveau régional ou national, il n'est pas utile d'estimer la population dans chacun des bâtiments, l'utilisation d'un maillage régulier kilométrique (carroyage) sur la zone d'étude est suffisante. La méthodologie retenue pour cette échelle est une correction des données carroyées INSEE à partir des données du recensement. Cette méthodologie est appelée méthodologie 'Carroyage'.

Celle-ci utilise différentes sources de données :

- Un carroyage de population selon un pas de 200 m a été élaboré par l'INSEE. Celui-ci recense tous les foyers fiscaux.
- Les populations du recensement officiel à une échelle infra-communale (IRIS).
- Une base de données SIG Contours...IRIS co-produite par l'IGN et l'INSEE reproduit les contours des IRIS.

Les données de population délivrées dans le carroyage INSEE 200 m ne sont pas nécessairement en cohérence avec le recensement officiel des populations : certaines classes de population n'y sont pas recensées (personnes résidant dans des collectivités et certaines catégories ne sont pas situées au même endroit (étudiants majeurs par exemple).

La méthodologie s'applique à utiliser l'information de population du carroyage INSEE comme pondérateur spatial pour répartir les populations officielles du recensement sur le territoire.

$$Pr_{c,m,i} = Pr_i * \frac{P_{c,m,i}}{\sum_{d,n,i} P_{d,n,i}} \quad \text{Équation 1}$$

$Pr_{m,c,i}$  : population estimée du recensement dans le morceau m du carreau c de l'IRIS i

$Pr_i$  : population du recensement de l'IRIS i.

$P_{c,m,i}$  : population carroyée calculée sur le morceau m du carreau c de l'IRIS

Les populations corrigées des carreaux 200m sont ensuite agrégées sur des carreaux kilométriques.

Les détails de cette méthodologie sont présentés au chapitre 4.4.

## 2 - Echelle locale :

A une échelle locale, par contre, il est recommandé d'affiner les travaux de spatialisation en positionnant les habitants sur les bâtiments. La méthodologie MAJIC est retenue comme méthode de référence, elle associe les données MAJIC aux données du recensement. La méthodologie BD Topo est choisie comme méthode de substitution : elle est appliquée sur les zones où la méthodologie MAJIC n'est pas applicable.

Les deux méthodes utilisent la surface habitable des bâtiments pour y répartir la population. Cette surface pour un bâtiment est estimée par le produit entre sa surface au sol et son nombre d'étages.

Le nombre d'étages du bâtiment doit être au préalable établi à partir de sa hauteur, définie dans la couche « bâti indifférencié » de la BD Topo.

A défaut d'informations locales précises, les règles suivantes pourront être utilisées pour estimer le nombre d'étages :

- Un bâtiment d'une hauteur inférieure à 5 mètres ne possède qu'un étage.
- La hauteur des étages à partir du premier étage se situe entre 2.5 et 3 m. Le LCSQA a utilisé une hauteur de 3 m pour évaluer les méthodes.

## **2.1 - Méthodologie MAJIC**

Les fichiers fonciers MAJIC, délivrés par la DGFIP, 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.

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 désigne 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 CETE Nord-Picardie a en charge la récolte et la mise en forme des fichiers fonciers pour les autres services de l'écologie. Le CETE délivre les données MAJIC sous la forme d'une base de données PostgreSQL associée à une composante spatiale PostGIS. Ce format facilite grandement l'utilisation de ces informations car un lien direct est créé avec la BD Parcellaire de l'IGN et facilite leur exploitation dans les logiciels SIG. Cette base de données sera appelée BD MAJIC CETE.

La méthodologie MAJIC utilise diverses sources de données :

- La BD MAJIC CETE,
- Les populations du recensement officiel à une échelle infra-communale (Fichiers infra-communaux et fichiers détail anonymisés),
- LA BD Contour...IRIS,
- La BD Parcellaire,
- Et la BD Topo.

La méthodologie se décompose en différentes étapes :

- extraction et répartition des locaux à caractère d'habitation de la BD MAJIC CETE sur les bâtiments de la BD Parcellaire, en distinguant les habitations selon leur type d'habitat : maison ou appartement,
- cumul par IRIS du nombre de locaux MAJIC par type d'habitat,
- calcul du nombre de résidents par IRIS habitant dans des locaux de type maison et de type appartement à partir des données du recensement officiel INSEE,
- croisement des données MAJIC et INSEE par IRIS pour calculer sur chacun des bâtiments les populations résidentes pour chaque type d'habitat.

$$P_{b,i,t} = P_{i,t} * \frac{L_{b,i,t}}{\sum_a L_{a,i,t}} \quad \text{Équation 2}$$

$P_{b,i,t}$  : 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.

NB :  $L_{b,i,t}$  est pondéré par la somme des  $L_{a,i,t}$ .

Les détails de cette méthodologie sont présentés au chapitre 4.5.2.

La BD MAJIC CETE ne couvrant pas l'ensemble du territoire, il sera utile d'utiliser des critères de validation pour estimer si la méthodologie est valide sur chacun des IRIS étudié (cf. 4.6.1).

## **2.2 - Méthodologie BD Topo**

La méthodologie proposée s'attache à regrouper toutes les bonnes pratiques relevées au sein des AASQA.

La méthodologie BD Topo consiste à déterminer

- les bâtiments susceptibles d'accueillir de la population sur la zone d'étude,
- la surface ou le volume habitable de ces habitations,
- et enfin le nombre d'habitants dans chacun des bâtiments.

Différentes couches thématiques de la BD Topo sont utilisées pour définir les bâtiments accueillant de la population.

La couche 'bâti indifférencié' comportent tous les bâtiments à caractère d'habitation plus d'autres bâtiments à usage commercial ou administratif. Ces autres bâtiments vont être écartés en utilisant le thème 'Zone d'activité' de la BD Topo. Ce thème présente les points d'activité ou d'intérêt (PAI), qui localisent certains bâtiments ou sites ayant des caractères particuliers (administratif, religieux, sportif...). L'utilisation croisée de la couche 'bâti indifférencié' avec cette thématique permet d'écarter un certain nombre de constructions n'ayant pas pour usage l'habitation. Seules certaines classes sont retenues :

- La classe surfacique 'surface activité' répertorie les zones ayant une activité spécifique : administratif, enseignement, industriel ou commercial, cultures et loisirs,... Ces zones peuvent regrouper, par exemple, tous les bâtiments d'une zone d'activités tertiaire.
- Les classes ponctuelles représentent par un point certains bâtiments ayant une activité spécifique : science et enseignement, santé, religieux (rubrique culte), industriel ou commercial, gestion des eaux, cultures et loisirs et administratif et militaire.

Une validation humaine est préconisée lors du filtrage des bâtiments car certains bâtiments d'habitation importants peuvent être supprimés par erreur.

D'autres critères sont utilisés pour écartier certains bâtiments :

- Une surface au sol trop faible indique que le bâtiment n'est pas habitable (entre 20 et 50 m<sup>2</sup>),
- De même pour une hauteur de bâtiment inférieure à 3 mètres.

La population résidant dans un bâtiment est enfin estimée en multipliant la population totale de l'IRIS par le poids normalisé du bâtiment à l'échelle de l'IRIS (surface habitée du bâtiment/surface habitée totale des bâtiments dans l'IRIS) :

$$P_{b,i} = P_i * \frac{S_{b,i}}{\sum_a S_a} \quad \text{Équation 3}$$

- $P_{b,i}$  : population estimée dans le bâtiment  $b$  de l'IRIS  $i$
- $P_i$  : population de l'IRIS  $i$
- $S_{b,i}$  : surface habitable du bâtiment  $b$  de l'IRIS  $i$
- $S_{a,i}$  : surface habitable d'un bâtiment  $a$  de l'IRIS  $i$ .
- NB :  $S_{b,i}$  est pondéré par la somme des  $S_{a,i}$ .

Les détails de cette méthodologie sont présentés au chapitre 4.5.3.

## **1. INTRODUCTION**

La législation européenne sur l'évaluation de la qualité de l'air impose, lors du dépassement ou du risque de dépassement d'une valeur limite, de déterminer la zone de couverture du dépassement et la population impactée<sup>2</sup>. Les textes précisent également que les populations concernées sont les personnes résidant sur cette zone de dépassement<sup>3</sup>. Afin de répondre au mieux à ces exigences réglementaires, diverses méthodes sont actuellement employées au sein des AASQA.

Dans l'objectif d'harmoniser les méthodes et les données employées au niveau national et local, une étude est menée actuellement par le LCSQA sur l'évaluation des zones exposées à un dépassement de seuil d'une part et sur la répartition spatiale de la population d'autre part. Elle inclura également une réflexion sur la manière de croiser ces deux informations. L'objectif de la présente étude est de réaliser un état des lieux des méthodes employées en France pour cartographier les populations, analyser les différentes méthodes à disposition et proposer une harmonisation de celles-ci. Le LCSQA s'est attaché à proposer des méthodologies adaptées à différentes résolutions spatiales en fonction du contexte et du type de cartographie visé. Ces méthodologies s'appuient sur les études existantes menées par les AASQA mais également par d'autres organismes tels que les CETE<sup>4</sup>. Un séminaire LCSQA a présenté ces propositions méthodologiques aux AASQA en octobre 2012<sup>5</sup>.

Dans une première partie, la présente note décrit 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 deuxième partie présente un tour d'horizon des méthodologies employées en France. La dernière partie propose un guide méthodologique sur la répartition spatiale des populations à l'échelle régionale et à l'échelle locale. Les méthodologies proposées utilisent le recensement officiel de la population délivré par l'INSEE. Ce guide fournit également des recommandations sur l'utilisation de données SIG et des traitements cartographiques associés.

---

<sup>2</sup> Directive 2008/50/CE, Décret no 2010-1250 du 21 octobre 2010

<sup>3</sup> Décision 2011/850/EU

<sup>4</sup> Centres d'Etudes Techniques de l'Equipement

<sup>5</sup> <http://www.lcsqa.org/reunion/2012/traitements-numeriques/seminaire-lcsqa-estimation-repartition-spatiale-population>

## **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.

### **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, 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<sup>6</sup>, 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 population légale de chaque collectivité territoriale et de chaque circonscription administrative.

Mis en place en 2004, les premières statistiques ont été disponibles en 2009 pour l'année de référence 2006. Depuis cette date, chaque nouvelle année 'n' voit apparaître une nouvelle version établie sur 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.

---

<sup>6</sup> Loi n° 2002-276 du 27 février 2002

Le décret du 5 juin 2003<sup>7</sup> 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.

---

<sup>7</sup> Décret n° 2003-485 du 5 juin 2003 relatif au recensement de la population

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.

Le graphique Figure 2 présente les différentes catégories de population légale dans le recensement.

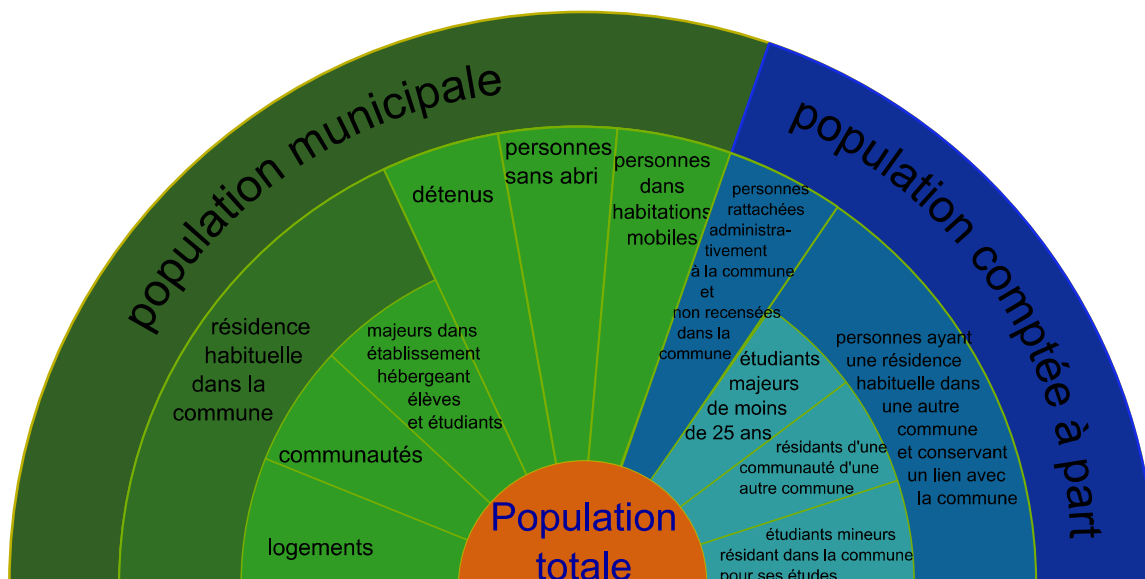


Figure 2 : catégories de population légale

### 2.1.2 LE DECOUPAGE 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. Ce sigle signifie « Ilots Regroupés pour l'Information Statistique » et qui 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 temps. 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.

- 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 RESULTATS 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 2010 sont publiés au début de l'année 2013. 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.recensement.insee.fr/basesInfracommunales.action>.

Il est à noter qu'il existe les fichiers détail anonymisés. Ceux-ci visent à fournir un complément d'informations 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/publics/default.asp?page=communication/recensement/particuliers/diffusion-fd.htm>

## **2.2 LES BASES DE DONNEES 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 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 REFERENTIEL A GRANDE ECHELLE (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 1<sup>er</sup> 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 est composé de quatre composantes qui fournissent des informations géographiques superposables entre elles : ortho photographique, topographique, parcellaire, adresse et altimétrie (Figure 3).

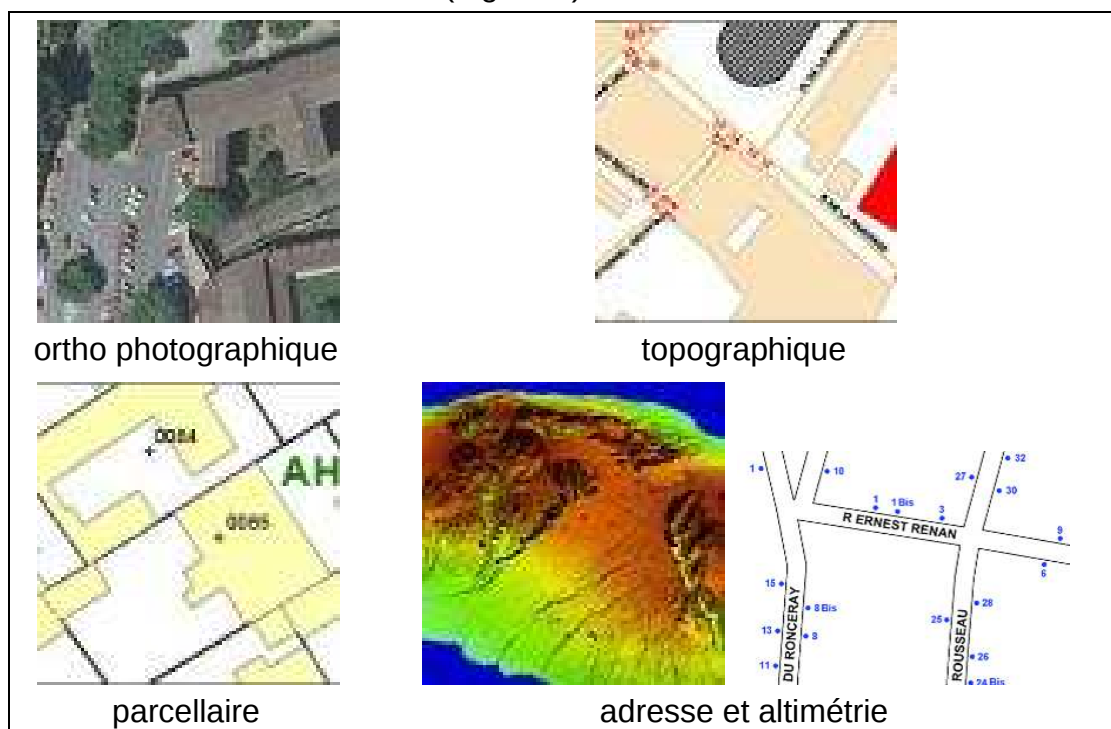


Figure 3 : Les composantes du RGE

Nous nous attachons à présenter dans ce document deux de ces composantes utiles pour la spatialisation des populations : la composante topographique représentée par la BD Topo et la composante parcellaire représentée par la BD Parcellaire.

#### 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 4). Elle est fournie dans un format vecteur. Cette base couvre la totalité du territoire français. Elle se décompose en différentes thématiques :

- réseau routier,
- voies ferrées et autres,
- transports énergie,

- hydrographies,
- bâti,
- végétation,
- orographie (lignes de rupture des pentes),
- administratif,
- zone d'activité,
- et toponymes.

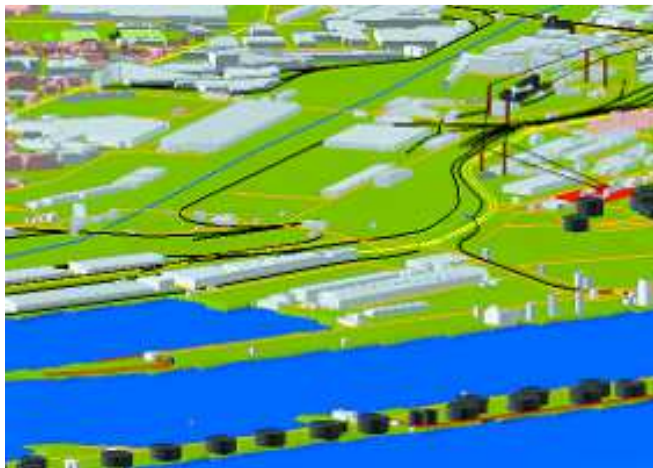


Figure 4 : représentation en 3D du paysage grâce à la BD Topo

#### 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 :

- Un 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 5 : plan cadastral présenté dans la BD Parcellaire Image de l'IGN

- Une couche d'objets vecteurs décrivant entre autre :
  - o chaque parcelle cadastrale (couche PARCELLE)

- o chaque bâtiment (couche bâtiment) en distinguant les bâtiments en dur (bâtiment industriel ou possédant des fondations et fermé sur les 4 côtés) et les constructions légères.



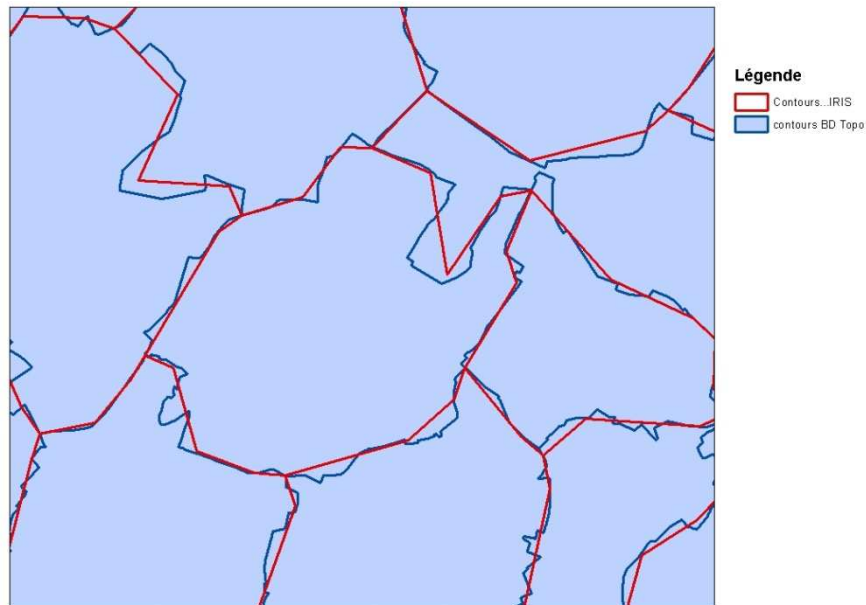
Figure 6 : plan cadastral présenté dans la BD Parcellaire Vecteur de l'IGN superposé avec la BD Ortho

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'agrandir cette couverture et dureront encore quelques années avant d'obtenir une couverture totale.

#### 2.2.1.3 LA BD CONTOURS...IRIS

Cette base de données, coproduite par l'INSEE et l'IGN, reproduit 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.

## **2.2.2 TRAITEMENT SPATIAL**

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

### **2.2.2.1 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 des carreaux INSEE avec les IRIS.

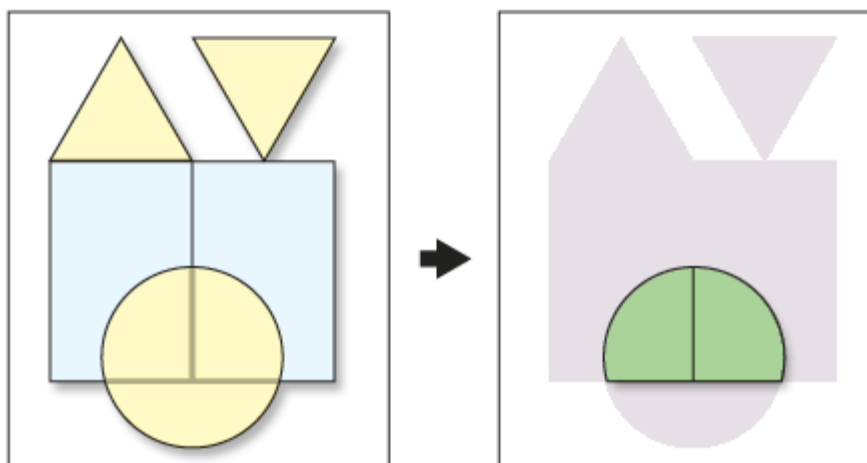


Figure 8 : A droite : résultat de l'intersection de couches géographiques de la partie gauche (une en bleu et l'autre en jaune)

#### 2.2.2.2 AGREGATION 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. Un exemple est présenté sur la Figure 9 :

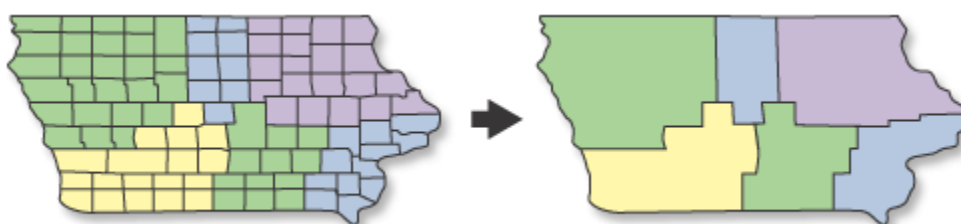


Figure 9 : Fusion d'entités géographiques en fonction d'un attribut commun

#### 2.2.2.3 UTILISATION D'UN BUFFER

L'application d'un zone tampon ou buffer sur des entités géographiques est souvent employée dans les traitements SIG pour rechercher des entités géographiques à proximité d'une entité de référence : 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 principalement utilisées pour rechercher des entités géographiques voisines d'une entité cible. Pour cela, le rayon de recherche utilisé est faible, de l'ordre du mètre.

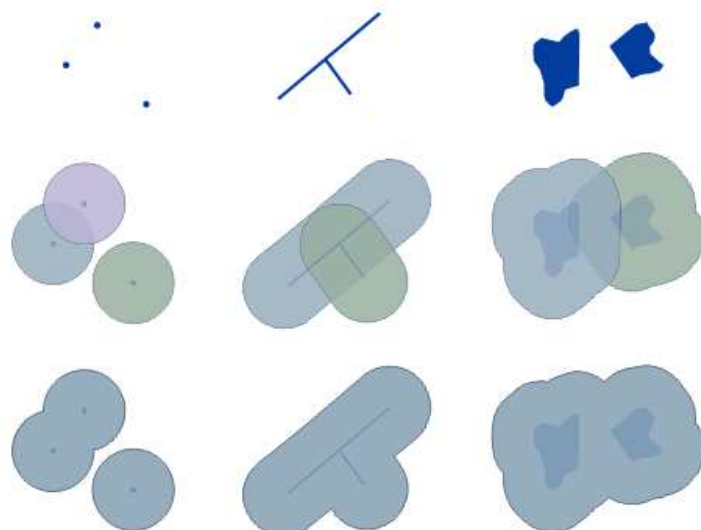


Figure 10 : application de zones tampon sur des entités ponctuelles, linéiques et surfaciques

### 3. ETAT DE L'ART

Dans l'objectif d'établir une méthodologie commune à toutes les AASQA et au LCSQA (pour le niveau national), il est utile en première étape d'inventorier les méthodologies employées par les AASQA et d'autres organismes en France et à l'étranger.

### **3.1 DONNEES ET METHODOLOGIES EMPLOYEES AU SEIN DES AASQA**

Une enquête a été menée par le LCSQA en 2012 auprès de toutes les AASQA pour évaluer les méthodes mises en œuvre. Les points abordés lors de cette consultation étaient :

- les bases de données disponibles : population et SIG,
- les années de référence des données de population utilisées,
- la ou les méthodologies employées à différentes échelles géographiques, pour répartir la population,
- quelques éléments sur les méthodes employées pour l'évaluation des populations exposées à un dépassement.

Ce dernier point n'est pas traité dans cette note, il avait juste pour objectif une meilleure compréhension des besoins des AASQA sur l'utilisation des données de population.

Il est à noter qu'un certain nombre d'AASQA (8) n'effectue pas actuellement de travaux de spatialisation de la population faute de moyens humains ou de compétences suffisantes en SIG.

#### **3.1.1 DONNEES EMPLOYEES**

Deux sources de données sont utilisées par les AASQA : les données de population issues de l'INSEE et des bases de données SIG servant de support géographique pour positionner la population.

Les données de population exploitées sont extraites des statistiques annuelles délivrées par l'INSEE. Selon la méthodologie employée, la résolution varie de l'échelle de l'IRIS à l'échelle communale. La mise à jour des données spatialisées de population peut pour certaines AASQA demander une durée de traitement non négligeable, l'année de référence de la population utilisée n'est donc pas nécessairement la plus récente.

Les bases de données spatiales exploitées sont :

- les contours communaux de la BD Topo ou BD Carto de l'IGN,
- les contours IRIS 1999 de l'INSEE ou les contours IRIS 2008 de la BD Contours...IRIS de l'IGN,
- certaines couches géographiques de la BD Topo (thématiques bâti et zone d'activité).

#### **3.1.2 METHODOLOGIES EMPLOYEES**

Les méthodes employées par les AASQA sont plus ou moins abouties suivant les régions. Une première méthode répartit uniformément les populations sur les

zones administratives (IRIS ou communes), une deuxième utilise la surface habitable des bâtiments comme pondérateur pour placer la population dans les habitations.

### 3.1.2.1 REPARTITION UNIFORME DE LA POPULATION

Cette première approche, utilisée par deux AASQA, suppose que les populations sont réparties uniformément sur l'ensemble du territoire d'une zone administrative. Cette technique exploite uniquement les contours communaux ou IRIS.

Un maillage régulier est constitué sur la zone d'étude. La population est ensuite répartie sur chacune des mailles en pondérant la population de la zone administrative (ZA) par la surface d'intersection maille/ZA divisée par la surface de la ZA qui la croise

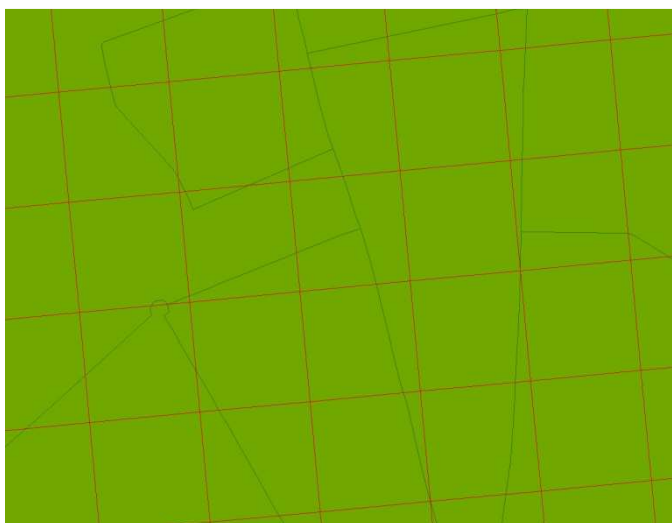


Figure 11 : Répartition uniforme de la population d'une zone administrative sur un maillage

$$P_m = \sum_i \frac{P_i * S_{m,i}}{S_i} \quad \text{Équation 4}$$

$P_m$  : population calculée dans la maille m

$P_i$  : population de la ZA i intersectant la maille

$S_{m,i}$  : surface d'intersection entre la maille m et la ZA i

$S_i$  : surface de ZA i

### 3.1.2.2 METHODOLOGIE BD TOPO

La deuxième approche, utilisée par les autres AASQA, croise la couche bâti de la BD Topo avec les contours IRIS. Cette méthode est pour partie documentée dans des fiches CERTU élaborées en 2005 et 2006. Par convention, ce type de méthode sera appelé méthodologie BD Topo dans la suite du document. Nous verrons par la suite que l'application de cette méthode peut être différente selon les AASQA. Cela peut induire des différences de répartition non négligeables.

La BD Topo est une base de données spatiale qui décrit la topographie et notamment les caractéristiques des bâtiments. Celle-ci se décompose en différentes thématiques dont les thèmes « Bâti » et « Zones d'activité ».

Le thème Bâti comporte 11 classes caractérisant la nature de certains bâtiments :

- bâtiments industriels, remarquables, ...
- bâtiments indifférenciés : cette classe regroupe tous les bâtiments dont le caractère n'a pas pu être décrit (logements, certains commerces, enseignement, ...) et qui possèdent une surface au sol supérieure à 20 m<sup>2</sup>.

La méthodologie BD Topo consiste à déterminer

- les bâtiments susceptibles d'accueillir de la population sur la zone d'étude,
- la surface ou le volume habitable de ces habitations,
- et enfin le nombre d'habitants dans chacun des bâtiments.

La méthodologie se décompose en plusieurs étapes :

- **Etape 1** : caractérisation de la nature des bâtiments

L'objectif de cette étape est de recenser les bâtiments susceptibles d'être des lieux d'habitation. Ce traitement peut être différent selon les AASQA :

- Tout le bâti indifférencié est retenu.
- Tous les bâtiments croisant un point d'activité d'intérêt ou une surface d'activité (couches PAI ponctuelles et couche SURFACE\_ACTIVITE du thème Zone d'activités de la BD Topo) sont retirés automatiquement.
- Des bases locales sont employées telles que les adresse des écoles ou les observatoires du bruit (travaux similaires déjà réalisés) pour supprimer certaines constructions.
- Des observations aériennes du type Géoportail, Google Maps ou Google Earth sont exploitées pour repérer les bâtiments à exclure (école ou administration). Cette tâche nécessite un temps de traitement relativement long car elle ne peut être automatisée. Elle est donc employée en général sur les agglomérations où une modélisation fine de la qualité de l'air est déployée.
- Certains bâtiments de surface faible (de 20 à 40 m<sup>2</sup>) sont supprimés car ceux-ci sont supposés être des bâtiments non appropriés pour un usage d'habitation.

- **Etape 2** : estimation de la surface d'habitation

La couche « bâtiment indifférencié » a la particularité de renseigner la hauteur des bâtiments. Elle permet donc d'estimer le volume ou la surface d'habitation des constructions. Différentes approches sont employées par les AASQA :

- o La hauteur du bâtiment n'est pas prise en compte : la surface habitable correspond à la surface au sol.
- o Le volume d'habitation du bâtiment est calculé directement par la formule : surface au sol multiplié par la hauteur du bâtiment.

- o Un nombre d'étages est déterminé pour évaluer la surface habitable du bâtiment :

- Tout bâtiment de hauteur inférieure à 2 ou 3 m est retiré. Ces bâtiments sont considérés comme étant des garages ou des dépendances.
- La BD Topo comporte un nombre non négligeable de bâtiments ayant une hauteur nulle. Pour ces habitations, leur hauteur est fixée à 3m pour le petit bâti, une observation satellitaire est utilisée pour les plus grandes surfaces.
- Le rez-de-chaussée des rues commerçantes est retiré.
- Les bâtiments de hauteur inférieure à 5 m possèdent un étage unique. Les étages suivants ont une hauteur variable entre 2m50 et 3 m.
- La surface habitable du bâtiment est calculée à partir de la formule suivante :

$$S_{hab} = \text{nombre d'étages} * \text{surface au sol} \quad \text{Équation 5}$$

- **Etape 3** : répartition de la population dans les bâtiments

La précédente étape a permis d'estimer la surface ou le volume habitable de chacun des bâtiments. Cette information est ensuite utilisée comme pondérateur pour ventiler la population de l'IRIS sur chaque habitation. La population d'un bâtiment est estimée en multipliant la population totale de l'IRIS par le poids normalisé du bâtiment à l'échelle de l'IRIS (surface du bâtiment/surface totale des bâtiments de l'IRIS) :

$$P_{b,i} = P_i * \frac{S_{b,i}}{\sum_a S_{a,i}} \quad \text{Équation 6}$$

$P_{b,i}$  : population estimée dans le bâtiment b de l'IRIS i

$P_i$  : population de l'IRIS i

$S_{b,i}$  : surface (ou volume) habitable du bâtiment b de l'IRIS i

$S_{a,i}$  : surface (ou volume) habitable d'un bâtiment a de l'IRIS i.

NB :  $S_{b,i}$  est pondéré par la somme des  $S_{a,i}$ .

- **Etape 4** (optionnelle) : Affectation de la population sur un maillage de restitution

Les sorties de modèles sont généralement disponibles sur des maillages réguliers plus ou moins fins : mailles de 20m à 100m de côté. Afin de travailler sur un même support géographique, certaines AASQA choisissent de projeter les données de population sur ces maillages.

La population d'une maille est alors égale à la somme des populations de chacun des bâtiments intersectant la maille pondérée par la surface d'intersection des bâtiments avec la maille :

$$P_m = \sum_{b,m} (P_{b,m} * \frac{SI_{b,m}}{S_{b,m}}) \quad \text{Équation 7 :}$$

$P_m$  : population estimée de la maille

$P_{b,m}$  : population du bâtiment b intersectant la maille m

$Sl_{b,m}$  : surface d'intersection entre le bâtiment b et la maille m

$S_{b,m}$  : surface au sol du bâtiment b intersectant la maille m

Il est à noter qu'Airparif utilise une base de données de population géo-localisée mise à disposition par l'IAU<sup>8</sup> d'Ile de France et n'a donc pas nécessité d'effectuer des travaux de répartition spatiale de la population. Ces données de population sont issues du recensement 2006. Cette population est recalculée aux niveaux des îlots INSEE 1999 (sous-partition des IRIS définis en 1999 mais abandonnée depuis par l'INSEE). Ce calcul de population tient compte de l'évolution de l'occupation des sols et de la proportion de logements inoccupés à l'îlot. Les populations sont ensuite réparties sur la couche bâtiment 2003 de l'IGN. La base d'occupation des sols de l'IAU permet de renseigner la destination du bâti (type d'habitats, ...). Une mise à jour de cette base était prévue en 2012. La population est ensuite projetée sur une grille de 50 m sur l'ensemble de l'Ile de France.

## 3.2 AUTRES METHODOLOGIES

D'autres méthodes plus ou moins fines sont exploitées en France et à l'étranger. Il n'est pas toujours aisé d'étudier les méthodologies employées dans d'autres pays car les bases de données de population ou les bases de données spatiales ne sont pas nécessairement en adéquation avec les données accessibles en France.

### 3.2.1 UTILISATION DE L'OCCUPATION DES SOLS

Une approche employée depuis de nombreuses années, que ce soit en France ou à l'étranger, s'appuie sur une information spatiale relative à l'occupation des sols. Cette information permet, à une résolution plus ou moins fine, de caractériser la nature des sols notamment la partie urbaine sur laquelle la population se situe principalement. Différentes méthodologies proposent de répartir différemment les populations selon le type d'occupation des sols. Ces méthodologies sont bien sûr très dépendantes de la qualité de l'information contenue dans ces bases de données spatiales.

Nous prenons ici pour exemple une méthodologie développée pour l'Agence Européenne de l'Environnement<sup>9</sup>. Celle-ci a été adaptée sur la France dans les travaux de thèse de Caudeville (2011). La méthodologie européenne utilisant les statistiques de population communale, Caudeville (2011) a adapté la méthode pour exploiter un découpage administratif plus fin à l'IRIS. Les données démographiques sont réparties sur le territoire national en utilisant la base de données Corine Land Cover (CLC). Celle-ci définit l'occupation du sol du territoire français au travers de 44 catégories.

---

<sup>8</sup> Institut d'Aménagement et d'Urbanisme

<sup>9</sup> Gallego F.J., 2009, A population density grid of the European Union, Submitted to Population and Environment. <http://www.eea.europa.eu/data-and-maps/data/population-density-disaggregated-with-corine-land-cover-2000-2>

Avant de procéder à une désagrégation des populations, les 44 postes (Cf. Annexe 2) sont regroupés en 10 classes (Tableau 1).

Tableau 1 : groupement des 44 postes de Corine Land Cover par Caudeville (2011)

| Groupement de classe                    | Codes Corine Land Cover   | k  |
|-----------------------------------------|---------------------------|----|
| Urbain continu                          | 1.1.1                     | 1  |
| Urbain discontinu                       | 1.1.2                     | 2  |
| Autres zones urbaines - Infrastructures | 1.3.3, 1.4                | 3  |
| Artificiel non résidentiel              | 1.2.2-1.2.4, 1.3.1, 1.3.2 | 4  |
| Zones agricoles                         | 2.1, 2.2, 2.3             | 5  |
| Hétérogène                              | 2.4.1-2.4.3               | 6  |
| Zones forestières                       | 2.4.4, 3.1                | 7  |
| Végétation naturelle                    | 3.2, 3.3                  | 8  |
| Zones industrielles et commerciales     | 1.2.1                     | 9  |
| Eau, Zones ouvertes                     | 3.4, 4, 5                 | 10 |

Un algorithme itératif a été développé pour estimer les coefficients de pondération de redistribution de la population à l'IRIS sur l'ensemble des classes de couverture de sol de la base de données CORINE Land Cover. Le résultat final de cette désagrégation est une grille kilométrique de densité de population sur l'ensemble du territoire français. L'étude est basée sur l'hypothèse que le ratio entre les densités de population de deux classes de couverture du sol est le même sur l'ensemble des IRIS étudiés.

Après la détermination des coefficients de pondération, la distribution spatiale de la population est calculée à partir de l'équation :

$$h_m = \sum_{i,k} \frac{h_i * u_k * S_{i,k,m}}{\sum_m S_{i,k,m} * u_k} \quad \text{Équation 8}$$

$h_m$  : nombre d'habitants dans la maille m

$h_i$  : nombre d'habitants dans l'IRIS i

$u_k$  : coefficient de pondération pour la classe Corine Land Cover k

$S_{i,k,m}$  : Surface d'intersection entre l'IRIS i, la classe k et la maille m

### 3.2.2 UTILISATION DES FICHIERS FONCIERS MAJIC

Hormis l'emploi de la BD Topo pour répartir les populations sur le bâti, des travaux ont été menés récemment par les CETE de l'Ouest et de Lyon pour affiner la localisation dans les communes.

Ces travaux exploitent les fichiers fonciers MAJIC 2 fournis par la DGFIP.

Les fichiers MAJIC 2 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.

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 depuis logements distincts.

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

Une convention a été signée entre la DGFIP et le Ministère en Charge de l'Ecologie sur la mise à disposition de ces fichiers pour la sphère Ecologie. Certaines de ces informations peuvent se révéler utiles pour localiser des données de population.

L'exploitation des fichiers fonciers a évolué ces dernières années grâce à l'amélioration des bases de données spatiales fournies par l'IGN. En effet, l'IGN propose la BD Parcellaire vecteur. Cette base de données fournit une information cadastrale numérique, géoréférencée sur le territoire français. Elle se décompose en deux groupes (cf. §2.2.1.2) :

- une couche de dalle « raster » de 1km de côté,
- et un groupe de couche d'objets 'vecteur', notamment les couches 'Parcelle' et 'Bâtiment'.

La couche vecteur 'Parcelle' de la BD Parcellaire permet par l'intermédiaire d'un identifiant unique de fournir une information géographique aux parcelles référencées dans les fichiers MAJIC. Toutefois, la BD Parcellaire vecteur ne couvrant pas intégralement le territoire, la localisation de ces parcelles n'est pas toujours réalisable.

Le CETE Nord-Picardie a en charge la récolte et la mise en forme des fichiers fonciers pour les autres services de l'écologie. A l'origine, les données foncières étaient délivrées sous la forme d'imposants fichiers 'texte' difficiles à traiter. Dorénavant, le CETE prétraite et délivre les données MAJIC sous la forme d'une base de données PostgreSQL associée à une composante spatiale PostGIS. Ce format facilite grandement l'utilisation de ces informations car un lien direct est créé avec la BD Parcellaire et facilite l'exploitation de ces données dans les logiciels SIG.

Deux exploitations des données MAJIC effectuées par les CETE sont présentées dans ce document.

### 3.2.2.1 OBSERVATOIRE NATIONAL DES SERVICES DE COMMUNICATIONS ELECTRONIQUES

Dans le cadre de l'Observatoire National des Services de Communications Electroniques, le CETE de l'Ouest avait pour mission de constituer une base nationale de locaux géolocalisés fiscalisés pour faciliter l'aménagement numérique raisonné du territoire.

Lors de la constitution de cette base, les données MAJIC n'étaient pas disponibles avec leur composante spatiale actuelle. Le CETE s'était donc employé à spatialiser les locaux en utilisant les coordonnées des centres de parcelles et les adresses renseignées pour ces locaux.

Les étapes principales du traitement étaient :

- calcul du nombre de locaux d'habitations et professionnels pour chaque adresse postale de chaque parcelle,
- géocodage des adresses : appariement des adresses MAJIC avec la BD Adresse de l'IGN. Cette base de données contient l'ensemble des adresses ponctuelles sur le territoire national (26 millions d'adresses).
- recherche du bâtiment d'accueil le plus probable dans la couche 'bâtiments indifférenciés' de la BD Topo.

Cette base de données ne fournit pas directement des chiffres de populations mais donne un nombre de foyers positionnés sur des points. Des traitements SIG complémentaires seraient nécessaires pour appliquer ces logements sur des bâtiments.

### 3.2.2.2 OBSERVATOIRE REGIONAL DES POINTS NOIRS ENVIRONNEMENTAUX DE RHONE-ALPES

L'observatoire points noirs environnementaux (PNE) de Rhône-Alpes a pour mission d'établir une cartographie de l'air et du bruit sur sa région de compétence, de déterminer des zones de co-expositions excessives. Dans ce cadre, le CETE de Lyon devait constituer une base de données spatiale recensant finement la population pour calculer un indicateur d'exposition.

La méthodologie 'BD Topo' ne répondait pas parfaitement à la problématique car celle-ci prend en compte un trop grand nombre de bâtiments n'ayant pas pour caractère l'habitation. Il a donc été développé une nouvelle méthodologie croisant la base de données MAJIC avec la couche 'bâtiment' de la BD Parcellaire et les données de population INSEE communales pour évaluer la population dans chacun des bâtiments.

Les principales étapes de cette méthodologie sont les suivantes :

- La couche 'bâtiment' de la BD Parcellaire ne renseigne pas la hauteur des bâtiments. Celle-ci est donc estimée en croisant la couche 'bâti' de la BD Topo et la couche 'bâtiment' Parcellaire'. Les mêmes règles que la méthodologie BD Topo (cf. §3.1.2.2, étape 2) sont exploitées pour corriger les hauteurs nulles de bâtiment et estimer ensuite le nombre d'étages et la surface d'habitation. Les bâtiments sont au préalable filtrés : on ne retient

que les bâtiments de type 'DUR'<sup>10</sup> et ayant une superficie au sol inférieure à 5 000 m<sup>2</sup><sup>11</sup>.

- Les statistiques INSEE fournissent le nombre d'habitants, le nombre de foyers, le nombre de logements vacants et le nombre de résidences secondaires par commune. Pour tenir compte des communes ayant un attrait touristique, les logements vacants et les résidences secondaires sont retranchés au nombre de foyers pour obtenir le nombre de foyers supposés abriter des résidents permanents. Un taux d'occupation des résidences principales (Tx) est ensuite calculé par commune :

$$Tx = \frac{\text{Population communale}}{\text{Nombre de foyers ayant des résidents permanents}} \quad \text{Équation 9}$$

- L'exploitation des données MAJIC fournit un nombre de locaux par parcelle. Ce nombre de locaux est multiplié par le taux d'occupation Tx pour obtenir un nombre de résidents par parcelle.
- Les parcelles MAJIC comportant des locaux de type habitation sont croisées avec la couche 'bâtiment' de la BD Parcellaire précédemment complétée pour calculer un nombre d'habitants par bâtiment au prorata de la surface d'habitation de chacun des bâtiments.

Le CETE de Lyon a comparé les chiffres de population obtenus au total par commune du recensement INSEE. Les résultats obtenus sont corrects : 70 % des communes présentent un écart compris entre -5 et +5 % et près de 90 % un taux compris entre -10 et +10 %. Les erreurs les plus importantes sont trouvées sur des communes à fort taux de vacances ou de résidences secondaires.

Comme indiqué précédemment, la BD Parcellaire ne couvre pas la totalité du territoire, le CETE a donc employé la méthode 'BD Topo' sur les zones parcellaires vides.

### 3.2.3 CARROYAGE INSEE

Depuis 2011, l'INSEE met à disposition en téléchargement sur son site internet un carroyage<sup>12</sup> de la population (nombre d'habitants par maille) à deux niveaux d'échelles : 200m et 1 km. Cette géolocalisation de population est effectuée à partir du croisement des données du RFL<sup>13</sup> (année 2009) et de la BD Parcellaire. Ces données sont disponibles uniquement sur la Métropole, la Martinique et la Réunion au format MapInfo.

Cette base de données SIG de population ne fournit qu'une vision partielle de la population car certaines populations n'y sont pas référencées. Les personnes sans domicile ou habitant dans des communautés ne sont pas retenues. Un écart méthodologique avec le recensement officiel apparaît principalement sur la

---

<sup>10</sup> Ne sont retenus que les bâtiments fermés des ayant des fondations

<sup>11</sup> Les bâtiments de surface supérieure à 5000 m<sup>2</sup> sont supposés être des constructions industrielles

<sup>12</sup> Technique de quadrillage utilisée en topographie, afin de rassembler et de traiter des données en vue d'une exploitation cartographique ou statistique. Il consiste à délimiter une surface en carrés identiques et localisés.

<sup>13</sup> RFL : dispositif des Revenus Fiscaux Localisés des ménages (<http://www.insee.fr/fr/methodes/default.asp?page=sources/sou-adm-rf-menages.htm>)

localisation des étudiants majeurs de moins de 25 ans. Ceux-ci sont généralement localisés au domicile de leurs parents alors que dans le recensement officiel ils sont localisés sur le lieu de leurs études. Ces différences créent un différentiel de 800 000 personnes environ entre les deux sources de données.



Figure 12 : Cartes des estimations carroyées de la population (résolution km) issues du RFL 2009 (source INSEE)

#### **4. PROPOSITIONS METHODOLOGIQUES**

Les chapitres précédents ont présenté différentes méthodologies pour répartir la population sur un territoire. Le dispositif de surveillance de la qualité de l'air doit répondre à des problématiques différentes qui nécessitent des échelles d'étude plus ou moins fines : suivant le cycle de vie des polluants surveillés, une approche locale ou régionale est recommandée. Les méthodes de répartition spatiale de la population doivent s'adapter à ces exigences.

Les méthodes proposées dans la suite du document s'appuient sur les méthodologies recensées en conservant les meilleures techniques employées et incluent de nouveaux éléments lorsque des données complémentaires sont disponibles. Elles s'appliquent à être en cohérence avec le recensement officiel de la population en utilisant les statistiques au niveau de l'IRIS.

Pour une évaluation à un niveau régional ou national, il n'est pas utile d'estimer la population dans chacun des bâtiments, l'utilisation d'un maillage régulier kilométrique sur la zone d'étude est suffisante. La méthodologie retenue pour cette échelle est une correction des données carroyées INSEE à partir des données du recensement.

A une échelle locale, par contre, il est recommandé d'affiner les travaux de spatialisation en positionnant les habitants sur les bâtiments. La méthodologie MAJIC est retenue comme méthode de référence, elle associe les données MAJIC aux données du recensement. La méthodologie BD Topo est choisie

comme méthode de substitution : elle est appliquée sur les zones où la méthodologie MAJIC n'est pas applicable.

Actuellement, il n'existe pas de base de données « exacte » de recensement de la population dans laquelle les personnes seraient géolocalisées parfaitement à leur domicile. Il n'est donc pas aisé de valider directement ces nouvelles méthodologies. Malgré cela, la méthodologie MAJIC ressort au niveau local comme pouvant être la méthode la plus adaptée car elle utilise les données les plus précises. Cette méthodologie sera donc employée comme repère pour évaluer les autres méthodes proposées.

#### **4.1 RECOMMANDATIONS SUR LES TRAITEMENTS ET VOCABULAIRE**

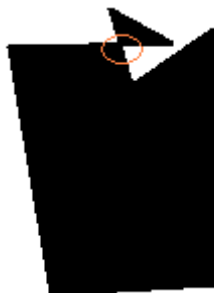
La mise en œuvre de ces méthodologies nécessite l'utilisation de logiciels SIG et pour certaines l'emploi de bases de données ayant une volumétrie conséquente.

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 nécessite de nombreuses étapes de traitement de vérifier au fil de ces étapes la conservation des surfaces ou des populations et valider les résultats finaux avec des bilans communaux, régionaux ou nationaux.

Cette partie s'attache à fournir quelques éléments pour aider à la validation des traitements à effectuer :

- 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 :
  - o 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 donc 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.

- o 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.
- o 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 fournis avec les logiciels SIG pour détecter et corriger ces erreurs.



*Figure 13 : auto-intersection sur un polygone (zone orange)*

- Il est nécessaire de travailler sur des couches SIG comportant le 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 les couches dans un référentiel unique avant d'appliquer des traitements sur ces données.

Dans l'enquête menée auprès des AASQA, certaines ont remonté la difficulté de mettre à jour les bases de données spatiales de population. En effet, lors de la création de ces bases, la future mise à jour de ces données n'est pas anticipée ou le travail associé est sous-évalué. Afin d'optimiser les ressources employées à la réalisation de bases spatiales de population, il est donc recommandé de faciliter la mise à jour des données avec les futurs millésimes du recensement de la population. Cette mise à jour pourrait être réalisée en conservant les ratios de calcul pour le passage de l'IRIS à un périmètre géographique plus fin (maille ou bâtiment) ou en conservant certaines étapes intermédiaires de calcul.

## **4.2 CORRECTION DE LA BD CONTOURS... IRIS**

Les données de population à l'IRIS étant utilisées, il est donc nécessaire d'employer une base de données spatiales des contours IRIS. Les bases à disposition sont la BD Contours...Iris et la BD Topo...Iris (cf. 2.2.1.3). La BD Topo...Iris, la plus fine en résolution, est la base recommandée par l'IGN pour une résolution fine. Celle-ci est onéreuse et ne couvre pas la totalité du territoire français ; le LCSQA a donc choisi d'utiliser la BD Contours...Iris. Le LCSQA propose des correctifs pour améliorer les imperfections de cette base, à savoir une résolution « grossière » et des superpositions d'IRIS.

#### 4.2.1 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>. 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.

#### 4.2.2 CORRECTION DE LA SUPERPOSITION DES IRIS

Lors de la première utilisation de cette base de données pour des traitements spatiaux, le LCSQA a constaté l'existence dans cette base de chevauchements de frontières entre des IRIS voisins. 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 14 : Chevauchement d'IRIS sur l'agglomération de LILLE

N'ayant pas 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.

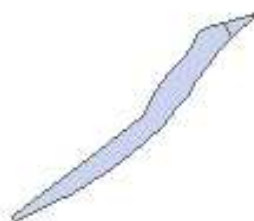


Figure 15 : Extraction des intersections entre deux IRIS

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

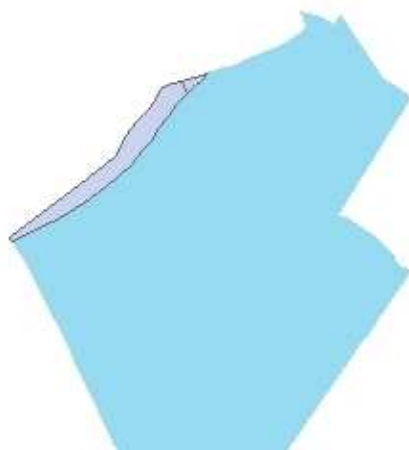


Figure 16 : Association du morceau d'intersection avec le premier IRIS rencontré par l'algorithme

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

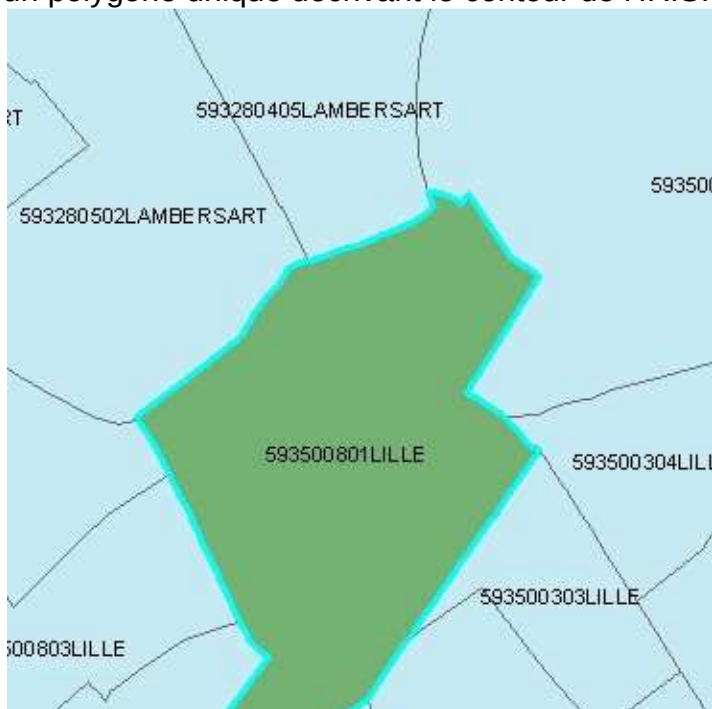


Figure 17 : Agrégation des morceaux d'IRIS pour obtenir le contour final de l'IRIS

### 4.2.3 CORRECTION DES CONTOURS COMMUNAUX

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). Une solution pour affiner la définition des contours IRIS est le remplacement des 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 20 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 restant (B).

Suite à la correction et agrégation des contours en fonction du code IRIS effectuée à l'étape 8, 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 (étape 8 à 11). La Figure 19 présente un exemple d'îlot (entouré en orange) séparé de la partie principale de l'IRIS colorée en rose. Cet îlot sera rattaché à l'îlot voisin (en bleu) appartenant à la même commune.

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

La Figure 18 présente un exemple de rendu de la correction des contours IRIS.

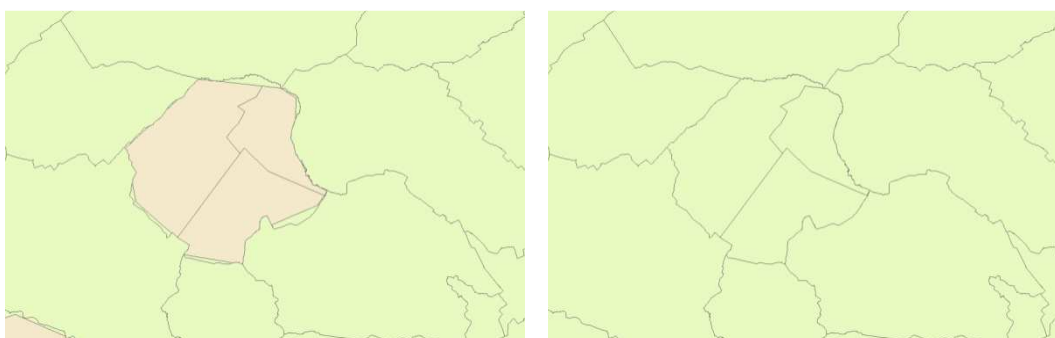
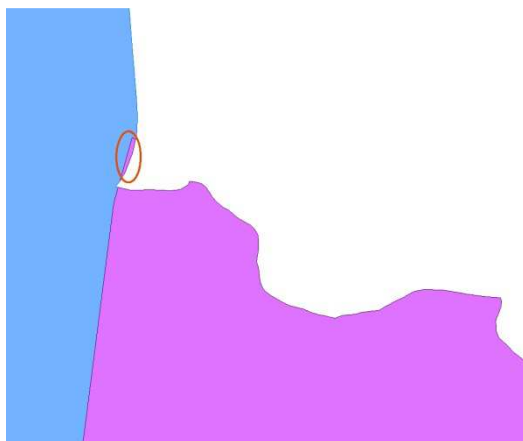


Figure 18 : A gauche, en rose, contour IRIS d'une commune avant correction ; à droite contours corrigés avec les contours communaux de la BD topo



*Figure 19 : Un îlot (entouré en orange) est séparé de l'IRIS principal coloré en rose*

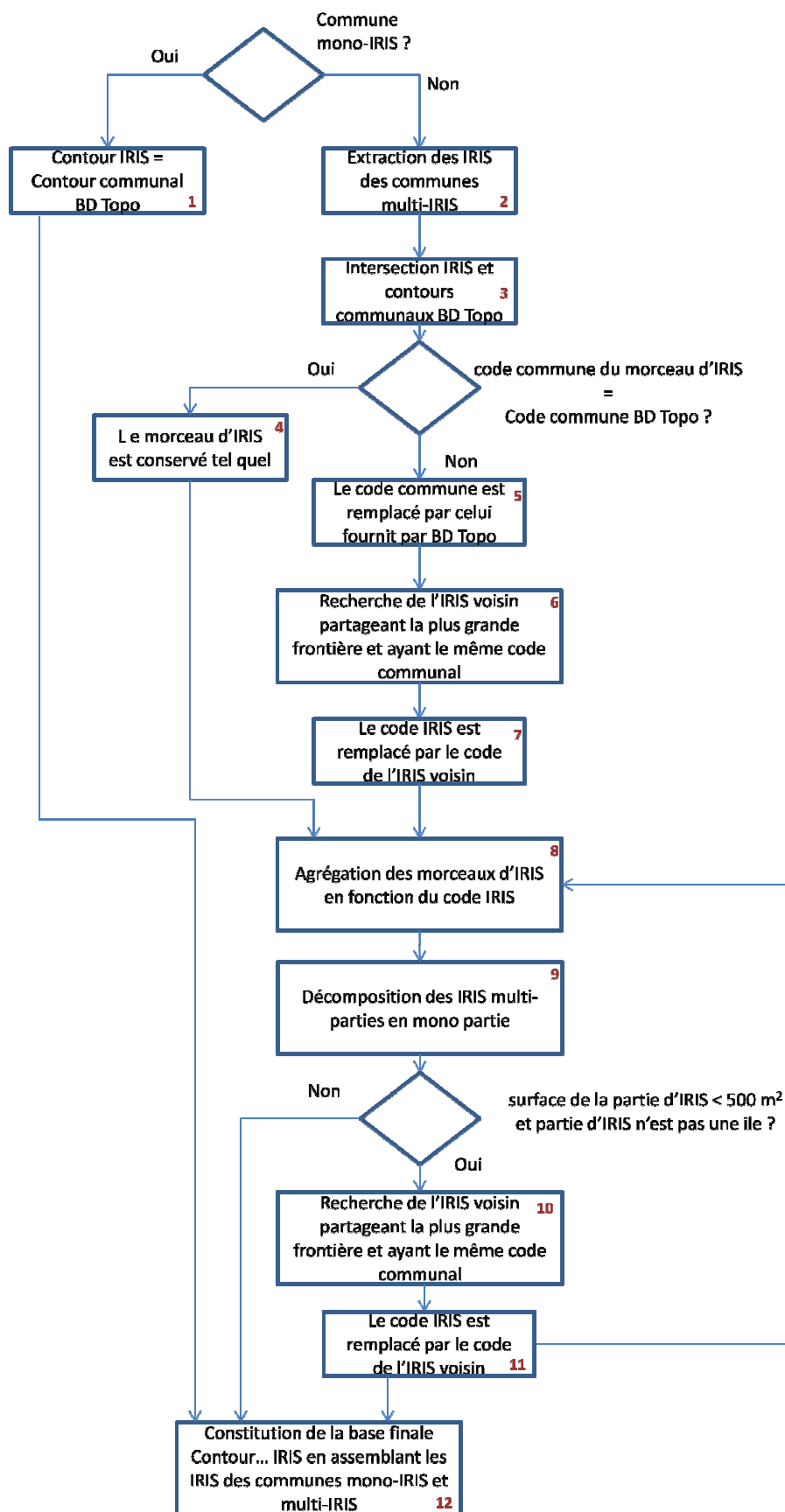


Figure 20 : Algorithme de correction des contours IRIS par les contours communaux de la BD Topo

### 4.3 CORRECTION DES POPULATIONS IRIS

L'INSEE délivre en début d'année le recensement communal des populations pour l'année n-3. Le recensement détaillé à l'IRIS est seulement délivré, quant à lui, lors du quatrième trimestre de la même année.

Afin de profiter du nouveau recensement communal, le LCSQA a étudié la possibilité de corriger en début d'année le recensement à l'IRIS à partir de l'évolution des populations communales n-3 et n-4. Pour cela, l'évolution des populations d'une commune est supposée analogue à celle des IRIS qui la constituent. Pour vérifier cette hypothèse, une simulation est réalisée sur l'évolution des populations entre deux années sur lesquelles les données du recensement sont déjà connues. Les tests portent sur les années 2008 et 2009. Le facteur d'évolution 2008/2009 est calculé pour chacune des communes françaises et est ensuite affecté sur les populations IRIS 2008.

Deux scénarios sont testés :

- les populations des IRIS évoluent à l'identique des communes,
- les populations restent constantes entre les deux années.

La comparaison est effectuée principalement sur l'écart relatif calculé entre la population 2009 et :

- la population estimée 2009,
- ou la population 2008.

Deux écarts relatifs différents sont calculés :

- l'écart relatif simple : ses statistiques tiennent compte du signe des écarts.

$$E = \frac{P - \hat{P}}{\hat{P}}$$

$E$  : écart relatif

$P$  : population mesurée

$\hat{P}$  : population estimée

- l'écart relatif absolu : ses statistiques sont indépendantes du signe.

$$E_a = |E|$$

$E_a$  : écart relatif absolu

Tableau 2 : Statistiques des écarts relatifs et des écarts relatifs absolus entre la population 2009 et les populations 2009 estimées et 2008

|         | Ecart relatifs        |           | Ecart relatifs absolus |           |
|---------|-----------------------|-----------|------------------------|-----------|
|         | 2009 - 2009<br>estimé | 2009-2008 | 2009 - 2009<br>estimé  | 2009-2008 |
| minimum | -129.9                | -130.1    | 0                      | 0         |
| maximum | 1                     | 1         | 129.9                  | 130.1     |
| moyenne | -0.005                | 0.002     | 0.015                  | 0.03      |

Les résultats des comparaisons entre les deux scénarios montrent que l'application d'un facteur correctif aux données par IRIS de l'année 2008 permet de réduire le biais moyen (Tableau 3), celui-ci devient nul.

De plus, cette correction recentre les écarts relatifs autour de 0 (Figure 21) et diminue les écarts relatifs absolus moyens (

Tableau 2

Tableau 3 : Biais moyen entre la population 2009 et les populations 2009 estimées (à gauche) et 2008 (à droite)

|             | 2009 - 2009<br>estimé | 2009-2008 |
|-------------|-----------------------|-----------|
| Corrélation | 0.999                 | 0.998     |
| Biais Moyen | 0.000                 | 6.7       |

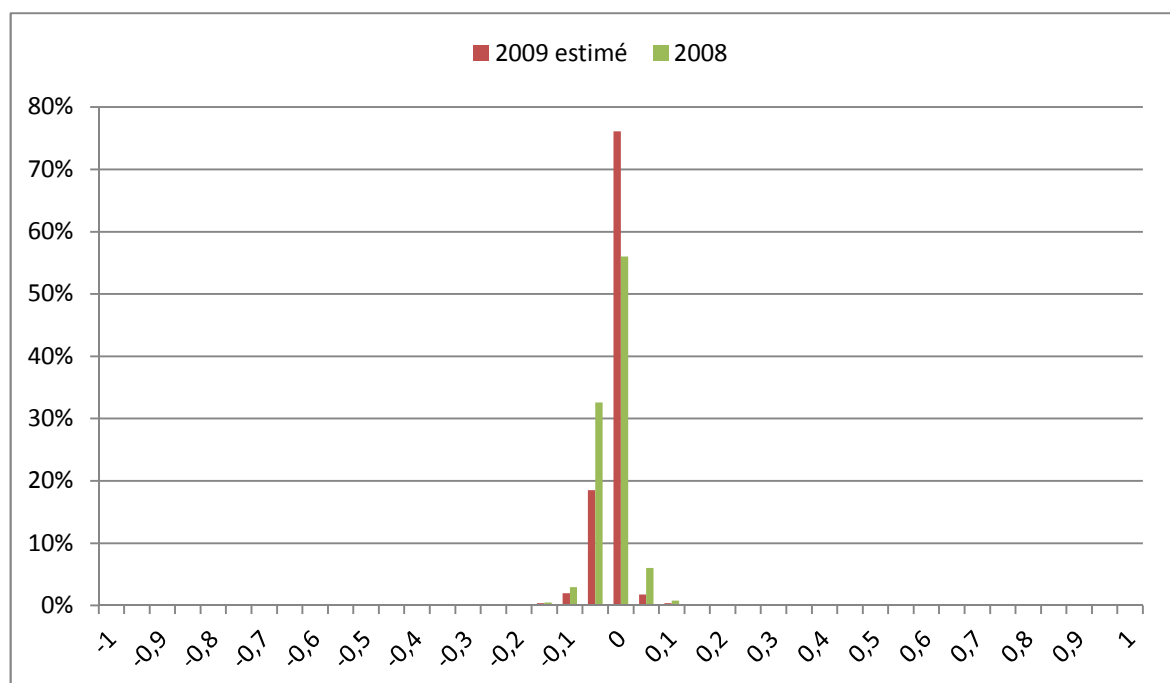


Figure 21 : Histogramme des écarts relatifs entre 2009 et 2009 estimé (en rouge) et 2008 (en vert).

Cette analyse justifie l'utilisation d'une correction des données à l'IRIS à l'aide des données communales plus récentes.

## 4.4 ECHELLE REGIONALE

### 4.4.1 LES DONNEES UTILISEES

Les données de population utilisées sont celles détaillées à l'IRIS, géoréférencées avec la BD contours...Iris et corrigées (cf. 4.2).

L'utilisation de données de population spatialisées à une échelle régionale ne nécessite pas une granularité fine. Une résolution kilométrique est convenable. L'INSEE ayant déjà créé un maillage régulier kilométrique sur la France, cette méthodologie utilisera le même maillage. Ce carroyage est défini dans le système de projection LAEA (code EPSG 3035) qui est conforme aux recommandations européennes de la directive INSPIRE (2007/2/CE).

### 4.4.2 METHODOLOGIE CARROYAGE INSEE

L'objectif est de constituer un recensement de la population agrégé au kilomètre sur l'ensemble du territoire. Une méthodologie simple est développée pour associer les données de population carroyées de l'INSEE avec les données du recensement de la population à l'IRIS.

Des écarts méthodologiques existent entre ces deux sources de données :

- certaines classes de population ne sont pas représentées dans le carroyage INSEE : les personnes sans domicile et les personnes habitant dans des collectivités ;
- certaines classes de population ne sont pas situées au même lieu d'habitation.

Le Tableau 4 présente les écarts méthodologiques. Un simple cumul sur la Métropole fait apparaître un manque de 800 000 habitants sur le carroyage INSEE.

Tableau 4 : Ecart méthodologique entre le carroyage INSEE et le recensement officiel de la population

|                                                                                                                              | Carroyage                                       | RGP                                             |
|------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------|-------------------------------------------------|
| <b>étudiants de moins de 25 ans<br/>habitant pour leurs études<br/>dans une autre commune<br/>que celle de leurs parents</b> | généralement<br>localisés chez leurs<br>parents | localisés sur<br>leur lieu<br>d'études          |
| <b>personne sans domicile</b>                                                                                                | non référencés                                  | rattachés à<br>une<br>commune                   |
| <b>collectivités</b>                                                                                                         | non référencés                                  | rattachés à la<br>commune de<br>la collectivité |

Du fait de ce manque important, les populations carroyées INSEE ne peuvent être utilisées comme référence pour localiser la population. Mais celles-ci sont probablement un bon indicateur de localisation des populations. Deux hypothèses différentes sont testées pour exploiter ces données carroyées :

- Hypothèse 1 : La donnée de population carroyée est utilisée comme pondérateur pour ventiler les populations à l'IRIS sur chacune des mailles.
- Hypothèse 2 : La donnée de population carroyée est uniquement un indicateur de présence de population sur la maille (0 ou 1). Les populations à l'IRIS sont réparties uniformément sur les mailles où une présence de population est relevée.

Les données carroyées sont mises à disposition par l'INSEE suivant deux résolutions : 200m et 1 km. Même si en général les IRIS couvrent une surface largement supérieure à 1 km<sup>2</sup>, les agglomérations où la densité de population est importante comportent des IRIS de superficie inférieure à 1 km<sup>2</sup>. Il est probablement plus judicieux d'utiliser les données carroyées à 200m pour une meilleure prise en compte de la dispersion spatiale de la population. Deux nouvelles hypothèses sont émises pour cette méthodologie :

- Hypothèse 3 : On utilise les données carroyées 200m. Les données de population calculées sur la grille de 200 m sont ensuite agrégées sur les mailles kilométriques.
- Hypothèse 4 : Les données carroyées 1 km sont employées comme pondérateur spatial de la population.

La Figure 22 présente les IRIS sur le quartier de Montmartre à Paris et démontre la nécessité d'employer un réseau maillé plus fin que le kilomètre.



Figure 22 : Iris et carroyages sur la zone de Montmartre : à gauche (en vert) carreaux de 1km de côté – à droite (en bleu) carreaux de 200 m de côté

Suite aux comparaisons (cf 4.6.2) effectuées entre les différentes hypothèses, les hypothèses 1 et 3 sont retenues dans la méthodologie

Les corrections sur le carroyage INSEE sont appliquées selon plusieurs étapes. Nous appellerons  $P_c$  la population initiale du carreau fourni par l'INSEE :

- **Etape 1** : Une intersection géographique est effectuée entre les contours IRIS et le carroyage 200m. Le résultat du processus est un ensemble de

polygones (morceaux) comportant chacun une double information numéro du carreau/code de l'IRIS.

- **Etape 2** : On répartit ensuite les populations  $P_c$  des mailles dans chaque morceau de maille en pondérant cette population par le ratio surfacique du morceau par rapport à la surface totale de la maille.

$$P_{c,m} = P_c * \frac{S_{c,m}}{\sum_n S_{c,n}} \quad \text{Équation 10}$$

$P_{c,m}$  : population carroyée calculée sur le morceau m du carreau c

$P_c$  : population carroyée du carreau c

$S_{c,m}$  : surface du morceau de carreau m du carreau c

$S_{c,n}$  : surface du morceau de carreau n du carreau c

- **Etape 3** : La population « carroyée » INSEE de chaque carreau est utilisée comme pondérateur pour appliquer une partie de la population de l'IRIS sur chacun des carreaux ou morceaux de carreaux contenus dans cet IRIS :

$$Pr_{c,m,i} = Pr_i * \frac{P_{c,m,i}}{\sum_{d,n} P_{d,n,i}} \quad \text{Équation 11}$$

$Pr_{c,m,i}$  : population estimée du recensement dans le morceau m du carreau c de l'IRIS i

$Pr_i$  : population du recensement de l'IRIS i.

$P_{c,m,i}$  : population carroyée calculée sur le morceau m du carreau c de l'IRIS i

$P_{d,n,i}$  : population carroyée calculée sur le morceau n du carreau d de l'IRIS i

- **Etape 4** : Cette étape cumule les populations des morceaux de carreaux pour reconstituer le carroyage 200m :

$$Pr_c = \sum_{m,i} Pr_{c,m,i}$$

$Pr_c$  : population estimée du recensement dans le carreau c

$Pr_{c,m,i}$  : population estimée du recensement dans le morceau m du carreau c de l'IRIS i

- **Etape 5** : Les populations des carreaux 200m sont enfin agrégées dans les carreaux kilométriques.

## 4.5 ECHELLE LOCALE

Comme évoqué précédemment, les données MAJIC apportent une information complémentaire très fine sur la répartition des locaux habités en France sur chaque parcelle cadastrale. Cette source de données n'étant pas exhaustive sur l'ensemble du territoire, il est essentiel de conserver une méthodologie de remplacement sur les zones non couvertes. La méthodologie BD Topo sera utilisée sur ces zones dites blanches.

L'estimation de la surface habitable des bâtiments étant commune à ces deux méthodes, celle-ci est décrite en premier.

#### **4.5.1 ESTIMATION DE LA SURFACE HABITABLE D'UN BATIMENT**

Au vu des différentes méthodes employées en France, l'utilisation du nombre d'étages et de la surface au sol du bâtiment paraît le moyen le plus adapté pour déterminer une surface habitable.

La couche « bâti indifférencié » de la BD Topo possède un champ numérique qui caractérise la hauteur de chaque bâtiment. Cette hauteur est utilisée pour déterminer le nombre d'étages d'un bâtiment.

La définition de hauteurs « type » par étage est délicate, ces hauteurs sont dépendantes de nombreux paramètres : région, zone rurale ou urbaine, âge de l'habitat difficiles à caractériser... A défaut d'informations précises, des observations par exemple, les hauteurs suivantes pourront être utilisées :

- Un bâtiment d'une hauteur inférieure à 5 mètres ne possède qu'un étage.
- La hauteur des étages à partir du premier étage se situe entre 2.5 et 3 m. Le LCSQA a utilisé une hauteur de 3 m pour évaluer les méthodes.

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 peut être fixée à un étage, pour ceux présentant une surface plus importante, une observation humaine est conseillée.

#### **4.5.2 METHODOLOGIE MAJIC**

La méthodologie MAJIC présentée dans ce document s'appuie sur les travaux menés par le CETE de Lyon dans le cadre de l'observatoire des PNE de la région Rhône-Alpes. Certaines limites ont été relevées dans l'étude du CETE :

- Aucun bouclage n'est réalisé sur les populations du recensement. De ce fait, l'utilisation de ratios moyens par habitation crée un biais lors du calcul des populations.
- L'affectation de certains locaux sur les parcelles peut être améliorée en utilisant des traitements SIG complémentaires.
- La méthode ne distingue pas les locaux selon leur type d'habitat, maison ou appartement.

Cette méthodologie utilise différentes sources de données : fichiers MAJIC, BD Parcellaire, BD Topo, Contours...IRIS corrigés et données du recensement INSEE.

##### **4.5.2.1 LES DONNEES MAJIC**

Le CETE Nord-Picardie met à disposition les données foncières sous la forme de base 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.

Ces données sont constituées en 18 tables. Seulement deux sont utilisées pour la répartition spatiale de la population :

- Une première table fournit des informations sur les parcelles :
  - o Le champ idpar fournit l'identifiant unique de la parcelle.
  - o Le champ idprocpte donne l'identifiant du propriétaire.
  - o Le champ idparef donne l'identifiant de la parcelle de référence.
  - o Le champ nbat indique le nombre de locaux référencés sur la parcelle.
  - o Le champ cmp indique si la parcelle fait partie d'une copropriété multi-parcelles (valeur 1).
  - o Le champ vecteur indique si la parcelle est vectorisée (valeur 'V').
  - o Le champ source\_geo indique la source de la géométrie (BD Parcellaire + année ou Cadastre.gouv.fr + année).
  - o La géométrie de la parcelle si elle existe.
- Une seconde table fournit des informations sur les locaux :
  - o Le champ idpar fournit l'identifiant de la parcelle sur laquelle se trouve le local.
  - o Le champ dteloc fournit le type du local : si 1 local de type habitation maison – si 2 local de type habitation appartement.
  - o Le champ stoth indique la surface totale des pièces d'habitation.

Avant l'exploitation de la BD MAJIC CETE, il est utile de d'analyser la constitution de ces données 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 (champ cmp =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 23 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 101. Il est nécessaire de prendre en compte cette particularité en répartissant les locaux référencés dans la parcelle de référence sur l'ensemble des parcelles constituant la copropriété.



*Figure 23 : 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. 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 sera 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 sera traité dans la méthodologie.

#### 4.5.2.2 ETAPES DE LA METHODOLOGIE

La méthodologie MAJIC se décompose en plusieurs étapes :

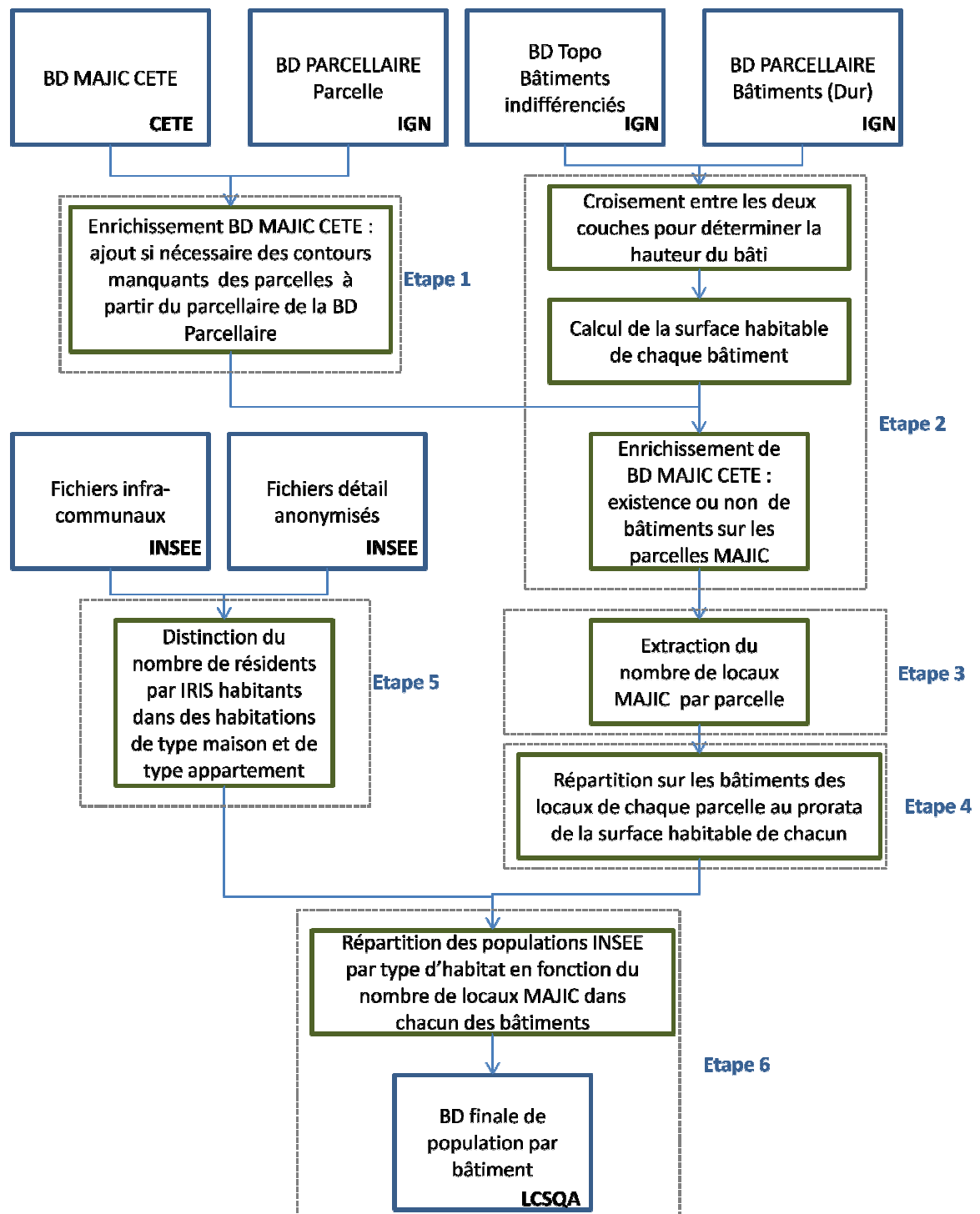


Figure 24 : présentation des différentes étapes de la méthodologie MAJIC

- **Etape 1 :** Lors de la constitution de la BD MAJIC CETE, le CETE a exploité les fichiers fonciers DGFIP et la BD Parcellaire disponible en 2011. Lors de l'application de la méthodologie MAJIC de spatialisation de la population, il est possible que la BD Parcellaire disponible soit plus récente que celle employée par le CETE et possède alors des parcelles vectorisées 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.
- **Etape 2 :** Nous avons vu dans les cas particuliers du 4.5.2.1 qu'il serait utile de recenser les parcelles possédant des bâtiments.

Les bases BD Parcellaire et BD Topo sont exploitées pour caractériser les bâtiments et déterminer les parcelles contenant des bâtiments :

- o Ne sont retenus dans la couche 'bâtiment' de la BD Parcellaire que les bâtiments en dur.
  - o Afin d'exclure les bâtiments à caractère industriel, tous les bâtiments de surface au sol inférieure à 20 000 m<sup>2</sup> sont retenus.
  - o La couche 'bâtiment' ne possède pas d'information sur la hauteur des bâtiments, elle est donc croisée avec la BD Topo pour obtenir la hauteur de ces bâtiments.
  - o Le nombre d'étages et la surface habitable sont ensuite calculés pour chacun des bâtiments (cf 4.5.1).
  - o Les bâtiments sont ensuite croisés avec la couche 'Parcelles' de la BD MAJIC CETE pour ajouter des informations complémentaires aux parcelles : existence ou non de bâtiments sur la parcelle, surface habitable cumulée sur la parcelle.
- **Etape 3 :** Les deux précédentes étapes ont enrichi de nouvelles informations la BD MAJIC CETE. Cette base est exploitée pour positionner les locaux à caractère d'habitation sur chacune des parcelles. La Figure 25 présente l'algorithme qui affecte les locaux sur les parcelles. Les cas particuliers présentés en 4.5.2.1 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, la parcelle est alors considérée comme un bâtiment unique d'un étage. 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 comptées séparément. Lors des prochaines étapes, cette distinction est conservée.

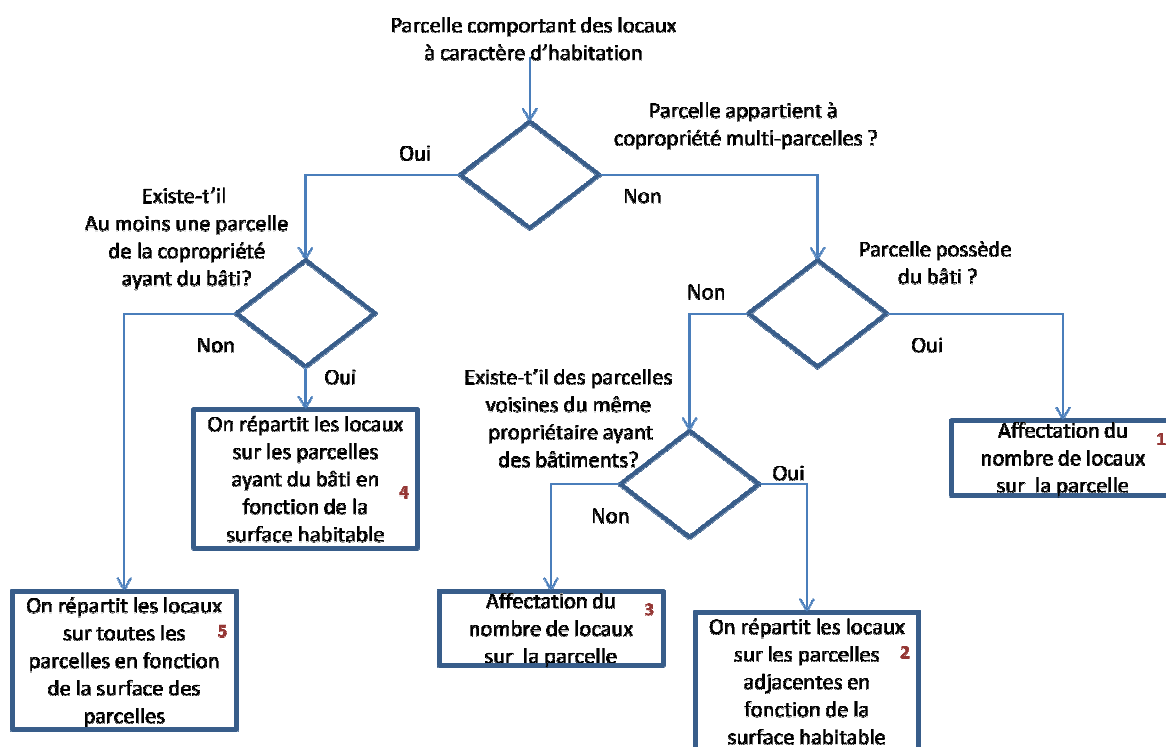


Figure 25 : Algorithme d'affectation des locaux dans les parcelles

- Etape 4 :** Suite à l'étape précédente, les locaux sont ventilés sur les bâtiments des parcelles retenues. Si une parcelle ne contient pas de bâtiment BD Parcellaire mais possède des locaux MAJIC, cette parcelle sera considérée comme étant un bâtiment unique d'un étage. On crée alors un bâtiment unique sur toute la surface de la parcelle pour que la population associée à cette parcelle puisse être spatialisée. 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 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 bâtiment contenant le nombre de locaux. Nous appellerons cette base BD Locaux.
- Etape 5 :** Avant de répartir les populations INSEE dans les bâtiments, il est nécessaire de calculer sur chacun des IRIS le nombre d'habitants résidant dans des habitations de type maison ou de type appartement. Pour cela, l'utilisation croisée des fichiers infra-communaux et des fichiers détail anonymisés de l'INSEE permet d'obtenir ce type de 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 ne comptabilisent uniquement que les populations des ménages. Ils ne comprennent pas les personnes vivant dans des habitations mobiles, les mariniers, les sans-abris et les personnes vivant en communauté. Le cumul des populations par IRIS

et par type de logement (maison ou appartement) permet de connaître le nombre d'habitants et le ratio associé à chaque type de logement par IRIS.

- Les ratios obtenus précédemment ne sont calculés que sur les populations des ménages : faute d'informations complémentaires disponibles, ces ratios sont appliqués sur les populations municipales (population des ménages + population hors ménages) pour obtenir un nombre d'habitants par IRIS résidant dans des logements de type maison ou de type appartement.

- **Etape 6 :** Suite à la spatialisation du nombre de locaux MAJIC, ces données sont croisées avec les données de population du recensement à l'IRIS. Pour cela, la couche BD Locaux est intersectée avec la couche Contours...IRIS.

Certains bâtiments se trouvent à cheval sur deux IRIS, il est donc nécessaire de recalculer le nombre de locaux MAJIC sur chacun de ces morceaux de bâtiments. Ce calcul est un calcul classique d'agrégation par ratio surfacique où les locaux sont répartis uniformément sur les morceaux des bâtiments en fonction de leur surface au sol.

Le nombre de locaux MAJIC contenus dans les bâtiments est ensuite cumulé par IRIS pour estimer sur chaque bâtiment le ratio du nombre de locaux sur le total à l'IRIS.

Le nombre de personnes habitant des habitations de type maison et de type appartement dans un IRIS est ainsi connu. Ce nombre de personnes est réparti sur tous les bâtiments contenus dans l'IRIS au prorata du nombre de locaux MAJIC de chaque bâtiment :

$$P_{b,i,t} = P_{i,t} * \frac{L_{b,i,t}}{\sum_a L_{a,i,t}} \quad \text{Équation 12}$$

$P_{b,i,t}$  : 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.

NB :  $L_{b,i,t}$  est pondéré par la somme des  $L_{a,i,t}$ .

Cette étape permet de constituer la base de données finale de population répartie sur chacun des bâtiments. Cette base sera appelée BD Bâti Population.

### 4.5.3 METHODOLOGIE BD TOPO

La méthodologie MAJIC ne peut encore s'appliquer à toutes les zones françaises du fait de la couverture spatiale incomplète des données foncières MAJIC. La méthodologie BD Topo peut pallier ce manque et être utilisée sur les zones concernées.

La méthodologie proposée s'attache à regrouper toutes les bonnes pratiques relevées au sein des AASQA. Cette méthodologie a été appliquée par le LCSQA sur le département d'Indre et Loire.

#### 4.5.3.1 LES DONNEES UTILISEES

La méthodologie BD Topo proposée s'appuie sur deux thématiques de la BD Topo : le thème 'Bâti' et le thème 'Zone d'activité'.

La première de ces thématiques regroupe toutes les classes répertoriant les bâtiments. La classe 'bâti indifférencié' décrit les bâtiments qui ne possèdent pas de fonction particulière décrite dans les autres classes. Cette classe intègre tous les bâtiments de type habitation mais également d'autres secteurs comme l'enseignement, les hôpitaux, les bâtiments administratifs. A la différence de la BD Parcellaire, un polygone de la BD Topo ne va pas nécessairement représenter un bâtiment unique mais parfois un groupe de bâtiments quand ceux-ci sont collés les uns aux autres. Ceci crée une difficulté pour distinguer la partie administrative ou tertiaire de la partie habitation d'un bâtiment. La Figure 26 présente un exemple d'agrégat de bâtiments sur une commune rurale de la Marne. Ces agrégats sont couramment rencontrés dans les zones de construction dense.



Figure 26 : agrégat de bâtiments dans la BD Topo

La deuxième thématique présente les points d'activité ou d'intérêt (PAI), qui localisent certains bâtiments ou sites ayant des caractères particuliers (administratif, religieux, sportif...). L'utilisation croisée de la couche 'bâti

indifférencié' avec cette thématique permet d'écarter un certain nombre de constructions n'ayant pas pour usage l'habitation. Seules certaines classes sont retenues :

- La classe surfacique 'surface activité' répertorie les zones ayant une activité spécifique : administratif, enseignement, industriel ou commercial, cultures et loisirs,... Ces zones peuvent regrouper, par exemple, tous les bâtiments d'une zone d'activités tertiaire.
- Les classes ponctuelles représentent par un point certains bâtiments ayant une activité spécifique : science et enseignement, santé, religieux (rubrique culte), industriel ou commercial, gestion des eaux, cultures et loisirs et administratif et militaire.

#### 4.5.3.2 ETAPES DE LA METHODOLOGIE

La méthodologie BD Topo s'articule en différentes étapes : sélection des bâtiments d'habitation, calcul de la surface d'habitation de ces bâtiments puis répartition de la population dans les habitats au prorata de leurs surfaces.

- **Etape 1 :** Cette étape définit les bâtiments accueillant la population. L'utilisation croisée de la classe 'bâti indifférencié' avec la thématique PAI facilite la sélection des bâtiments à exclure. Les classes ponctuelles sélectionnent des bâtiments ou agrégats de bâtiments unitaires. La classe surfacique 'surface activité' recouvre des zones et permet ainsi d'éliminer des groupes de bâtiments. Suite aux essais effectués sur le département d'Indre et Loire, cette couche paraît en général plus précise que certaines couches ponctuelles car elle paraît délimiter parfaitement certaines zones d'activité et de découper ainsi les bâtiments en fonction de leur utilisation.

Une validation humaine est tout de même préconisée sur le filtrage des bâtiments car les classes présentées ne sont pas exemptes d'erreurs. Des bâtiments d'habitation importants peuvent être supprimés par erreur. La Figure 27 présente un bâtiment d'habitation exclu par erreur.

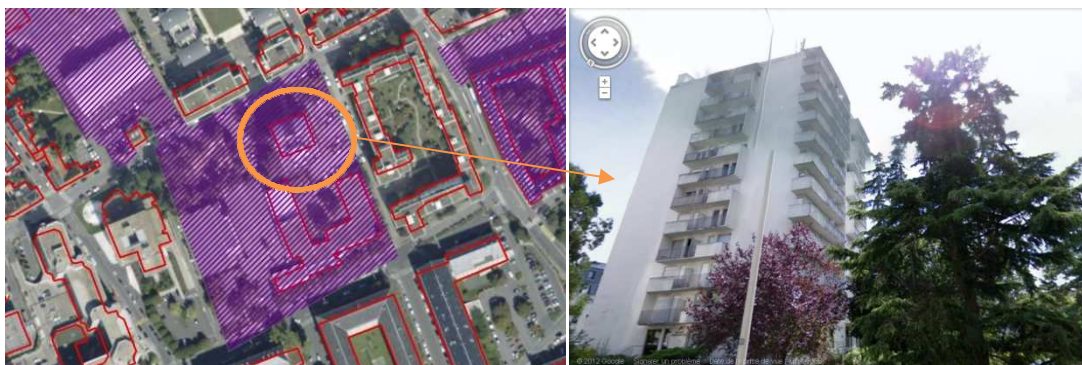


Figure 27 : A gauche : en violet, zone d'activité de type 'enseignement', entouré en orange, un immeuble d'habitation sélectionné par erreur - A droite, vue de la rue du bâtiment d'habitation supprimé.

Suite à l'application d'un premier filtre sur les bâtiments, deux caractéristiques sont utilisées pour exclure d'autres constructions : leur surface au sol et leur hauteur. En effet dans les différentes méthodes employées en France, une

surface trop faible ou une hauteur trop basse signifie généralement que le bâtiment ne peut être habitable. Pour qu'un bâtiment soit considéré comme habitable, ces méthodes définissent un seuil entre 20 et 50 m<sup>2</sup> pour la surface et une hauteur minimale entre 3 et 4 m.

Des essais ont été menés pour comparer ces seuils en utilisant une validation visuelle sur des photographies aériennes. Les résultats ne permettent pas de conclure sur la surface minimale à retenir. Celle-ci est vraisemblablement dépendante du milieu régional voir même local.

Par contre, ces tests indiquent que l'emploi d'une hauteur minimale de 4 mètres élimine un nombre non négligeable de maisons individuelles. Une valeur minimale de 3 mètres est en conséquence préconisée.

- **Etape 2** : Le nombre d'étages et la surface habitable sont estimés sur chacun des bâtiments (cf 4.5.1).
- **Etape 3** : La population d'un bâtiment est estimée en multipliant la population totale de l'IRIS par le poids normalisé du bâtiment à l'échelle de l'IRIS (surface habitée du bâtiment/surface habitée totale des bâtiments dans l'IRIS) :

$$P_{b,i} = P_i * \frac{S_{b,i}}{\sum_a S_a} \quad \text{Équation 13}$$

- $P_{b,i}$  : population estimée dans le bâtiment  $b$  de l'IRIS  $i$
- $P_i$  : population de l'IRIS  $i$
- $S_{b,i}$  : surface habitable du bâtiment  $b$  de l'IRIS  $i$
- $S_{a,i}$  : surface habitable d'un bâtiment  $a$  de l'IRIS  $i$ .
- NB :  $S_{b,i}$  est pondéré par la somme des  $S_{a,i}$ .

## 4.6 VALIDATION DES METHODES

### 4.6.1 LA METHODOLOGIE MAJIC

Le LCSQA a mis en œuvre et automatisé cette méthodologie sur plusieurs régions. Ces travaux ont montré la grande inégalité qui existe entre les départements sur la couverture de la BD Parcellaire. Même si cette méthode est plus précise que la méthodologie BD Topo, elle n'est pas exempte d'incertitudes. En effet, certaines parcelles n'étant pas vectorisées, la spatialisation des locaux attachés n'est pas appliquée. Des zones blanches apparaissent donc à l'intérieur de certains IRIS. De plus, les données des locaux MAJIC n'ayant pas la même source que celles du recensement INSEE, des écarts dus à cette différence peuvent apparaître.

Des comparaisons ont été menées entre les données des locaux INSEE et MAJIC. Les Figure 28 et Figure 29 présentent un exemple de comparaison entre les locaux MAJIC et INSEE de type maison et de type appartement. Le détail des résultats obtenus est présenté en Annexe 3.

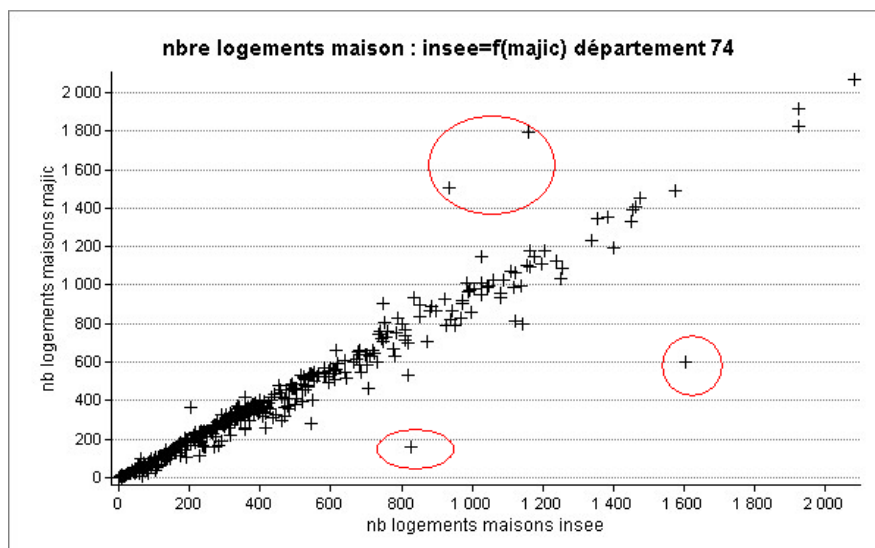


Figure 28 : comparaison par IRIS du nombre de logements INSEE de type maison avec le nombre de logements MAJIC

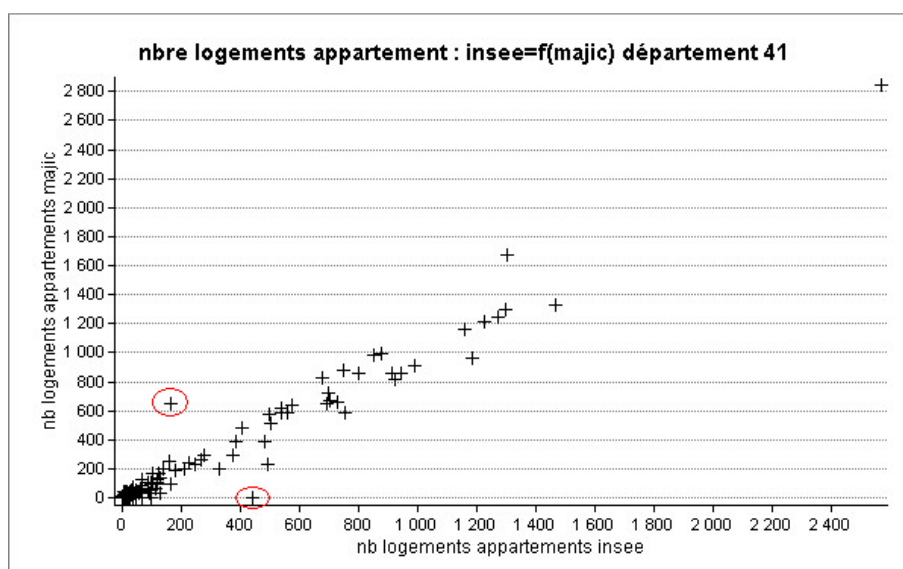


Figure 29 : comparaison par IRIS du nombre de logements INSEE de type appartement avec le nombre de logements MAJIC

Même si certains écarts sont constatés (en rouge Figure 28 et Figure 29) pour certains IRIS, une bonne cohérence entre les sources INSEE et MAJIC est relevée. La validation de la méthode est en cours auprès des AASQA des régions concernées. Des analyses complémentaires seront menées ultérieurement sur les IRIS considérés comme valides.

Ces premiers résultats montrent la nécessité d'une validation approfondie des données issues de la base de données constituée.

Certaines informations issues des différentes étapes de la méthodologie sont utiles pour évaluer la qualité de ces données :

- Les étapes 3 et 4 affectent les locaux et populations sur les bâtiments. Lors de ces étapes, lorsqu'une parcelle ne peut être localisée, tous les locaux associés à celles-ci ne peuvent être géolocalisés. Malgré cela, les données contenues dans MAJIC permettent de déterminer la commune

d'appartenance des locaux. Il est donc possible d'établir un taux de couverture des locaux MAJIC sur les communes. Si ce taux de spatialisation pour une commune est égal à 1, alors, la répartition spatiale des populations résultantes sera considérée valide, et tous les IRIS associés également. Par contre, si ce ratio est inférieur mais reste proche de 1, il sera utile d'évaluer si l'écart engendré est convenable. Des statistiques par parcelle apportent une aide à cette validation. Prenons l'exemple d'une commune présentée sur la Figure 30, les contours des parcelles sont représentés en vert. Cet exemple est une représentation simplifiée de la réalité car une commune comporte généralement plusieurs milliers de parcelles. De plus les parcelles non localisées ne devraient pas apparaître, mais elles sont affichées pour l'exemple. Chaque chiffre sur les cartes présente le nombre de locaux qui n'ont pu être localisés. On calcul le minimum, le maximum et la moyenne du nombre de locaux non positionnés sur la commune par parcelle. Différents cas sont rencontrés :

- o Partie gauche : La moyenne du nombre de locaux non localisés est faible, le nombre minimum et maximum sont faibles également : on peut estimer que l'erreur est diffuse et dans ce cas, la commune et les IRIS associés devraient être valides.
- o Au milieu : le nombre maximum de locaux est important, 200 locaux d'une même parcelle ne sont pas positionnés, il est probable qu'un groupe d'immeuble ne soit pas localisés. A charge de l'AASQA de déterminer si cette erreur de localisation est acceptable ou non. Pour des cas similaires, le LCSQA extraira de la BD MAJIC l'adresse des locaux associés. Ces adresses permettront pour certains cas de repérer l'IRIS sur lequel se porte l'erreur. Cet IRIS pourrait alors être invalidé et les autres IRIS de la commune retenus.
- o Partie droite : le nombre de locaux non localisés par parcelle n'est pas négligeable, il est utile d'approfondir l'analyse pour statuer sur la validité de la commune et des IRIS qui la constitue.

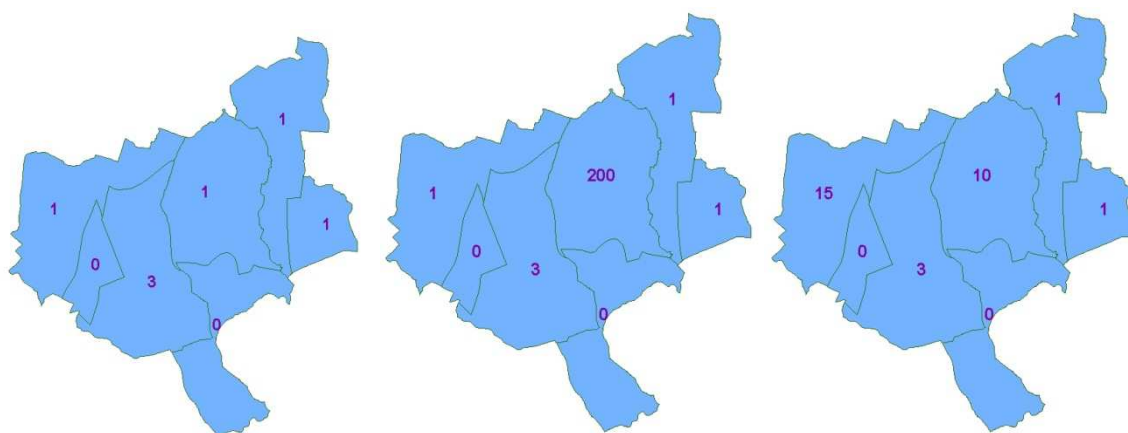


Figure 30 : commune présentant des parcelles où certains locaux n'ont pu être localisés – A gauche : seulement quelques locaux par parcelles ne sont pas localisés – Au milieu : une parcelle présente de nombreux locaux non localisés – A droite : les parcelles présentent en général un certain nombre non négligeable de locaux non localisés

- Pour les IRIS, où l'étape de validation précédente ne permet de conclure sur la validité, il est nécessaire d'utiliser d'autres éléments des étapes de

la méthodologie. L'étape 1 constitue le référentiel parcellaire sur lequel la méthodologie s'appuie. Le cumul des surfaces des parcelles sur chacun des IRIS fournit un indicateur de validité. La Figure 31 présente le taux de couverture de la BD Parcellaire sur les régions étudiées. Un faible taux de couverture n'implique pas nécessairement une mauvaise couverture des populations. En effet, certains IRIS sont traversés par des aléas naturels ou artificiels comme des fleuves ou des autoroutes. Ceux-ci couvrent une surface non négligeable de la superficie de l'IRIS et biaisent ainsi les résultats. Si les parcelles ne couvrent pas entièrement l'IRIS, il est nécessaire de vérifier visuellement que les zones non couvertes ne comportent pas ou peu d'habitation. La superposition des zones non couvertes avec des images satellitaires et avec le bâti de la BD Topo peut être une bonne stratégie de vérification. La BD topo n'étant pas non plus exempte de manques sur certaines zones, elle sera utilisée avec précaution.

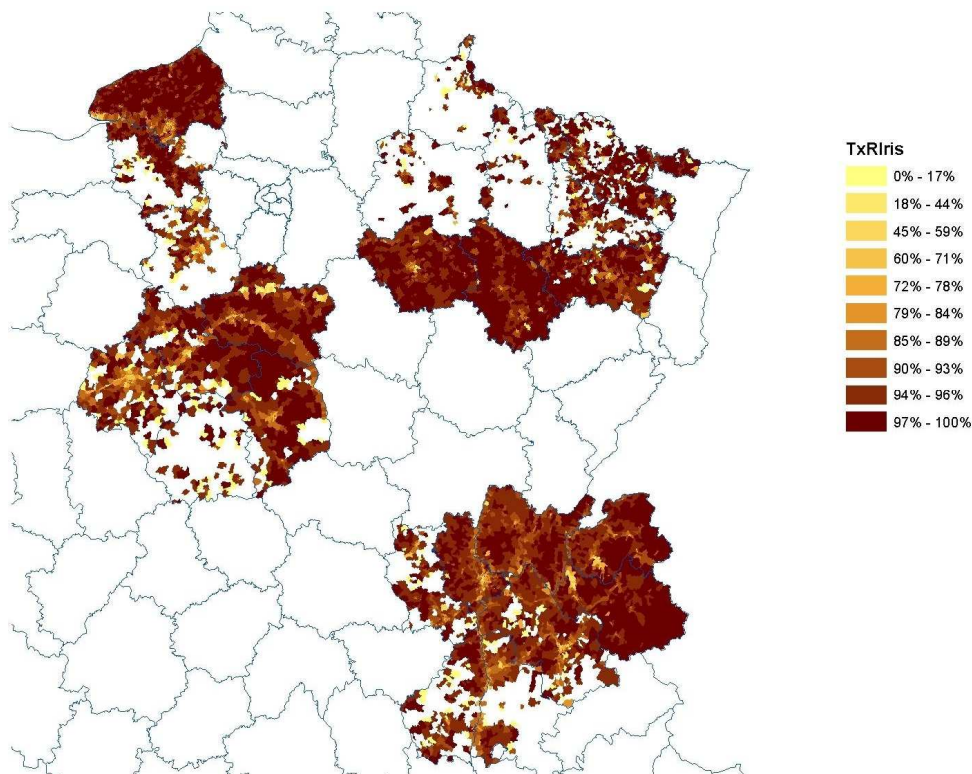


Figure 31 : Taux de couverture de la BD Parcellaire calculé sur 5 régions françaises

- L'étape 2 caractérise les bâtiments utilisés pour spatialiser les locaux MAJIC, puis les étapes 3 et 4 affectent les locaux et populations sur les bâtiments. Quelques questions peuvent alors être posées pour aider au contrôle de ces données :
  - o Des bâtiments comportent-ils un nombre trop important d'habitants ? Si c'est le cas, il serait utile d'analyser les zones concernées pour vérifier l'information.

- o Existe - t'il une cohérence entre le nombre de locaux MAJIC et le nombre de locaux INSEE ? Des écarts importants peuvent être constatés qui nécessitent une étude approfondie sur les zones concernées.

Une expertise forte des AASQA sera nécessaire pour valider ces données.

#### **4.6.2 LES AUTRES METHODOLOGIES**

Même si la BD Topo couvre, a priori, l'intégralité du territoire français, des écarts peuvent apparaître avec la réalité du terrain notamment sur les zones nouvellement construites. Certains nouveaux lotissements (maisons ou immeubles) ne sont pas encore géoréférencés dans cette base et, de ce fait, biaisent localement l'affectation de la population sur le bâti. Des écarts constatés avec la méthodologie MAJIC pourraient également apparaître avec la BD Topo. La recherche des bâtiments comportant un nombre trop important de résidents est certainement utile pour valider ces données.

La méthodologie carroyage INSEE est pour partie dépendante de la qualité des données issues des carreaux INSEE. En effet, l'INSEE s'appuie sur les adresses des contribuables pour les géo-référencer. Le géocodage d'adresse n'est pas aisé car il est parfois difficile d'associer une adresse à une base de données d'adresses spatialisées. Les terminologies employées peuvent différer et empêcher pour certaines adresses leur localisation exacte. Dans le pire des scénarios, elles sont localisées au centre de la commune. De plus, l'adresse est située en bord de rue alors que le bâtiment est parfois décalé de plusieurs dizaines de mètres pour des complexes type HLM par exemple. La position réelle d'un immeuble peut donc être située en dehors de la maille le référençant. Des écarts importants sont nécessairement constatés à des échelles fines (carreaux de 200 m).

Afin d'évaluer la pertinence des méthodologies BD Topo et Carroyage INSEE aux échelles attendues, leurs résultats sont comparés à ceux fournis par la méthode MAJIC. Les zones sur lesquelles s'est portée la comparaison sont celles où la méthodologie MAJIC a été évaluée comme correcte. Les supports de population étant différents pour les méthodologies, il a été choisi d'appliquer les populations sur des maillages réguliers pour les comparer maille à maille (au prorata des surfaces intersectées). Deux échelles sont retenues pour comparer les données : 200 m et 1km. Ces maillages correspondent aux carroyages utilisés par l'INSEE. A la vue des résultats sur les mailles de 200 m, il n'a pas été jugé utile d'évaluer la BD Topo sur un maillage plus fin. Seules les mailles entièrement incluses dans les zones valides sont retenus.

La comparaison est effectuée sur le département d'Indre et Loire.

Lors de l'élaboration de la méthodologie carroyage INSEE (4.4.2), plusieurs hypothèses ont été testées :

- **Hypothèse 1** : La donnée de population carroyée INSEE est utilisée comme pondérateur pour répartir les populations à l'IRIS sur chacune des mailles - **Hypothèse 2** : la donnée de population carroyée est uniquement un indicateur de présence de population sur la maille.
- **Hypothèse 3** : Les carreaux 200m de données de population carroyée INSEE sont utilisés - **Hypothèse 4** : Les carreaux 1 km de données de population carroyée INSEE sont utilisés.

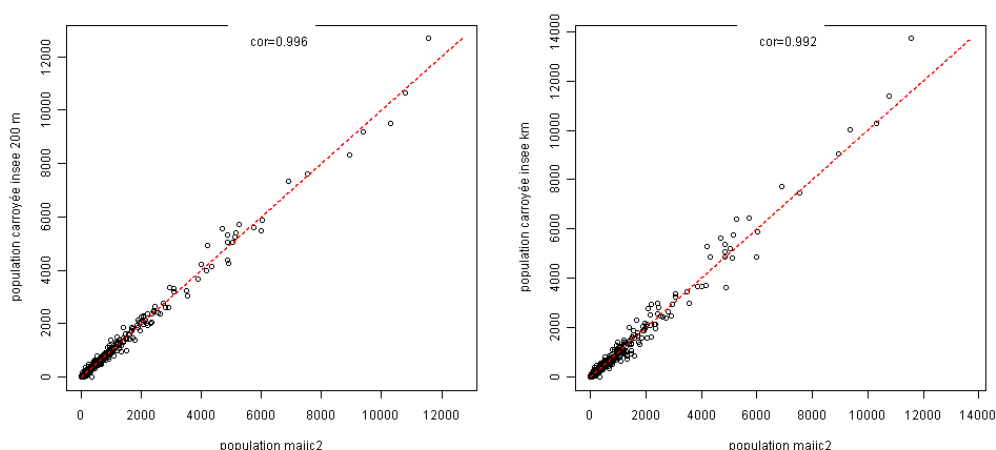


Figure 32 : Comparaison de la méthodologie MAJIC avec la méthodologie carroyage pour l'hypothèse 1 : à gauche, utilisation des carreaux 200 m agrégés au km – à droite, utilisation des carreaux 1km

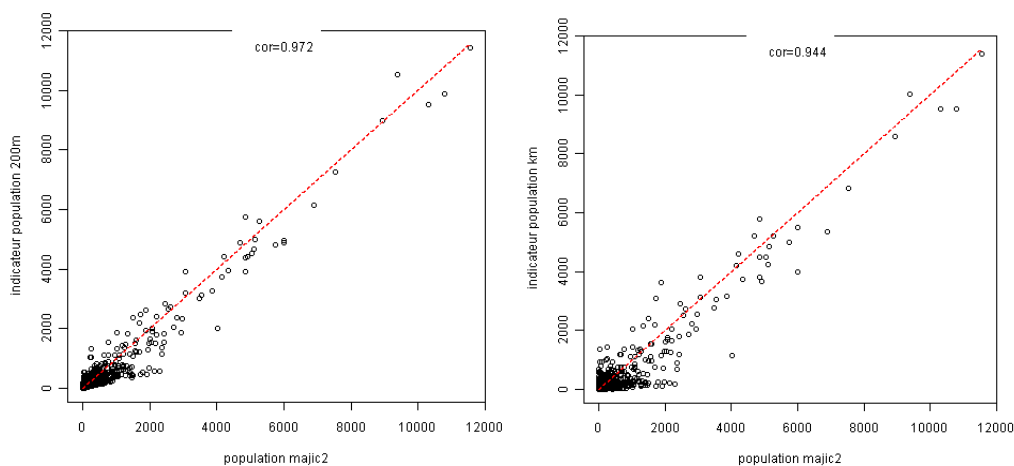


Figure 33 : Comparaison de la méthodologie MAJIC avec la méthodologie carroyage pour l'hypothèse 2 : à gauche, utilisation des carreaux 200 m agrégés au km – à droite, utilisation des carreaux 1km

Tableau 5 : comparaison des résultats de la méthodologie carroyage appliquée selon différentes hypothèses avec les résultats de la méthode MAJIC. Les chiffres du tableau sont des nombres d'habitants.

|                | MAJIC2 | Hypothèse 1          |                     | Hypothèse 2          |                     |
|----------------|--------|----------------------|---------------------|----------------------|---------------------|
|                |        | Hypothèse 3<br>200 m | Hypothèse 4<br>1 km | Hypothèse 3<br>200 m | Hypothèse 4<br>1 km |
| Minimum        | 0      | 0                    | 0                   | 0                    | 0                   |
| Maximum        | 11549  | 12673                | 13713               | 11424                | 11393               |
| Moyenne        | 143    | 143                  | 143                 | 141                  | 140                 |
| P95            | 511    | 522                  | 524                 | 431                  | 298                 |
| P5             | 0      | 0                    | 0                   | 0                    | 0                   |
| Ecart type     | 611    | 613                  | 638                 | 570                  | 552                 |
| biais<br>moyen |        | 0                    | 0.18                | -2.03                | -3.51               |
| Corrélation    |        | 0.996                | 0.992               | 0.972                | 0.944               |

L'utilisation conjointe des hypothèses 1 et 3 qui utilise des mailles de 200 m agrégées sur 1 km et répartit la population sur les mailles en fonction des données carroyées INSEE présente les meilleurs résultats. Le biais moyen est nul, la corrélation est très proche de 1 et le nuage de corrélation présente peu de dispersion. L'application de l'hypothèse 3 dégrade fortement les résultats.

Les paramètres utilisés pour la méthodologie BD Topo sont les suivantes : les bâtiments d'une superficie au sol inférieure à 50 m<sup>2</sup> et de hauteur inférieure à 3 m ne sont pas retenus. Les vues aériennes ont été utilisées pour :

- vérifier le type des bâtiments de grande surface au sol,
- vérifier l'utilisation des PAI (points d'activité ou d'intérêt cf 4.5.3.1) pour éliminer des bâtiments,
- évaluer la hauteur des bâtiments dont le champ 'hauteur' n'est pas renseigné.

Une comparaison conjointe BD Topo et carroyage est proposée aux deux échelles 200 m et 1 km.

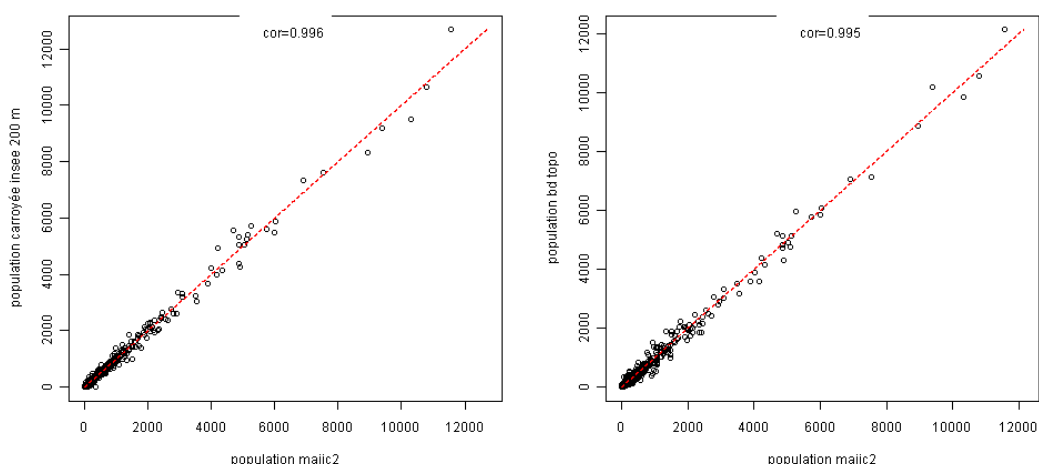


Figure 34 : Comparaison de la méthodologie MAJIC sur des mailles de 1 km avec : à gauche, la méthodologie carroyage pour l'hypothèse 1-3 - à droite, la méthodologie BD Topo

Tableau 6 : comparaison des méthodologies carroyage et BD Topo avec la méthode MAJIC sur des mailles kilométriques

|                | MAJIC2 | Hypothèse 1-3<br>200 m | Méthode BD Topo |
|----------------|--------|------------------------|-----------------|
| Minimum        | 0      | 0                      | 0               |
| Maximum        | 11549  | 12673                  | 12121.16507     |
| Moyenne        | 143    | 143                    | 142.6315052     |
| P95            | 511    | 522                    | 495.1971468     |
| P5             | 0      | 0                      | 0               |
| Ecart type     | 611    | 613                    | 608.6080545     |
| biais<br>moyen |        | 0                      | -0.65           |
| Corrélation    |        | 0.996                  | 0.995           |

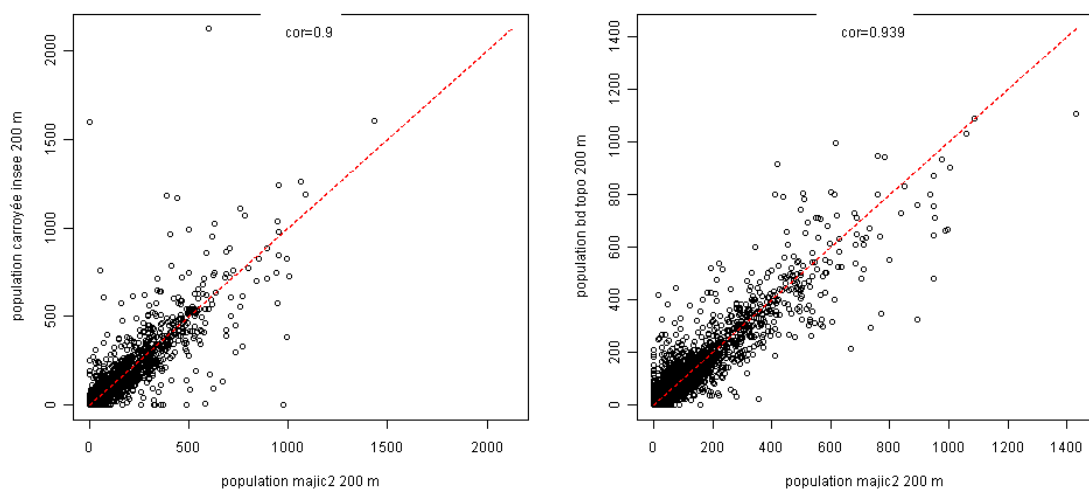


Figure 35 : Comparaison de la méthodologie MAJIC sur des mailles de 200 m avec : à gauche, la méthodologie carroyage pour l'hypothèse 1-3 - à droite, la méthodologie BD Topo

Tableau 7 : comparaison des méthodologies carroyage et BD Topo avec la méthode MAJIC sur des mailles 200 m

|                | MAJIC2     | méthode<br>« population<br>carroyée » 200 m | Méthode BD Topo |
|----------------|------------|---------------------------------------------|-----------------|
| Minimum        | 0          | 0                                           | 0               |
| Maximum        | 1430.04666 | 2124.670166                                 | 1106.806884     |
| Moyenne        | 5.14094827 | 5.1410906                                   | 5.134078196     |
| P95            | 20.2844722 | 21.19651                                    | 20.144502       |
| P5             | 0          | 0                                           | 0               |
| Ecart type     | 31.0674017 | 32.87726962                                 | 30.3149634      |
| biais<br>moyen |            | -0.01                                       | 0               |
| Corrélation    |            | 0.9                                         | 0.939           |

A l'échelle kilométrique, les méthodes BD Topo et population carroyée sont comparables (Figure 34 et

Tableau 6). La dispersion des nuages de points reste limitée et les corrélations sont excellentes.

Une nette dégradation des résultats est constatée sur une échelle plus fine (Figure 35 et Tableau 7) mais les corrélations restent malgré tout correctes. La méthodologie BD Topo apparaît meilleure que la méthode carroyage avec une dispersion du nuage de corrélation moins prononcée. Quelques écarts importants apparaissent sur la méthode carroyage (Figure 35). Ces écarts sont dus au mauvais positionnement de l'adresse de zones type HLM. Les populations sont alors situées sur une maille adjacente à celle attendue.

## **5. CONCLUSIONS ET PERSPECTIVES**

La législation européenne impose un suivi de la pollution dans l'air ambiant et de l'exposition des populations à cette pollution. Ces 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.

A l'échelle nationale et régionale, la méthodologie carroyage est proposée. Elle constitue un développement du carroyage de la population réalisé par l'INSEE dont elle corrige certains manques. La population est répartie sur des mailles kilométriques en fonction des données de revenus fiscaux et de la description cadastrale du territoire.

A l'échelle locale, la méthodologie MAJIC proposée par le CETE de Lyon et enrichie par le LCSQA est retenue comme méthode de référence. Elle permet une description très fine de la population grâce à une description détaillée des bâtiments, de la nature des locaux et de leur répartition spatiale.

Toutefois, du fait d'un manque de couverture sur l'ensemble du territoire, cette méthodologie ne peut actuellement être appliquée partout et doit, le cas échéant, être remplacée par la méthodologie BD Topo. L'application de la méthodologie MAJIC nécessite donc une expertise locale forte. Des validations sont en cours avec des ASQAA volontaires. Elles apporteront probablement quelques ajustements sur la méthode et fixeront en détail la manière de valider les données de population qui en résultent.

Toutes ces méthodologies sont conçues pour fournir des résultats cohérents avec les données du recensement national, qui constituent la référence à l'échelle de l'IRIS.

La méthodologie MAJIC, qui est considérée comme la plus précise, a permis de valider la méthodologie carroyage. Celle-ci a été développée pour l'exploitation des sorties de modèles régionaux des AASQA et la plateforme nationale PREV'AIR. En particulier, le LCSQA emploie déjà cette méthodologie dans le cadre de PREV'AIR Urgence, afin d'évaluer les populations exposées à des dépassements de seuil d'information ou d'alerte.

Le LCSQA met à disposition des AASQA qui le souhaitent une extraction des données carroyées kilométriques sur leur zone de compétences. Celles-ci sont mises à jour chaque année avec les nouvelles statistiques du recensement de la population.

Du fait de contraintes sur l'accès aux données MAJIC et de la complexité de la méthodologie MAJIC, le LCSQA assurera la mise en œuvre de cette approche et mettra à disposition des AASQA les données spatiales de la population qui en sont issues. Des échanges seront nécessaires entre les AASQA et le LCSQA pour la validation de ces données. Dans les zones où la méthodologie MAJIC n'est pas encore utilisable, la mise en œuvre de la méthodologie BD Topo sera laissée aux soins des AASQA avec, s'il est besoin, le concours du LCSQA. Pour les AASQA qui ne disposeraient pas des ressources suffisantes, le LCSQA pourra également appliquer cette méthode.

L'IGN continue d'améliorer continuellement la couverture de la BD Parcellaire 'Vecteur'. Il sera utile de suivre chaque année ses mises à jour<sup>14</sup> pour augmenter le nombre de zones éligibles à la méthodologie MAJIC.

Pour les zones où la BD Parcellaire 'vecteur' n'existe pas, le CETE Nord-Picardie utilise d'autres sources de données pour localiser le centre des parcelles. Cette action améliore grandement la couverture géographique des données MAJIC pour atteindre un taux de couverture des parcelles proche de 99 %. L'utilisation croisée des centres de parcelles, des adresses des locaux MAJIC et du thème 'bâti' de la BD Topo pourrait fournir une méthode alternative à la méthodologie BD Topo en apportant une qualité améliorée du positionnement des populations. Cette méthode est déjà pour partie décrite dans les travaux menés par le CETE de l'Ouest pour élaborer l'observatoire national des services de communication électronique (cf 3.2.2.1). Des travaux complémentaires seront nécessaires pour développer et appliquer cette méthode et évaluer le gain apporté sur la méthodologie Bd Topo.

---

<sup>14</sup> [http://professionnels.ign.fr/sites/default/files/Tableau\\_BDPARCELLAIRE\\_DEP.pdf](http://professionnels.ign.fr/sites/default/files/Tableau_BDPARCELLAIRE_DEP.pdf)

## 6. REFERENCES

CERTU, 2005, Méthodes d'estimation des populations - [http://www.observation-urbaine.certu.equipement.gouv.fr/IMG/pdf/Rapport\\_d\\_etude\\_cle28e914-1.pdf](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.

Caudeville J., 2011, Développement d'une plateforme intégrée pour la cartographie de l'exposition des populations aux substances chimiques: construction d'indicateurs spatialisés en vu d'identifier les inégalités environnementales à l'échelle régionale. Compiègne : Université Technologique de Compiègne, 198 pages.

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.

## **7. LISTE DES ANNEXES**

| Référence | Désignation                                                                          | Nombre de pages |
|-----------|--------------------------------------------------------------------------------------|-----------------|
| Annexe 1  | Fiche MOCA LCSQA 2012                                                                | 3               |
| Annexe 2  | Classes d'occupation du sol de la base Corine Land Cover                             | 3               |
| Annexe 3  | Comparaison par iris du nombre de logements INSEE et MAJIC (appartements et maisons) | 7               |



## **ANNEXE 1**

---

Fiche MOCA programme LCSQA 2012

## **Thème 6 Modélisation et traitements numériques**

### **ETUDE N 2 Cartographie des concentrations et des populations exposées aux dépassements de seuils**

**Institut responsable de l'étude: INERIS**

#### **Contexte et objectif**

Pour toute situation de dépassement constatée, la législation impose de délimiter la zone géographique ainsi que la population exposée à ce dépassement. La production de cartographies, aussi bien à grande échelle qu'à l'échelle locale, s'est considérablement développée afin de répondre à ces exigences. Plus précisément, l'estimation de l'exposition nécessite de cartographier préalablement deux types de variables: les concentrations d'une part, la densité de population d'autre part. Pour la première application, différentes approches sont utilisées (exploitation géostatistique de données de campagnes, modélisation, combinaison de données de mesure et de modélisation), pour la seconde, la répartition est généralement effectuée selon l'occupation du sol ou le bâti.

Cette étude a pour objet :

- de réaliser un état des lieux sur les méthodes employées par le LCSQA et les AASQA pour cartographier les concentrations de polluants, les zones exposées à des dépassements de seuils réglementaires et les populations exposées à ces dépassements ;
- d'en faire l'analyse et d'extraire de celle-ci des recommandations de méthode et d'usage, fonctions du contexte et du type de cartographie visé. Le cas échéant, une harmonisation des pratiques sera proposée.
- 

#### **Travaux antérieurs**

Depuis quelques années, des études ont été réalisées sur la cartographie des zones de dépassement et des populations exposées. A l'échelle nationale, le travail a été orienté sur les PM<sub>10</sub> pour les besoins liés à la demande de report de délai concernant le respect des valeurs limites, puis en appui à la définition des zones sensibles. Les données de mesure (stations fixes) et de modélisation (CHIMERE) ont été principalement exploitées. A l'échelle urbaine, les travaux ont porté plus spécialement sur le NO<sub>2</sub>, avec le traitement géostatistique de données de campagnes.

#### **Travaux proposés pour 2012**

##### **1) Revue de l'existant**

Cette partie comprendra une consultation de l'ensemble des AASQA ainsi qu'une étude bibliographique. Le GT *Evaluation de plans*, pour qui l'estimation de l'exposition représente une question essentielle, et le GT *Révision du guide d'implantation des stations de mesure*, qui a commencé à recueillir des informations sur les techniques de répartition spatiale de la population, seront également interrogés.

Les différents usages de la cartographie seront recensés en distinguant les points suivants : objectif des cartographies ; variables cartographiées ; méthodes et outils de cartographie ; données d'entrée ; avantages et limites constatés, etc.

## 2) Analyse et recommandations

L'état des lieux décrit ci-dessus fera l'objet d'une analyse. En confrontant les conclusions de celle-ci à l'expérience et aux évaluations qui ont pu être réalisées, on élaborera des recommandations sur le choix des méthodes et la façon de les mettre en œuvre. L'étude sur la modélisation urbaine et locale prévue pour 2012 alimentera également ce travail.

Lorsqu'il est possible, une approche harmonisée sera proposée. En particulier, les AASQA attendent du LCSQA la mise au point d'une méthodologie commune de répartition de la population. Il s'agit d'une demande urgente qui sera traitée au premier semestre 2012.

Dans tous les cas, différentes situations pourront être considérées selon l'échelle de travail (spatiale et temporelle) et les données d'entrée disponibles.

## 3) Application

Les préconisations et les méthodologies identifiées précédemment seront appliquées, en collaboration avec des AASQA volontaires, à quelques cas choisis parmi les applications de cartographie les plus courantes. Elles seront affinées s'il est besoin d'après les résultats obtenus.

## Renseignements synthétiques 2012

|                                    |                                                                                                      |              |         |
|------------------------------------|------------------------------------------------------------------------------------------------------|--------------|---------|
| Titre de l'étude                   | <b><i>Cartographie des concentrations et des populations exposées aux dépassements de seuils</i></b> |              |         |
| Personne responsable de l'étude    | INERIS : Laure Malherbe                                                                              |              |         |
| Travaux                            | pérennes                                                                                             |              |         |
| Durée des travaux pluriannuels     | 1 an                                                                                                 |              |         |
| Collaboration AASQA                | OUI                                                                                                  |              |         |
| Heures d'ingénieur                 | EMD :                                                                                                | INERIS : 500 | LNE : - |
| Heures de technicien               | EMD :                                                                                                | INERIS :     | LNE : - |
| Document de sortie attendu         | Rapport                                                                                              |              |         |
| Lien avec le tableau de suivi CPSe |                                                                                                      |              |         |
| Lien avec un groupe de travail     | GT <i>Evaluation de plans</i><br>GT <i>Révision du guide d'implantation des stations de mesure</i>   |              |         |
| Matériel acquis pour l'étude       |                                                                                                      |              |         |



## ANNEXE 2

---

### CLASSES D'OCCUPATION DU SOL DE LA BASE CORINE LAND COVER

#### Niveau 1

| Code     | Libellé français                |
|----------|---------------------------------|
| <b>1</b> | Territoires artificialisés      |
| <b>2</b> | Territoires agricoles           |
| <b>3</b> | Forêts et milieux semi-naturels |
| <b>4</b> | Zones humides                   |
| <b>5</b> | Surfaces en eau                 |

#### Niveau 2

| Code      | Libellé français                                                |
|-----------|-----------------------------------------------------------------|
| <b>11</b> | Zones urbanisées                                                |
| <b>12</b> | Zones industrielles ou commerciales et réseaux de communication |
| <b>13</b> | Mines, décharges et chantiers                                   |
| <b>14</b> | Espaces verts artificialisés, non agricoles                     |
| <b>21</b> | Terres arables                                                  |
| <b>22</b> | Cultures permanentes                                            |
| <b>23</b> | Prairies                                                        |
| <b>24</b> | Zones agricoles hétérogènes                                     |
| <b>31</b> | Forêts                                                          |
| <b>32</b> | Milieux à végétation arbustive et/ou herbacée                   |
| <b>33</b> | Espaces ouverts, sans ou avec peu de végétation                 |
| <b>41</b> | Zones humides intérieures                                       |
| <b>42</b> | Zones humides maritimes                                         |
| <b>51</b> | Eaux continentales                                              |
| <b>52</b> | Eaux maritimes                                                  |

### Niveau 3

| Code       | Libellé français                                                                     |
|------------|--------------------------------------------------------------------------------------|
| <b>111</b> | Tissu urbain continu                                                                 |
| <b>112</b> | Tissu urbain discontinu                                                              |
| <b>121</b> | Zones industrielles et commerciales                                                  |
| <b>122</b> | Réseaux routier et ferroviaire et espaces associés                                   |
| <b>123</b> | Zones portuaires                                                                     |
| <b>124</b> | Aéroports                                                                            |
| <b>131</b> | Extraction de matériaux                                                              |
| <b>132</b> | Décharges                                                                            |
| <b>133</b> | Chantiers                                                                            |
| <b>141</b> | Espaces verts urbains                                                                |
| <b>142</b> | Equipements sportifs et de loisirs                                                   |
| <b>211</b> | Terres arables hors périmètre' d'irrigation                                          |
| <b>212</b> | Périmètres irrigués en permanence                                                    |
| <b>213</b> | Rizières                                                                             |
| <b>221</b> | Vignobles                                                                            |
| <b>222</b> | Vergers et petits fruits                                                             |
| <b>223</b> | Oliveraies                                                                           |
| <b>231</b> | Prairies                                                                             |
| <b>241</b> | Cultures annuelles associées aux cultures permanentes                                |
| <b>242</b> | Systèmes cultureux et parcellaires complexes                                         |
| <b>243</b> | Surfaces essentiellement agricoles, interrompues par des espaces naturels importants |
| <b>244</b> | Territoires agro-forestiers                                                          |
| <b>311</b> | Forêts de feuillus                                                                   |
| <b>312</b> | Forêts de conifères                                                                  |
| <b>313</b> | Forêts mélangées                                                                     |
| <b>321</b> | Pelouses et pâturages naturels                                                       |
| <b>322</b> | Landes et broussailles                                                               |
| <b>323</b> | Végétation sclérophylle                                                              |
| <b>324</b> | Forêt et végétation arbustive en mutation                                            |
| <b>331</b> | Plages, dunes et sable                                                               |
| <b>332</b> | Roches nues                                                                          |
| <b>333</b> | Végétation clairsemée                                                                |
| <b>334</b> | Zones incendiées                                                                     |
| <b>335</b> | Glaciers et neiges éternelles                                                        |

|            |                      |
|------------|----------------------|
| <b>411</b> | Marais intérieurs    |
| <b>412</b> | Tourbières           |
| <b>421</b> | Marais maritimes     |
| <b>422</b> | Marais salants       |
| <b>423</b> | Zones intertidales   |
| <b>511</b> | Cours et voies d'eau |
| <b>512</b> | Plans d'eau          |
| <b>521</b> | Lagunes littorales   |
| <b>522</b> | Estuaires            |
| <b>523</b> | Mers et océans       |



## **ANNEXE 3**

---

### **COMPARAISON PAR IRIS DU NOMBRE DE LOGEMENTS INSEE** **ET MAJIC (APPARTEMENTS ET MAISONS)**

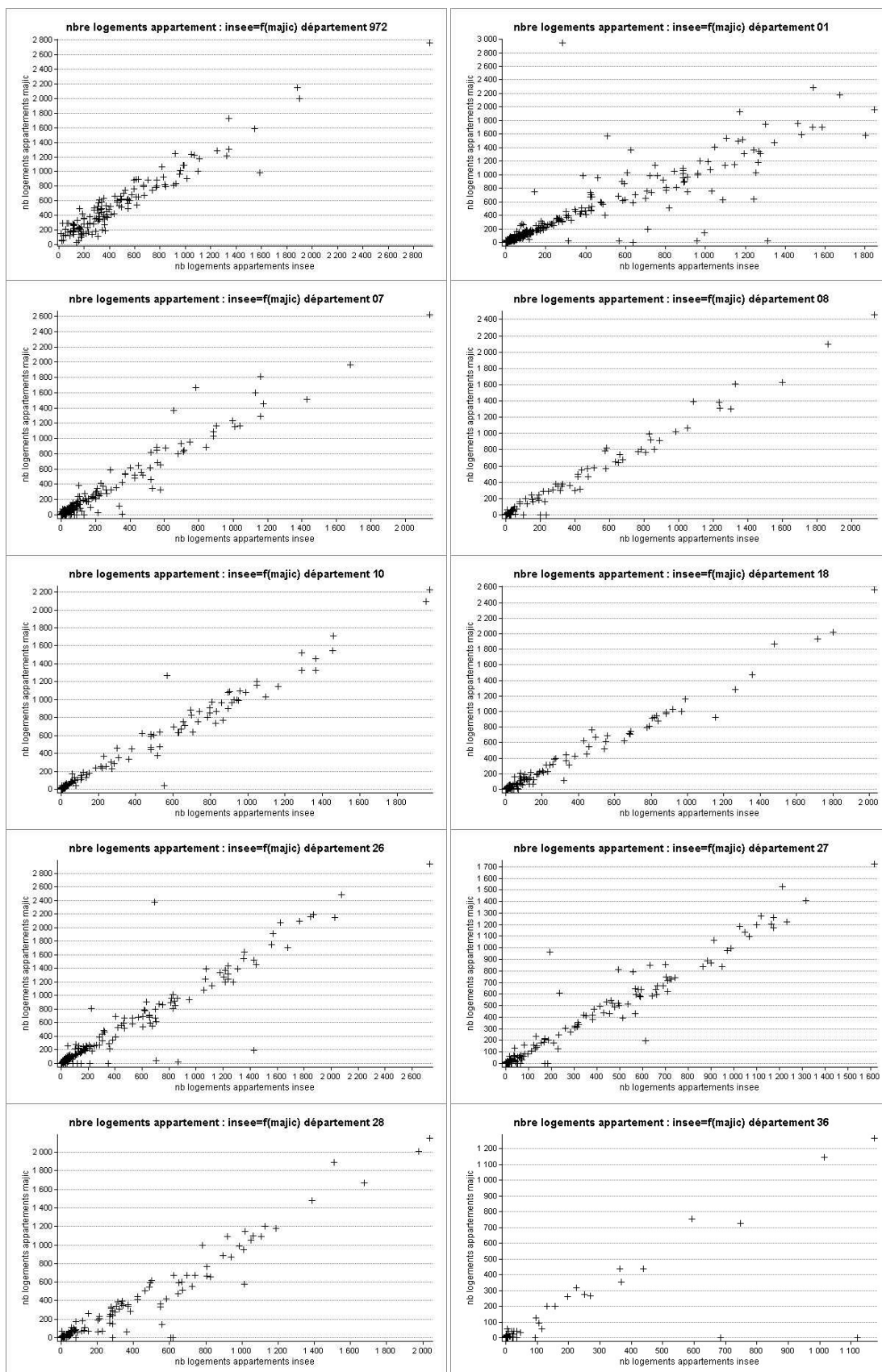


Figure 36 : comparaison par IRIS du nombre de logements INSEE de type appartement avec le nombre de logements MAJIC

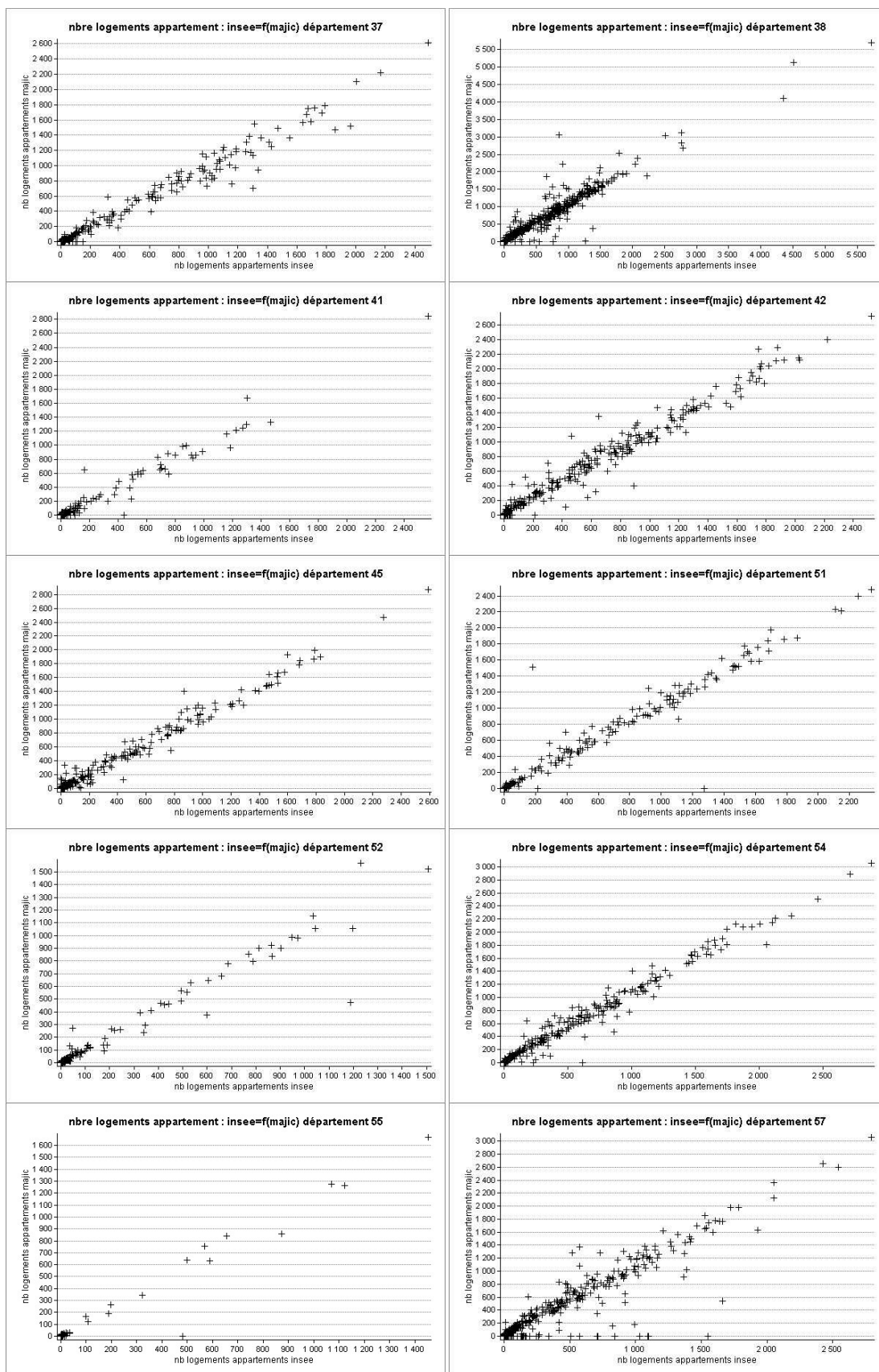


Figure 37 : comparaison par IRIS du nombre de logements INSEE de type appartement avec le nombre de logements MAJIC

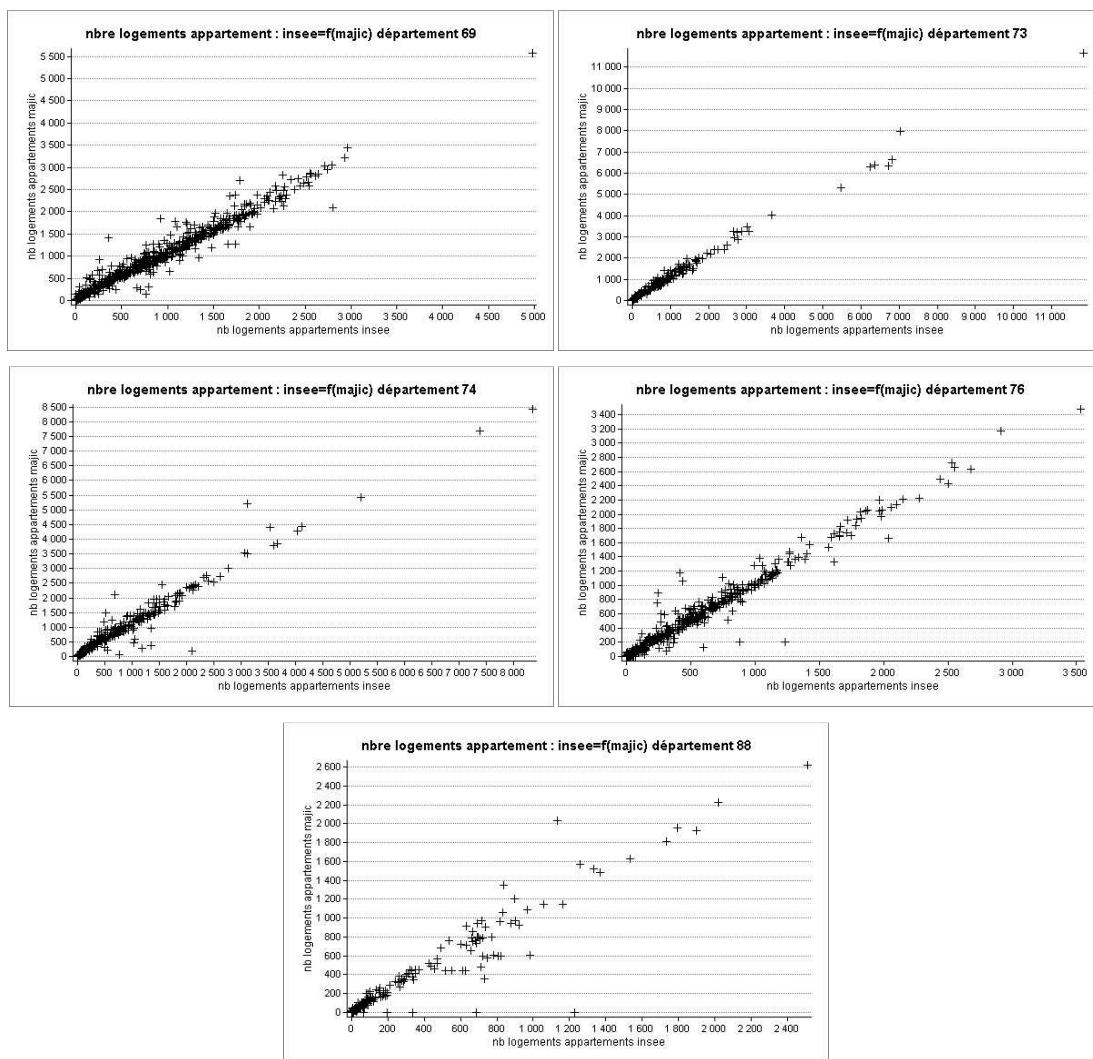


Figure 38 : comparaison par IRIS du nombre de logements INSEE de type appartement avec le nombre de logements MAJIC

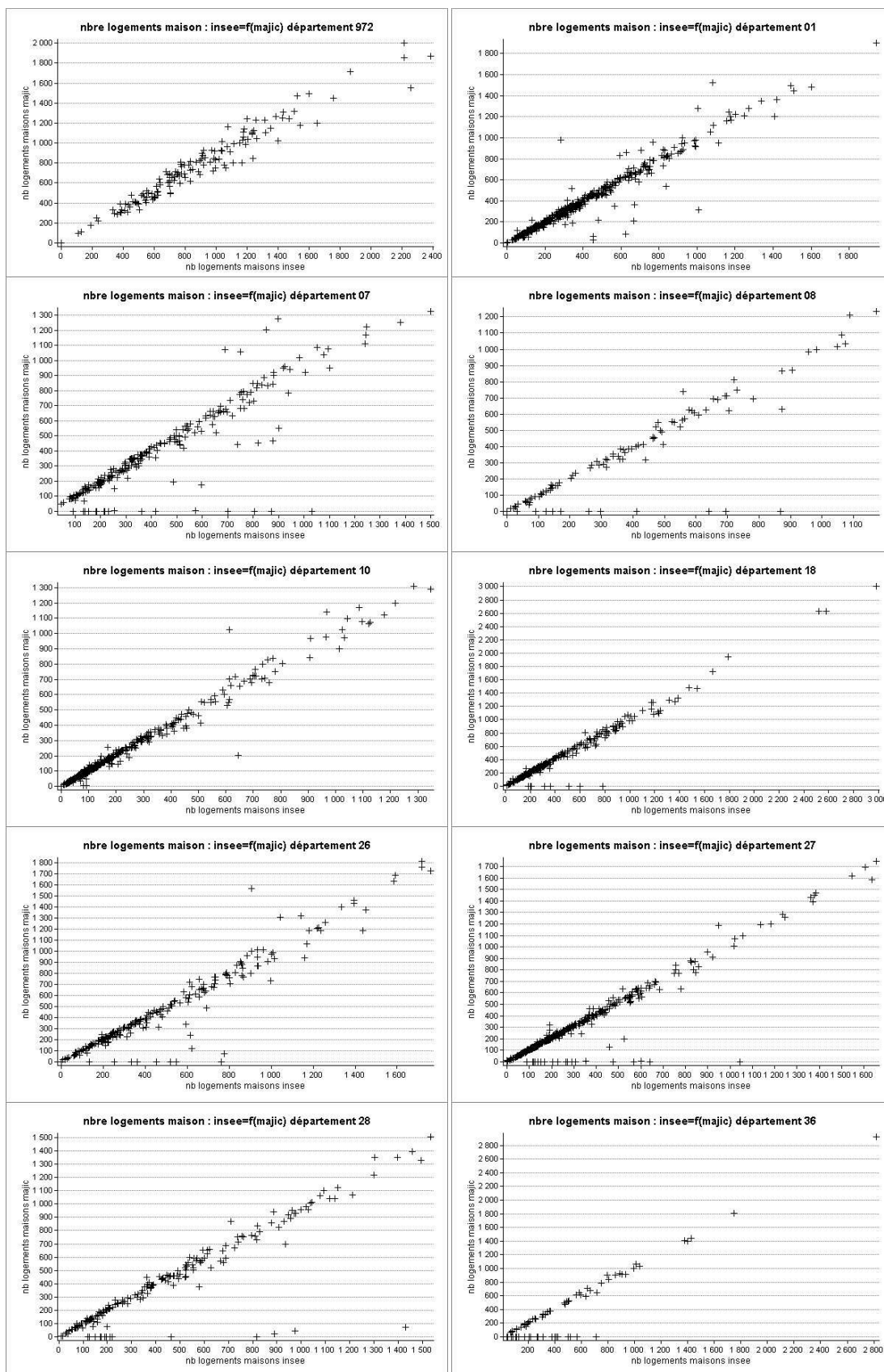


Figure 39 : comparaison par IRIS du nombre de logements INSEE de type maison avec le nombre de logements MAJIC

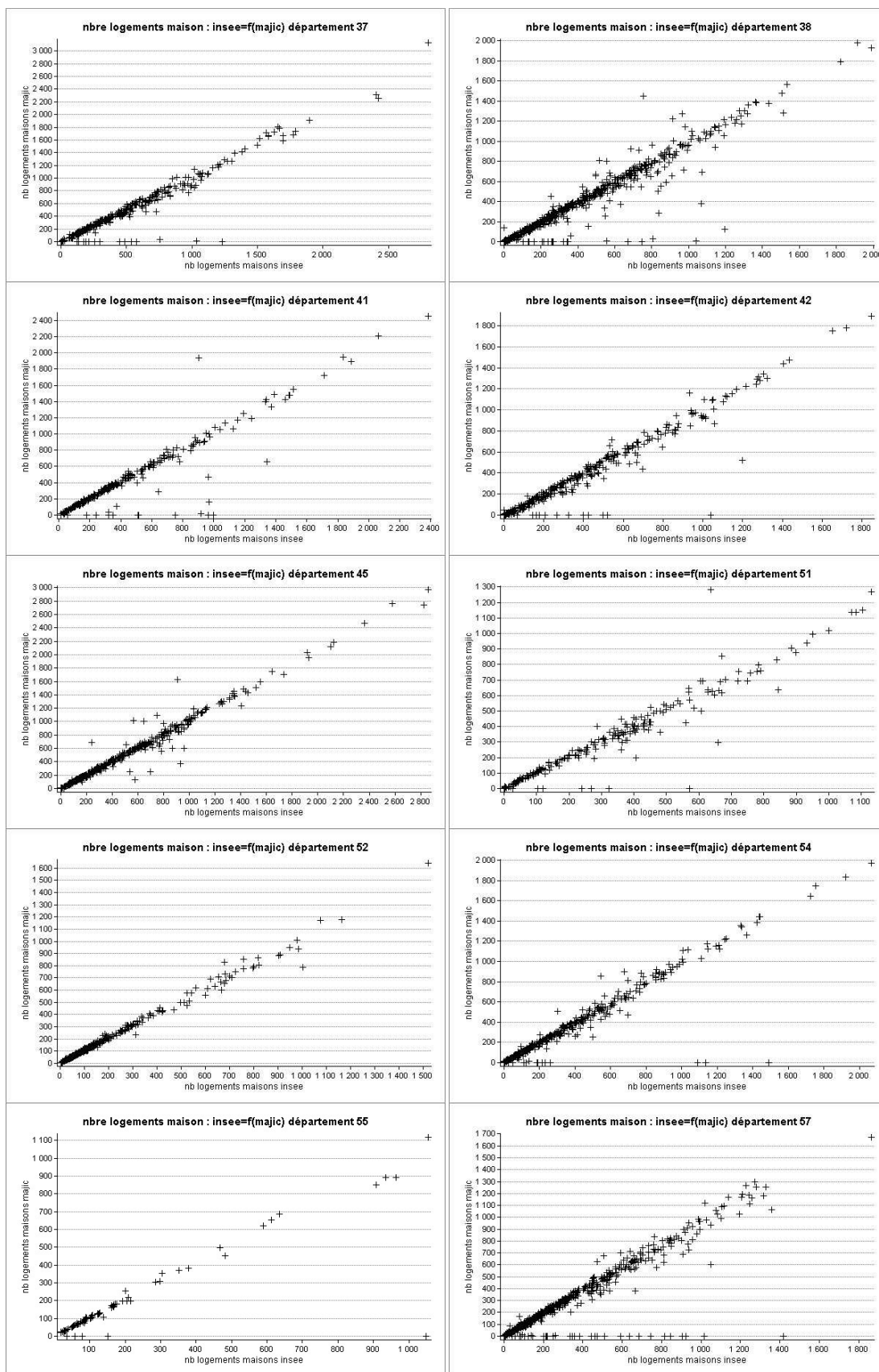


Figure 40 : comparaison par IRIS du nombre de logements INSEE de type maison avec le nombre de logements MAJIC

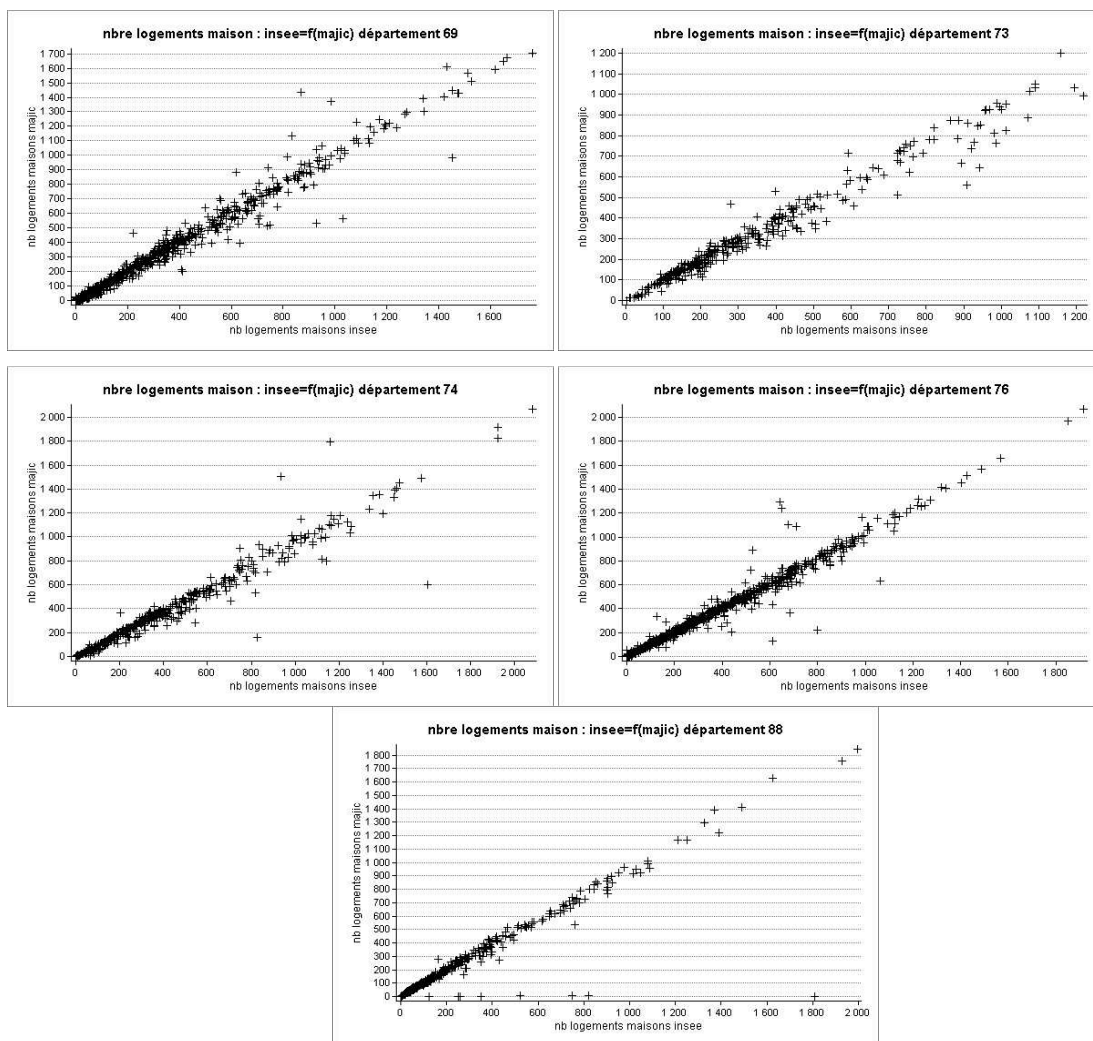


Figure 41 : comparaison par IRIS du nombre de logements INSEE de type maison avec le nombre de logements MAJIC