

# Document de travail n° 8

Rapport - Développement durable

## *Méthodes d'estimation de population et de logements en zone inondable*

## Sommaire

<b>Préambule .....</b>	<b>3</b>
<b>Données utilisées .....</b>	<b>3</b>
<b>1. Estimations à partir des données carroyées de population .....</b>	<b>8</b>
1.1. Données utilisées .....	8
1.2. Analyse des données carroyées de population.....	8
1.3. Estimation de population en zone inondable.....	10
<b>2. Méthodes d'estimation « Popéval »-BD TOPO® .....</b>	<b>11</b>
2.1. Données utilisées .....	11
2.2. Traitement des données SIG.....	11
2.3. Présentation des méthodes.....	12
2.3.1. Méthode simple (au prorata de la surface développée).....	12
2.3.2. Modèles économétriques linéaires .....	14
2.3.2.1. Estimation du modèle .....	14
2.3.2.2. Application du modèle .....	15
2.4. Qualité de l'estimation .....	16
2.4.1. Méthode simple .....	16
2.4.2. Modèles économétriques .....	16
<b>3. Comparaison des deux méthodes .....</b>	<b>16</b>
3.1. Tests sur la région Centre .....	17
3.2. Conclusion .....	18
<b>Annexes.....</b>	<b>19</b>
Annexe 1 : Problèmes de modélisation .....	19
Annexe 2: Méthodes d'estimation des populations au carreau.....	22

## Préambule

En 2011, la Direction générale de la prévention des risques (DGPR) a confié au Service de l'observation et des statistiques (SOeS) le soin de développer un outil d'évaluation permettant d'estimer la population et les logements situés dans l'enveloppe approchée des inondations potentielles (EAIP) dans le cadre de l'évaluation préliminaire des risques d'inondation (EPRI). Cette évaluation constitue la première étape de la Directive inondation 2007/60/CE ; elle correspond à une description des aléas et des enjeux pour la santé humaine, l'environnement et l'activité économique sur les 14 districts hydrographiques français. Un socle national d'une vingtaine d'indicateurs permet de traduire les conséquences négatives que les inondations futures pourraient provoquer et constitue la base de l'EPRI. Les territoires à risque important d'inondation (TRI) sont identifiés à partir de cette évaluation.

La DGPR s'est tournée vers le SOeS pour la réalisation de ces calculs dans un souci de traitement homogène et cohérent des données, mais également parce que le SOeS a depuis 2008 élaboré une méthode d'estimation de population et de logements (méthode dite « Popéval ») dans les zones à risques (retrait, gonflement des argiles, inondation) à l'aide de la base de données CORINE Land Cover (CLC). Cette méthode est présentée dans le document « Méthodologie d'estimation des enjeux exposés aux inondations » disponible en ligne : <http://www.statistiques.developpement-durable.gouv.fr> rubrique Publications > Documents de travail.

Par souci de recourir à une méthode plus robuste et afin de couvrir la quasi-totalité du territoire, y compris les DOM, il a été décidé de manière conjointe avec la DGPR, que la méthode d'évaluation de la population et des logements devait s'appuyer sur la base de données BD TOPO® de l'IGN. Les autres indicateurs de l'EPRI (emprise des bâtiments d'activité, nombre d'établissements de santé, emprise totale du bâti, etc.) reposent également le plus souvent sur cette base de données.

Dans un second temps, en 2012, le SOeS a souhaité mettre à jour les données de population et de logements en zone inondable qu'il diffuse sur son site internet. Les contours pris en compte sont établis à partir des Atlas des zones inondables (AZI) et des plans de prévention des risques d'inondation (PPRI) disponibles sur Cartorisque et auprès des DREAL. Les données carroyées de population, récemment diffusées par l'Insee, ont été utilisées pour l'estimation. Elles correspondent à un comptage de la population fiscale réalisé sur des carreaux de 200 mètres sur 200 mètres.

Ces deux nouvelles méthodes suivent deux logiques différentes. Pour les données carroyées, les données de base sont des données de population que l'on agrège sur un zonage à façon. Pour la méthode BD TOPO®, ce sont des bâtiments dans lesquels on estime un nombre de logements avant d'estimer la population.

Ce document résume les travaux méthodologiques réalisés pour quantifier le nombre de logements et la population dans les zones à risque suivant les deux méthodes présentées ci-dessus. Ces méthodes sont applicables sur n'importe quel zonage (zones inondables, EAIP, zones à risques retrait-gonflement des argiles, etc.).

## Données utilisées

**Plusieurs bases de données géographiques sont utilisées. Les comparaisons des différentes méthodes présentées ont été réalisées à partir des zones inondables ou des enveloppes approchées des inondations potentielles et d'autres données décrites ci-dessous.**

### **Zones inondables : AZI et PPRI disponibles sur Cartorisque (DGPR) et transmis par les DREAL.**

Les atlas des zones inondables (AZI), pour lesquels les circulaires de référence sont la circulaire du 24 janvier 1994, la lettre circulaire du 1er février 2002 et la circulaire aux préfets de région du 4 novembre 2003, constituent les premiers documents de connaissance des zones inondables. Ce sont des outils de référence pour l'élaboration de la politique de gestion des risques d'inondation qui guident les services de l'État dans la programmation de leurs actions pour la gestion des risques d'inondation, en particulier des PPRN inondation. L'AZI est réalisé sous maîtrise d'ouvrage de l'État (DDT ou DREAL).

La base de données Cartorisque ([http://cartorisque.prim.net./](http://cartorisque.prim.net/)) s'est enrichie progressivement des différents contours départementaux des zones inondables. En juillet 2012, la plus grande partie de ces

zones a été numérisée, mais il reste néanmoins des cours d'eau pour lesquels ces cartographies n'étaient pas encore disponibles.

Le zonage réalisé par le SOeS en juillet 2012 utilise les données des AZI et des enveloppes d'aléas des PPRI disponibles sur Cartorisque et transmises par les DREAL.

### **Enveloppe approchée des inondations potentielles**

L'objectif de cette enveloppe est d'approcher au mieux des connaissances disponibles sur l'événement extrême potentiel. L'effet des ouvrages de protection n'est pas considéré.

Deux enveloppes ont été constituées pour répondre aux besoins de l'EPRI :

- une enveloppe approchée des inondations potentielles par cours d'eau ;
- une enveloppe approchée des inondations potentielles par submersion marine.

La couche géographique concernant les cours d'eau assemble les trois types d'informations ci-dessous :

- la synthèse de l'ensemble de la connaissance cartographique concernant les zones inondables disponibles au format SIG au sein des services de l'État : Cartorisque, emprise des AZI, carte d'aléas des PPRI et assimilés, données historiques locales... ;
- des informations sur la géologie (couche des alluvions récentes) pour les cours d'eau importants ;
- de résultats de méthodes simplifiées à grand rendement géographique : les zones basses hydrographiques (résultats de l'application de la méthode EXZECO2 – extraction des zones d'écoulement - développée par le CETE Méditerranée et mise en œuvre conjointement avec le CETMEF, basée sur une approche topographique uniquement) principalement pour les têtes de bassin non couvertes.

L'enveloppe approchée des inondations potentielles est construite en assemblant l'ensemble de la connaissance cartographique disponible concernant les zones inondables (sur le même principe que la « couche de synthèse » de Cartorisque).

La couche géographique concernant les submersions marines assemble les trois types d'informations ci-dessous et utilise les éléments sur la prise en compte du changement climatique :

- l'étude de référence au niveau national « Vulnérabilité du territoire national aux risques littoraux » (VTNRL) diffusée aux services dans le cadre de l'application de la circulaire du 7 avril 2010 relative aux mesures à prendre suite à la tempête Xynthia du 28 février 2010 ;
- la synthèse de l'ensemble de la connaissance cartographique concernant les zones inondables par submersions marines disponible au format SIG au sein des services de l'État ;
- des informations sur la géologie (couche des alluvions maritimes récentes) disponibles sur le littoral.

### **Recensement de population (Insee)**

Jusqu'en 1999, les populations légales étaient déterminées à l'occasion de chaque recensement général de la population. Entre deux recensements elles pouvaient être modifiées par un recensement complémentaire.

À partir de 2006, la nouvelle méthode de recensement basée sur des enquêtes de recensement annuelles permet de calculer chaque année des populations légales actualisées. Ainsi par exemple, fin 2009 ont été publiées les populations légales qui sont entrées en vigueur le 1er janvier 2010. Ces populations légales étaient millésimées 2007 car elles étaient calculées à partir des informations collectées lors des enquêtes de recensement de 2005 à 2009 et ramenées à une même date : celle du milieu de la période. L'égalité de traitement entre les communes est ainsi assurée.

Le nouveau dispositif de recensement qui actualise les populations légales chaque année se substitue aussi aux recensements complémentaires.

### **Contours des Iris 2008 (IGN)**

Les îlots regroupés pour l'information statistique (Iris) sont un zonage de l'Insee, constitués de sommes d'îlots et regroupant environ 2 000 habitants. Ce zonage, infra-communal, pour lequel est diffusé le recensement de population, est une partition des communes en quartier. Seules les communes de plus de 10 000 habitants et la plupart des communes de 5 000 à 10 000 habitants sont découpées en Iris.

En 2008, une retouche très partielle du découpage est intervenue pour prendre en compte les évolutions importantes de la voirie et de la démographie.

La France compte environ 16 100 Iris dont 650 dans les DOM. Pour les 34 800 communes sans Iris, on considère que le contour de la commune est un Iris. Au total, le fond cartographique est donc constitué de 50 880 objets.

Le contour utilisé est le contour des Iris, référence 2008.

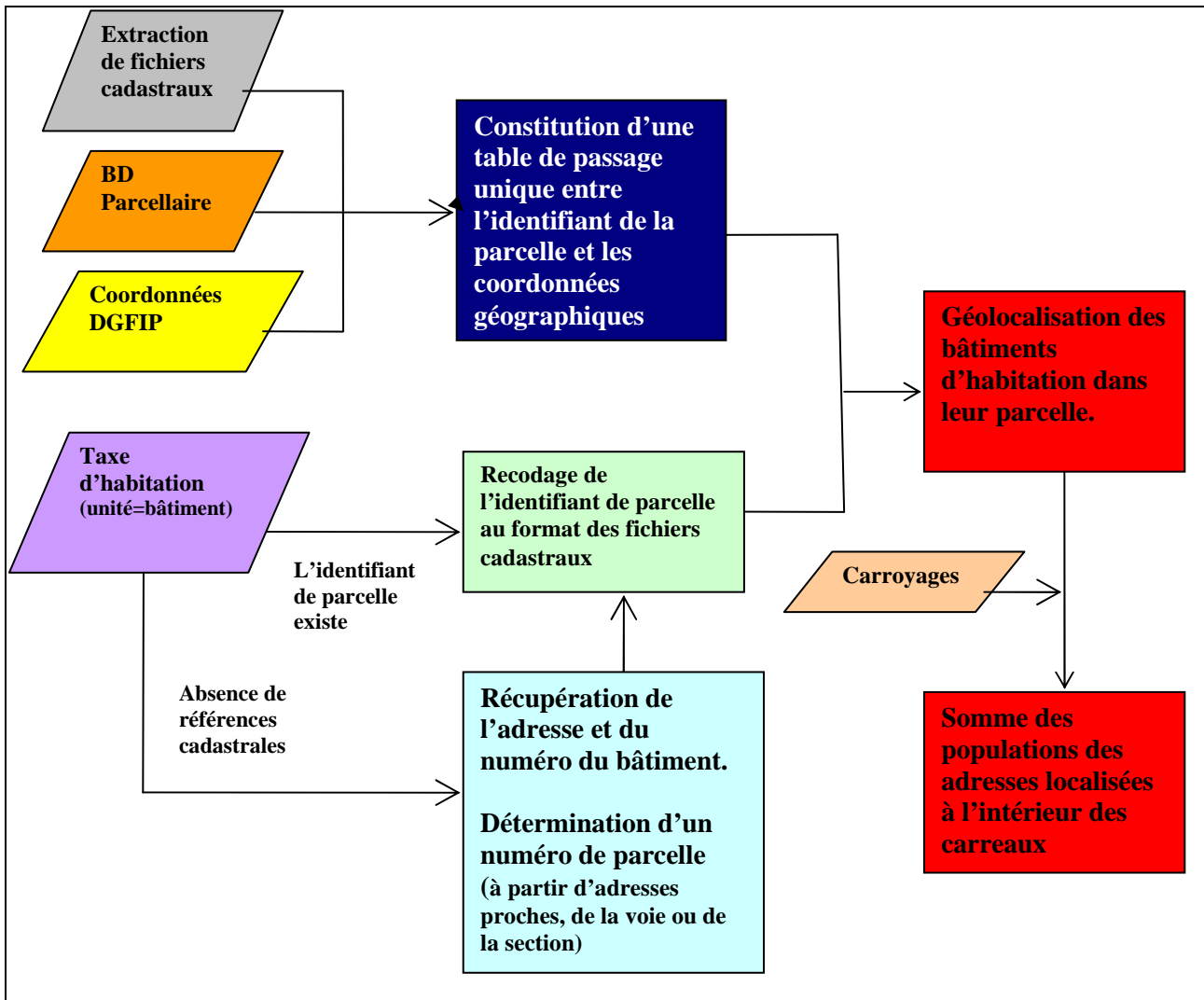
### **Données carroyées de la population (Insee, Revenus fiscaux localisés, 2009)**

Ces données, produites par l'Insee, dénombrent la population fiscale par carreau de 200 mètres X 200 mètres au 31 décembre de l'année de référence. Elles sont produites à partir des revenus fiscaux localisés (RFL), établis par rapprochement des fichiers des déclarations de revenus des personnes physiques et des fichiers de la taxe d'habitation (TH), fournis à l'Insee par la Direction générale des finances publiques (DGFIP).

Les ménages fiscaux, ainsi identifiés, sont localisés par le biais du local soumis à la taxe d'habitation, auquel chaque ménage est rattaché. À chaque ménage une position géographique précise est affectée à partir de la localisation de la parcelle cadastrale sur laquelle est situé le local. Les coordonnées de la parcelle proviennent des fichiers cadastraux de la DGFIP ou d'un rapprochement avec la BD PARCELLAIRE® de l'IGN. Une fois les ménages positionnés sur le territoire, le nombre de personnes rattachées par ménage est comptabilisé par carreau de 200 mètres x 200 mètres.

Les régions Guyane, Guadeloupe et Mayotte, ne pouvant pas être traitées actuellement, ne disposent pas de données carroyées de population.

**Schéma 1 : Résumé des traitements pour la constitution de bases de populations carroyées**



Pour plus de détails, vous pouvez consulter le site de l'Insee : [http://www.insee.fr/fr/themes/detail.asp?reg\\_id=0&ref\\_id=donnees-carroyees&page=donnees-detaillees/donnees-carroyees/donnees-carroyees-200m.htm](http://www.insee.fr/fr/themes/detail.asp?reg_id=0&ref_id=donnees-carroyees&page=donnees-detaillees/donnees-carroyees/donnees-carroyees-200m.htm)

D'autres travaux sont actuellement menés pour produire des données carroyées de population à partir du recensement associé à des sources administratives. Les bases ne sont pas encore disponibles mais ces méthodes sont présentées en annexe 2.

**Données BD TOPO® (IGN)**

La BD TOPO® est la composante topographique des bases de données du référentiel à grande échelle (RGE®).

La BD TOPO® est une base de données vectorielles, de précision métrique, disponible sur l'ensemble du territoire français. Elle détaille différents thèmes du territoire : réseau routier, réseau ferré, réseau hydrographique, bâtiments, découpage administratif...

Le thème des bâtiments détaille les bâtiments à caractère industriel, des bâtiments à caractère remarquable, des bâtiments indifférenciés. La couche « Bâtiment indifférencié » du thème « Bâtiment » est utilisée pour réaliser les estimations de population et de logements.

Le bâti indifférencié de la BD TOPO® comprend les bâtiments de plus de 20 m<sup>2</sup> ne possédant pas de fonction particulière. Il inclut les bâtiments d'habitation mais aussi les bergeries, les bories, les bungalows, les bureaux, les chalets, les bâtiments d'enseignement, les garages individuels, les bâtiments hospitaliers,

les immeubles collectifs, les lavoirs couverts, les musées, les prisons, les refuges, les villages de vacances.

### **CORINE Land Cover 2006 (UE-SOeS)**

La base de données géographiques CORINE Land Cover est un inventaire biophysique européen de l'occupation des sols. Elle est issue de l'interprétation visuelle d'images satellitaires, avec des données complémentaires d'appui. L'échelle de production est le 1/100 000 et la surface de la plus petite unité cartographiée est de 25 hectares.

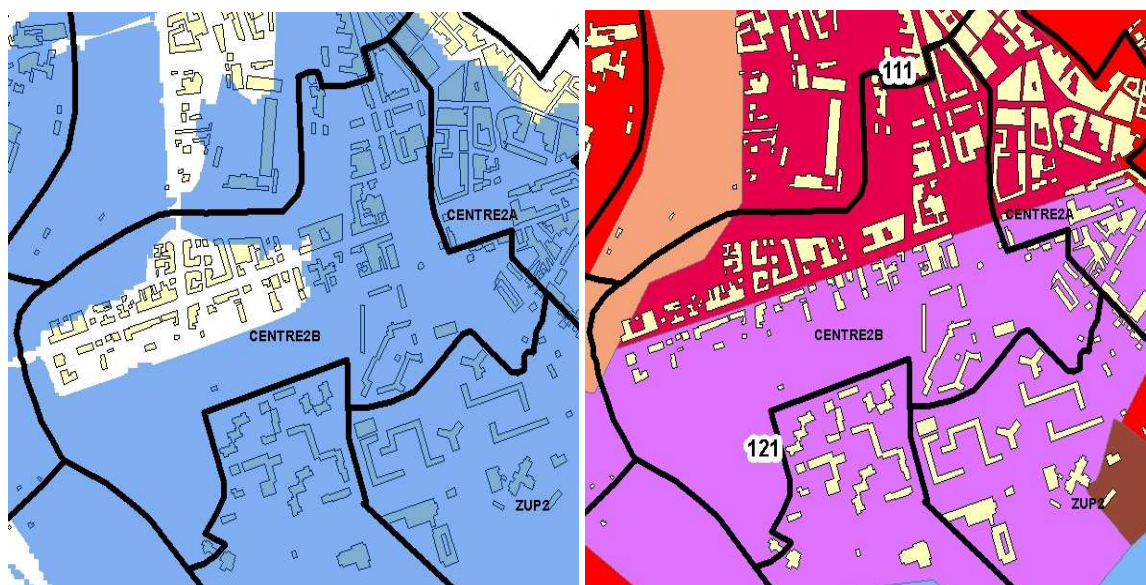
Cette base est constituée selon une nomenclature standard et hiérarchisée sur 3 niveaux imbriqués. Cette nomenclature comporte, au 1er niveau, 5 grands types d'occupation du sol : les territoires artificialisés, les territoires agricoles, les forêts et milieux semi-naturels, les zones humides et les surfaces en eau. Elle est constituée de 44 postes au niveau 3.

Les territoires artificialisés sont décomposés en zones urbanisées (11) et en zones industrielles ou commerciales ou réseaux de communications (12).

Les zones urbanisées caractérisent des espaces structurés autour de bâtiments ; elles sont elles-mêmes détaillées en 2 catégories : le tissu urbain continu (111) et le tissu urbain discontinu (112).

Les zones industrielles ou commerciales et réseaux de communications sont quant à eux détaillés en 4 catégories : les zones industrielles et commerciales (121), les réseaux routiers et ferroviaires (122), les zones portuaires (123) et les aéroports (124).

Pour illustrer ces sources, voici un exemple :



**Contours des zones inondables et bâtiments**

**Occupation du sol de l'Iris CENTRE 2B**

Cet Iris (CENTRE 2B) situé dans un centre-ville est en grande partie en zone inondable (couleur bleue, sur l'image de gauche) avec une majorité de bâtiments de type indifférencié.

La partie nord de l'Iris est en tissu urbain continu (modalité 111) tandis que sa partie sud est en zone industrielle ou commerciale (modalité 121). Une petite partie à l'ouest de l'Iris est en espaces verts urbains (modalité 141).

Dans les encadrés situés dans les pages suivantes du document, figurent les résultats de l'application de :

- la méthode utilisant les carroyages de population sur les zones inondables de juillet 2012, constituées des zones inondables disponibles sur Cartorisque et des données transmises au SOeS par les DREAL ;
- la méthode BD TOPO® utilisant l'enveloppe approchée des inondations potentielles (EAIP) et ayant permis de réaliser l'EPRI.

## 1. Estimations à partir des données carroyées de population

### 1.1 Données utilisées

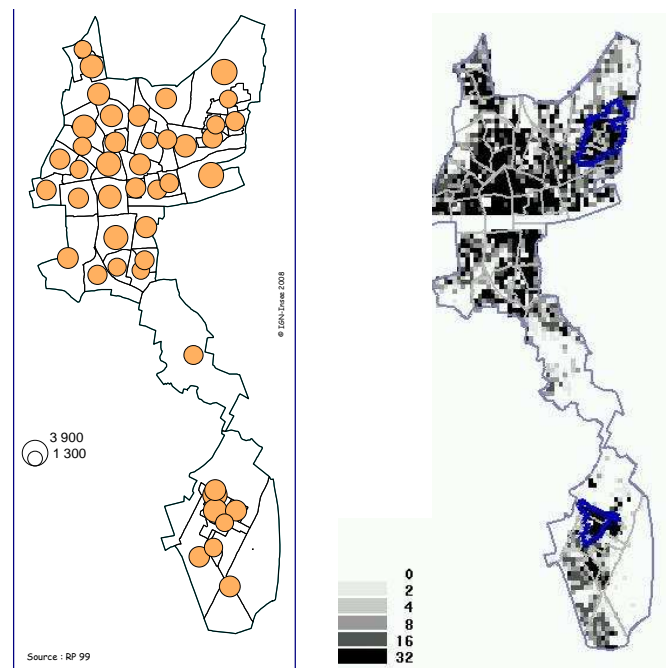
Cette méthode utilise plusieurs bases de données présentées dans le paragraphe « Données utilisées » :

- les contours des zones inondables ;
- le nombre de logements et la population par Iris, du recensement de population (RP) 2006 de l'Insee ;
- les contours Iris 2008 de l'IGN ;
- les données carroyées de population de l'Insee.

### 1.2 Analyse des données carroyées de population

Les données carroyées de population, de l'Insee, présentent la population fiscale par carreaux de 200 mètres. Cette donnée détaille plus finement la répartition de population sur le territoire, mais n'est pas comparable aux données de recensement de population, établi par sondage.

#### Carte 1 : Comparaison du recensement et des données carroyées de population sur la commune



Source : Insee, Recensement de population & RFL.

L'exemple ci-dessus montre que la répartition de la population est plus détaillée avec les données carroyées de population (à droite) qu'avec les données du recensement à l'Iris (à gauche).

En cumulant les populations par carreaux de 200 mètres, l'estimation est de 61 653 533 habitants en métropole. Moins de 1 % de la population municipale (au RP 2008) est manquante dans la base carroyée.



Il reste une imprécision sur environ 0,3 % de la population comptabilisée, particulièrement dans les régions du sud.

En raison du champ couvert par les sources fiscales, rappelons que certaines populations non soumises à la taxe d'habitation ne sont pas présentes.

La population est plus faible que celle du recensement dans la majorité des régions, plus encore en Corse (- 13 %), dans le Limousin (- 5 %) ou en Alsace et en Lorraine (- 3 %).

Seules les régions Île-de-France, PACA et Languedoc-Roussillon ont une population plus nombreuse dans la base carroyée.

**Tableau 1 : Somme des populations carroyées par région**

Région	Population municipale 2008	Somme des populations carroyées	Ecart de population		Population avec localisation imprécise	Part
Île-de-France	11 659 260	11 741 942	82 682	0,7%	33 822	0,29%
Champagne-Ardenne	1 338 004	1 306 852	-31 153	-2,4%	6 286	0,48%
Picardie	1 906 601	1 886 725	-19 876	-1,1%	5 890	0,31%
Haute-Normandie	1 825 667	1 806 449	-19 218	-1,1%	8 243	0,46%
Centre	2 531 588	2 511 973	-19 615	-0,8%	4 968	0,20%
Basse-Normandie	1 467 522	1 442 071	-25 451	-1,8%	4 214	0,29%
Bourgogne	1 638 588	1 601 166	-37 422	-2,3%	3 449	0,22%
Nord-Pas-de-Calais	4 024 490	3 985 158	-39 333	-1,0%	12 702	0,32%
Lorraine	2 346 361	2 275 454	-70 908	-3,1%	9 260	0,41%
Alsace	1 837 087	1 781 973	-55 114	-3,1%	7 298	0,41%
Franche-Comté	1 163 931	1 149 025	-14 906	-1,3%	3 585	0,31%
Pays de la Loire	3 510 170	3 509 817	-353	0,0%	7 729	0,22%
Bretagne	3 149 701	3 130 194	-19 508	-0,6%	6 347	0,20%
Poitou-Charentes	1 752 708	1 730 232	-22 476	-1,3%	5 071	0,29%
Aquitaine	3 177 625	3 136 598	-41 028	-1,3%	11 771	0,38%
Midi-Pyrénées	2 838 228	2 775 161	-63 068	-2,3%	10 780	0,39%
Limousin	740 743	706 111	-34 632	-4,9%	1 500	0,21%
Rhône-Alpes	6 117 229	6 100 814	-16 415	-0,3%	21 090	0,35%
Auvergne	1 341 863	1 315 762	-26 102	-2,0%	2 360	0,18%
Languedoc-Roussillon	2 581 718	2 588 533	6 815	0,3%	15 431	0,60%
Provence-Alpes-Côte d'Azur	4 882 913	4 904 026	21 113	0,4%	33 909	0,69%
Corse	302 966	267 502	-35 465	-13,3%	2 066	0,77%
France métropolitaine	62 465 709	61 653 533	-481 430	-0,8%	217 765	0,35%

Source : Insee, RFL, 2009 - Insee, recensement de population, 2008.

Concernant les DOM, les données carroyées de population ne sont disponibles que sur la Réunion et la Martinique.

Pour la Réunion, 786 174 habitants sont dénombrés (dont 4,4 % avec une localisation imprécise), soit 22 000 habitants de moins qu'au recensement 2008 (- 2,8 %). Pour la Martinique, ce sont 371 653 habitants (5,2 % imprécis), 26 000 habitants de moins qu'au RP (- 7 %).

Un dénombrement a été également effectué sur les aires urbaines 2010.

L'écart entre le recensement et les carroyages est de +1,4 % : les populations carroyées sont un peu plus sous estimées sur les aires urbaines que sur l'ensemble des communes.

Les écarts s'échelonnent de - 25 % (Uzerche) à + 17 % (Agde).

Il en ressort une surestimation du nombre d'habitants (entre + 10 % et + 20 %) dans des aires urbaines comme Saint-Tropez, Biscarrosse, Le Grau-du-Roi ou Agde, et dans une moindre mesure Nice, Toulon et Avignon.

Les aires urbaines dans lesquelles les populations carroyées sont supérieures à celles du recensement ont un point commun : elles sont plutôt touristiques et littorales.

La forte présence des résidences secondaires explique en partie ces écarts.

À l'inverse, des aires urbaines de plus petite taille ont une population carroyée inférieure à celle du recensement. C'est également le cas des aires urbaines corses Bastia et Ajaccio (- 10 % environ) mais

aussi de Poitiers et Nancy (- 7 %), Reims, Limoges, Dijon, Strasbourg, Amiens, Besançon ou Montpellier (- 5 %).

### 1.3. Estimation de population en zone inondable

Afin d'estimer la population en zone inondable, les données carroyées de population sont utilisées en entrée. Elles sont croisées avec les contours des Iris et des zones inondables afin d'être agrégées.

Pour répartir la population fiscale des carreaux qui sont à cheval sur plusieurs contours (Iris et/ou zones inondables), la population du carreau est répartie dans chacune des zones au prorata surfacique. La population se répartit donc proportionnellement à la superficie de chaque zone sur laquelle le carreau se situe.

Les données de population ainsi réparties en fonction de leur appartenance à un Iris et à une zone inondable, peuvent être agrégées à l'Iris afin de connaître la population fiscale totale et celle en zone inondable de chaque Iris.

Afin d'estimer la population et non la population fiscale en zone inondable, il est proposé de caler les données de population fiscale à l'Iris sur la population du recensement de population à l'Iris, de la manière suivante :

$$PopZI_i = PopZI_i^{RFL} \cdot \frac{Pop_i^{RP}}{Pop_i^{RFL}}, \text{ avec :}$$

- $PopZI_i$  : la population estimée en zone inondable sur l'Iris  $i$  ;
- $PopZI_i^{RFL}$  la population comptabilisée en zone inondable dans l'Iris  $i$  à partir des données carroyées ;
- $Pop_i^{RP}$  la population de l'Iris  $i$ , selon le recensement ;
- $Pop_i^{RFL}$  la population comptabilisée dans l'Iris  $i$  à partir des données carroyées.

**Tableau 4 : Estimations de population en zone inondable dans les communes de plus de 10 000 habitants**

Commune	codegeo	Population RP 2008	Population carroyée	Ecart relatif entre le RP et la population carroyée	Population carroyée en zone inondable	Population carroyée en zone inondable après calage	Ecart relatif après calage
Tours	37261	135 480	117 645	15%	71 705	81 814	14%
Orléans	45234	113 257	105 451	7%	23 293	24 430	5%
Saint-Pierre-des-Corps	37233	15 528	15 322	1%	15 322	15 528	1%
La Riche	37195	10 080	9 684	4%	9 052	9 353	3%
Blois	41018	46 834	44 259	6%	7 893	8 305	5%
Olivet	45232	20 103	18 582	8%	5 956	6 235	5%
Vendôme	41269	16 763	16 411	2%	4 460	4 753	7%
Chartres	28085	39 159	37 979	3%	2 284	2 367	4%
Amboise	37003	12 436	11 897	5%	2 127	2 102	-1%
Dreux	28134	31 212	30 589	2%	2 077	2 229	7%
Gien	45155	15 337	14 464	6%	1 997	1 960	-2%
Châteauroux	36044	46 026	45 736	1%	1 862	1 800	-3%
Nogent-le-Rotrou	28280	11 124	10 152	10%	1 698	2 022	19%
Montlouis-sur-Loire	37156	10 444	10 424	0%	1 351	1 307	-3%
Romorantin-Lanthenay	41194	17 395	17 619	-1%	1 245	1 219	-2%
Issoudun	36088	13 477	12 484	8%	1 085	1 118	3%
Saint-Avertin	37208	14 002	14 202	-1%	905	924	2%
Saint-Jean-de-Braye	45284	18 933	19 525	-3%	849	777	-8%
Fondettes	37109	10 270	10 466	-2%	664	673	1%
Châteaudun	28088	13 905	12 524	11%	589	585	-1%
Saint-Cyr-sur-Loire	37214	15 841	16 310	-3%	525	499	-5%
Saint-Jean-de-la-Ruelle	45285	16 669	16 591	0%	410	423	3%
Joué-lès-Tours	37122	35 839	34 951	3%	125	113	-10%
Vernouillet	28404	11 794	11 587	2%	34	35	1%

Source : Insee, RFL, 2009 - Insee, recensement de population, 2008 - DGPR-DREAL, AZI et PPRI, 2010.  
Traitements : SOeS, 2012.

Dans les principales communes de la région, les deux écarts relatifs, entre la population du RP et la population fiscale et entre la population fiscale et la population estimée en zone inondable, sont relativement proches : le calage compense presque entièrement l'erreur due à l'estimation à l'Iris.

Mais les tendances peuvent diverger : à Gien et Châteauroux, la population au recensement est supérieure à celle des carroyages mais le calage fait baisser la population en zone inondable.

On peut également estimer, le nombre de logements en zone inondable en se basant sur l'hypothèse que la taille des ménages (population moyenne par logement) est uniforme sur chaque Iris. Ainsi, on peut estimer le nombre de logement en zone inondable par Iris de la manière suivante :

$$\text{LogZI}_i = \text{PopZI}_i \cdot \frac{\text{Log}_i^{\text{RP}}}{\text{Pop}_i^{\text{RP}}}, \text{ avec :}$$

- $\text{LogZI}_i$  : le nombre de logements estimés en zone inondable sur l'Iris  $i$ ,
- $\text{PopZI}_i$  : la population estimée en zone inondable sur l'Iris  $i$ ,
- $\text{Log}_i^{\text{RP}}$  : le nombre de logement de l'Iris  $i$ , selon le recensement ;
- $\text{Pop}_i^{\text{RP}}$  la population de l'Iris  $i$ , selon le recensement.

### Estimation de la population en zone inondable à partir des données carroyées de population

En superposant les données carroyées de population (RFL 2009) et les zones inondables de juillet 2012 constituées par le SOeS, la population des carreaux est répartie au prorata surfacique des surfaces situées en zone inondable, puis calée sur celle du recensement de population. L'estimation est alors de :

- 6 238 932 personnes en zone inondable en métropole, soit 10 % de la population ;
- 138 494 personnes en zone inondable à la Réunion, soit 17 % de la population.

Source : DGPR-DREAL, AZI et PPRI, 2012 – Insee, RFL, 2009 - ©IGN, Contours Iris, 2008. Traitements SOeS, 2012.

## 2. Méthodes d'estimation « Popéval » - BD TOPO®

### 2.1. Données utilisées

Les méthodes utilisent plusieurs bases de données présentées dans le paragraphe « Données utilisées » :

- les contours des zones inondables ;
- le nombre de logements et la population par Iris, du recensement de population 2006 de l'Insee ;
- les contours des Iris ;
- le bâti indifférencié de la BD TOPO® ;
- CORINE Land Cover, 2006.

### 2.2. Traitements des données SIG

Dans un premier temps, l'objectif est de repérer parmi les bâtiments de type indifférencié, de la BD TOPO®, les bâtiments d'habitation. Il y a en effet un risque de dénombrer un grand nombre de logements dans de vastes bâtiments qui ne sont pas d'habitation. Cette opération repose sur des traitements géographiques.

**Sélection** : Le but de la sélection est de conserver un maximum de bâtiments susceptibles de contenir des logements (résidences principales et secondaires).

**Sont exclus les bâtiments de plus de 100 mètres de haut ou d'une surface inférieure à 20 m<sup>2</sup>** (on considère arbitrairement que les bâtiments ont une surface minimum de 20 m<sup>2</sup>, en cohérence avec la définition du bâti indifférencié de la BD TOPO®).

**Ceux de moins de 2 mètres de hauteur sont mis à une hauteur de 3 mètres.** En effet, les bâtiments de moins de 3 mètres recensés dans la base BD TOPO® ne sont pas exclus du fait d'imprécisions sur le

paramètre « hauteur » des polygones. Certains polygones, constitués pourtant d'habitations, ont une hauteur à 0 mètre dans la BD TOPO®. En les ôtant, il y a sous-estimation des logements et donc du nombre de personnes en zone inondable.

**Découpe :** Puis, les contours des zones inondables et des Iris sont croisés avec les bâtis retenus de la BD TOPO®.

Un bâtiment peut être ainsi découpé en plusieurs polygones s'il est situé sur plusieurs Iris et sur une zone inondable.

À l'issue des traitements géographiques, une base ayant la forme suivante est obtenue :

Identifiant bâtiment	HAUTEUR	Surface	CODE Iris	CODE CLC	ZI
00000001	5	673	452340000	111	1
00000001	5	240	452340000	111	0
.....					

Dans cette table, le bâtiment 00000001 situé dans l'Iris 452340000 est haut de 5 mètres et situé dans le tissu urbain continu (Code 111). La partie en zone inondable de ce bâtiment a une emprise au sol de 673 m<sup>2</sup>, celle en dehors de la zone inondable a une emprise de 240 m<sup>2</sup>.

Dans la suite du document, on ne parlera plus de bâtiments mais de polygones dans les Iris.

Pour tous les polygones, on estime une surface développée (somme des surfaces de tous les niveaux) en considérant en moyenne des niveaux de 3 mètres de hauteur.

$$\text{Surface développée} = \frac{\text{Surface} \cdot \text{Hauteur}}{3}$$

### 2.3. Présentation des méthodes

L'approche la plus simple, voire la plus intuitive, est d'estimer un nombre de logements dans une zone à partir de deux informations connues : le nombre total de logements recensés dans l'Iris et les surfaces développées des polygones présents.

#### 2.3.1. Méthode simple (au prorata de la surface développée)

Cette méthode estime, par Iris, le nombre de logements en zone inondable proportionnellement à la surface développée des polygones qui s'y trouvent.

Dans un Iris donné, le nombre de logements en zone inondable est déterminé de la manière suivante :

$$\text{Log}_{ZI} = \text{Log}_{IRIS} \times \frac{\sum_{\text{polygones}} SD_{ZI}}{\sum_{\text{polygones}} SD_{IRIS}}$$

où  $\text{Log}_{ZI}$  et  $\text{Log}_{IRIS}$  sont le nombre de logements respectivement en zone inondable et dans l'ensemble de l'Iris,  $SD_{ZI}$  et  $SD_{IRIS}$  la surface développée des polygones situés respectivement en zone inondable et dans l'ensemble de l'Iris.

En supposant que la densité de population par logement est uniforme pour l'ensemble des bâtiments de l'Iris, la population en zone inondable se déduit du nombre de logements, selon la formule suivante :

$$\text{Pop}_{ZI} = \text{Log}_{ZI} \times \frac{\text{Pop}_{IRIS}}{\text{Log}_{IRIS}}$$

où  $\text{Pop}_{IRIS}$  et  $\text{Log}_{IRIS}$  sont la population et le nombre de logements au recensement.

Ainsi, comme l'indique la formule ci-dessus, la part de population en zone inondable ( $Pop_{ZI} / Pop_{IRIS}$ ) coïncide avec celle des logements ( $Log_{ZI} / Log_{IRIS}$ ) à l'échelle des Iris.

Conformément au traitement amont des bâtiments, détaillé au 2.2. « Traitements des données SIG », un bâtiment à cheval sur deux Iris et partiellement en zone inondable est découpé en 4 polygones selon son appartenance à un Iris et à une zone inondable.

On obtient ainsi 4 surfaces développées : SD1 et SD2 dans l'Iris 1 et SD3 et SD4 dans l'Iris 2.

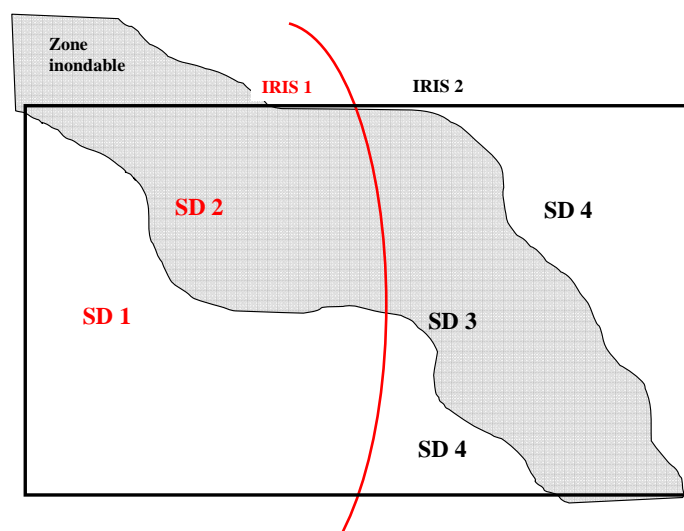
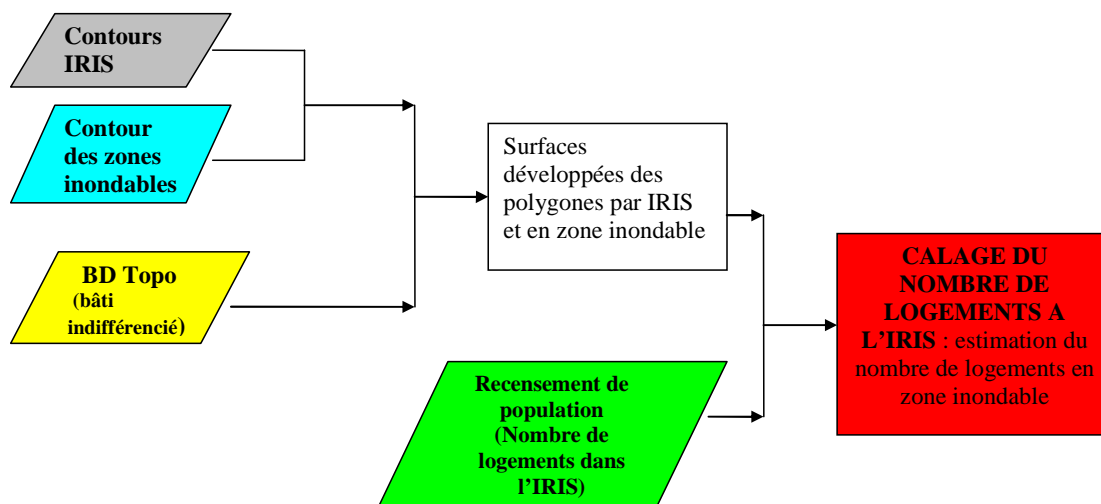


Schéma 2 : Résumé des traitements de la méthode simple BD TOPO®



Relativement simple d'application, elle présente l'inconvénient d'attribuer une même densité de logements aux bâtiments présents dans un Iris quelles que soient les caractéristiques de la zone.

De plus, on fait l'hypothèse que la densité de logements (relativement à la surface développée) est identique dans les zones inondables et en dehors, ce qui est loin d'être le cas surtout dans certaines zones littorales.

**Application de la méthode simple BD TOPO® aux enveloppes approchées des inondations potentielles : résultats de l'Évaluation préliminaire des risques d'inondation.**

Sont situés dans l'enveloppe cours d'eau :

- 16,8 millions de résidents permanents en métropole, soit 27 % de la population ;
- 326 000 résidents permanents dans les DOM (Guyane, Guadeloupe, Martinique, Réunion, Mayotte), soit 18 % de la population.

Sont situés dans l'enveloppe submersion marine :

- 1,4 million de résidents permanents en métropole, soit 2 % de la population de métropole ou 23 % de la population permanente des communes littorales ;
- 92 000 résidents permanents dans les DOM (Guyane, Guadeloupe, Martinique, Réunion, Mayotte), soit 5 % de la population.

Note : pour ces calculs, un traitement complémentaire a été réalisé pour trier les polygones de bâti de la BD TOPO® : les polygones ont été découpés avec les zones de la classe « surface activité » de la BD TOPO®, et les portions de polygones contenus dans cette classe ont été exclus.

Source : CETE Méditerranée, EAIP, 2011 - Insee, RP 2006 - ©IGN, BD TOPO®, 2009 - ©IGN, Contours Iris®, 2008. Traitements, 2011.

**2.3.2. Modèles économétriques linéaires**

La méthode d'estimation simple, se base sur le postulat que la densité d'habitants par logement est identique au sein d'un Iris. Pour améliorer les estimations, il faudrait appliquer une densité différente selon l'occupation du sol. En effet, dans un bâtiment situé en zone industrielle, il y a généralement moins de logements par m<sup>2</sup> de surface développée que dans le centre d'une grande ville. D'autant qu'il est plus probable que ce bâtiment ne soit pas d'habitation.

De plus, seul un tiers des Iris métropolitains est recouvert par un seul poste d'occupation des sols selon CORINE Land Cover. Les Iris entièrement composés de tissu urbain continu ou discontinu se concentrent dans certaines zones comme l'Île-de-France.

La modélisation économétrique permet d'estimer les densités de logements en fonction de l'occupation des sols ; elle doit être utilisée avec prudence, en choisissant les variables explicatives de manière appropriée.

**2.3.2.1. Estimation du modèle**

Dans la continuité de la méthode simple, le nombre de logements dans un Iris est exprimé en fonction des surfaces développées des polygones selon la nature d'occupation du sol.

Les 4 types d'occupation du sol pris en compte sont :

- le tissu urbain continu (111) ;
- le tissu urbain discontinu (112) ;
- les modalités 121 (zones industrielles et commerciales), 123 (zones portuaires) et 133 (chantiers) : ces zones se distinguent en effet généralement des autres types (zones agricoles, forêts, zones humides...) par la présence de bâtiments spacieux qui ne sont pas d'habitation ;
- et une catégorie « Autres ».

$$NbLog_i = a_{111} SD_{i,111} + a_{112} SD_{i,112} + a_{zoncom} (SD_{i,121} + SD_{i,123} + SD_{i,133}) + a_{autres} SD_{i,autres} + \mathcal{E}_i$$

où **NbLog<sub>i</sub>** est le nombre de logements dans l'Iris **i**, **SD<sub>i,k</sub>** est la surface développée des polygones de l'Iris **i** dans la modalité **k** d'occupation du sol, **a<sub>k</sub>** est la densité de logements dans la modalité **k** d'occupation du sol, **a<sub>zoncom</sub>** celle pour les modalités cumulées 121,123 et 133. et **a<sub>autres</sub>** celle des autres types d'occupation du sol.

**ℰ<sub>i</sub>** est la perturbation du modèle, d'espérance nulle et de variance constante, la covariance entre deux observations différentes est nulle.

### 2.3.2.2. Application du modèle

Une fois les coefficients du modèle  $\hat{\alpha}$  déterminés, dans la phase d'estimation, on applique ces derniers aux surfaces développées en zone inondable différenciées selon l'occupation du sol.

Afin d'être cohérent avec les données du recensement, on applique un coefficient de calage, ratio entre le nombre de logements au recensement et celui estimé par la méthode dans la totalité de l'Iris.

**Si l'un des coefficients estimés  $\hat{\alpha}$  est négatif, il est mis automatiquement à zéro afin de ne pas estimer un nombre de logements négatif dans l'Iris.**

$$\text{NbLog}_{i,ZI} \text{ estimé} = [\hat{\alpha}_{111} \text{SD}_{i,111,zi} + \hat{\alpha}_{112} \text{SD}_{i,112,zi} + \hat{\alpha}_{\text{zoncom}} (\text{SD}_{i,121,zi} + \text{SD}_{i,123,zi} + \text{SD}_{i,133,zi}) + \hat{\alpha}_{\text{autres}} \text{SD}_{i,\text{autres},zi}] \times \left( \frac{\text{NbLog}_i}{\text{NbLog}_i \text{ estimé}} \right)$$

**NbLog<sub>i,ZI</sub> estimé** est le nombre de logements dans l'Iris *i* en zone inondable et **NbLog<sub>i</sub> estimé** est le nombre de logements dans l'Iris *i* estimé en appliquant les coefficients  $\hat{\alpha}$ .

**SD<sub>i,k,ZI</sub>** est la surface développée totale des polygones de l'Iris *i* en zone inondable pour la modalité *k* d'occupation du sol.

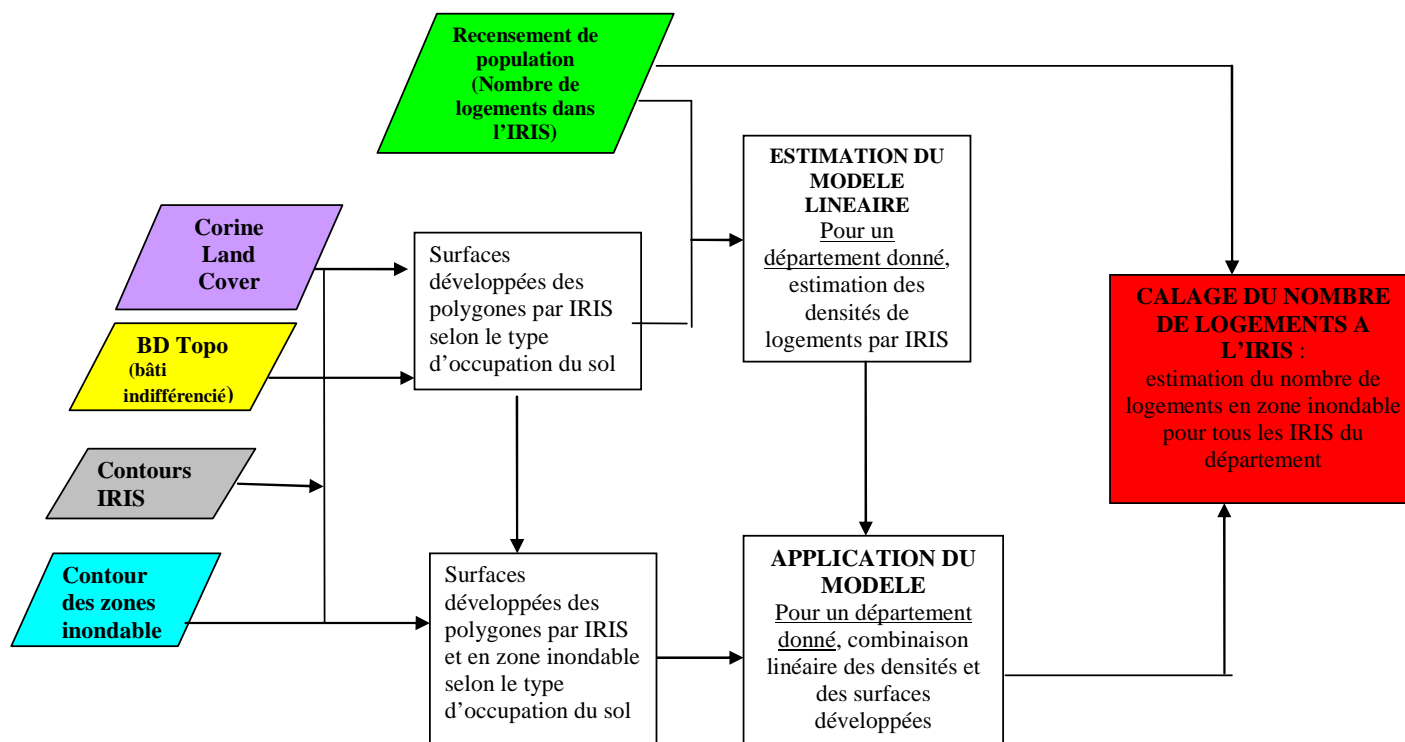
Cette modélisation se justifie pour des Iris avec une occupation du sol très diversifiée. Elle n'a aucun intérêt pour les Iris qui n'ont qu'une seule nature d'occupation du sol (on n'aurait qu'une seule variable explicative dans le modèle). Dans ce cas, la méthode d'estimation simple est suffisante.

Mais, il est difficile d'appliquer un même modèle à tous les Iris de métropole voire d'outre-mer. En effet, la diversité de nature des Iris et les spécificités locales ne permettent pas d'attribuer partout une même densité de logements.

D'où la nécessité de différencier le même modèle de régression selon des critères pertinents sur les Iris.

**Cette méthode est adaptée dans le cas où la zone sur laquelle se fait l'estimation est relativement restreinte et comporte plusieurs types d'occupation du sol (ex : une commune de moins de 10 000 habitants, un canton...).**

**Schéma 3 : Résumé des traitements de la méthode BD TOPO® avec modélisation**



## 2.4 Qualité de l'estimation

### 2.4.1 Méthode simple

Plusieurs éléments influent sur l'estimation du nombre de logements en zone inondable par la méthode simple. Il convient d'être vigilant sur certains points :

- Tout d'abord, les surfaces développées sont automatiquement cumulées sur les Iris. Rapportées au nombre de logements connu dans l'Iris, elles traduisent des surfaces développées moyennes qui peuvent s'élever à plusieurs milliers de m<sup>2</sup>. C'est le cas d'Iris peu peuplés avec de vastes bâtiments.  
Plus cet indicateur est élevé, plus l'estimation inclut des bâtiments qui ne sont pas exclusivement d'habitation. Si de surcroît, le nombre total de logements est élevé, l'estimation sera d'autant plus imprécise.
- Le nombre de logements en zone inondable est calculé à partir des surfaces développées en zone inondable. Ainsi, la part des surfaces développées des bâtiments en zone inondable est également un indicateur à observer.  
Si cette part est très faible (ou très forte) dans l'Iris, le nombre de logements estimés en zone inondable sera proche de zéro (respectivement proche du nombre total de logements) : le risque d'erreur est assez faible. Il est en revanche plus élevé si cette part avoisine les 50 %.

### 2.4.2 Modèles économétriques

La qualité de l'estimation est liée en premier lieu à la spécification du modèle (voir annexe 1).

Les modèles départementaux ont un bon pouvoir explicatif ( $R^2$  compris entre 0,8 et 0,96) excepté ceux du Languedoc-Roussillon (entre 0,7 et 0,8 dans l'Hérault, l'Aude et le Gard).

Mais ils semblent inappropriés pour plusieurs départements en raison d'une plus forte densité de logements dans les zones commerciales de l'Aude, de l'Hérault et des Pyrénées-Orientales ou de la faible proportion des autres types d'occupation du sol comme dans certains départements franciliens.

La modélisation pose d'autres problèmes purement statistiques.

#### Variance des résidus

Dans tout modèle économétrique, l'erreur issue des résidus du modèle n'est pas constante. Elle croît plutôt en fonction de la variable à expliquer, ici les logements (hétéroscédasticité des résidus). Il n'y a pas obligation de corriger l'hétéroscédasticité lors de la modélisation puisqu'elle ne biaise pas l'estimation des coefficients. En revanche, ces derniers ne sont plus de variance minimum : l'estimateur n'est plus optimal car les Moindres Carrés Ordinaires font l'hypothèse que les résidus connaissent un niveau de variance constant.

#### Autocorrélation des résidus

L'estimation de logements par un modèle économétrique entraîne inévitablement ce qu'on appelle une autocorrélation spatiale des résidus. Les modèles produisent des résidus dont le signe varie selon les zones mais qui ont plutôt tendance à se regrouper (autocorrélation positive). Pour simplifier, l'estimation produit une erreur qui n'est pas répartie aléatoirement mais a tendance à être plus ou moins concentrée géographiquement.

**Si ces éléments ont été évalués, ils n'ont pas été corrigés dans les modèles, compte tenu des difficultés pratiques de mise en œuvre (voir annexe 2).**

## 3. Comparaison des deux méthodes

Pour comparer les méthodes d'estimation de population et de logements à partir des données carroyées de population et à partir de la BD TOPO®, une analyse a été menée spécifiquement sur la région Centre, en prenant en compte les zones inondables (AZI et PPRI).



### 3.1. Tests sur la région Centre

En région Centre, la population s'élève en 2006 à 2 513 500 habitants pour 1 258 607 logements. Les données carroyées de population dénombrent une population inférieure (2 511 973 habitants), soit un écart de - 0,1 % par rapport au recensement.

**Tableau 2 : Comparaison des estimations de population et de logements en région Centre**

	Méthode BD TOPO® simple	Méthode à partir des données carroyées	Ecart	Ecart relatif
Logements en ZI	155 078	157 664	2 586	1,6%
Population en ZI	299 315	303 131	3 816	1,3%
Part des logements en ZI	12,3%	12,6%		
Part de la population en ZI	11,9%	12,1%		

Source : DGPR-DREAL, AZI et PPRI, 2010 – Insee, RP, 2006 - Insee, RFL, 2009 - ©IGN, BD TOPO®, 2009 - ©IGN, Contours.Iris®, 2008.

**Tableau 5 : Comparaison des estimations de population et de logements en zone inondable, par départements en région Centre**

Département	Logements carroyés en zone inondable (calage à l'IRIS)	Estimation de logements en zone inondable (BDTopo simple)	Estimation de logements en zone inondable (modèle économétrique)	Population carroyée en zone inondable (calage à l'IRIS)	Estimation de population en zone inondable (BDTopo simple)	Estimation de population en zone inondable (modèle économétrique)
Cher	2 470	2 836	2 663	4 306	4 979	4 667
Eure-et-Loir	13 882	12 350	12 368	28 796	25 813	25 852
Indre	5 372	5 004	5 125	9 613	8 835	9 004
Indre-et-Loire	76 816	76 751	76 896	140 517	140 906	141 198
Loir-et-Cher	17 460	17 117	17 353	31 166	30 854	31 240
Loiret	41 664	41 020	40 824	88 735	87 929	87 501
Centre	157 664	155 078	155 230	303 131	299 315	299 463

Source : DGPR-DREAL, AZI et PPRI, 2010 – Insee, RP, 2006 - Insee, RFL, 2009 - ©IGN, BD TOPO®, 2009 - ©IGN, Contours.Iris®, 2008. Traitements : SOeS, 2012.

Dans le Cher, l'Indre-et-Loire et le Loir-et-Cher, les méthodes BD TOPO® estiment davantage de population et de logements que les données carroyées. C'est l'inverse dans les trois autres départements de la région.

Par la méthode avec calage, la population en zone inondable est plus élevée. Elle s'élève à près de 303 131 habitants (soit +1,3 % par rapport à la méthode BD TOPO® simple) et 157 664 logements (+ 1,6 %).

Mais les différences avec les estimations BD TOPO® sont localement réduites (en valeur absolue).

**Tableau 6 : Comparaison des estimations du taux de population et de logements en zone inondable, par départements en région Centre**

Département	Part des logements en zone inondable			Part de la population en zone inondable		
	Carroyages (après calage)	BD Topo méthode simple	BD Topo modèle économétrique	Carroyages (après calage)	BD Topo méthode simple	BD Topo modèle économétrique
Cher	1,5%	1,7%	1,6%	1,4%	1,6%	1,5%
Eure-et-Loir	7,0%	6,3%	6,3%	6,8%	6,1%	6,1%
Indre	4,2%	3,8%	3,9%	4,2%	3,8%	3,9%
Indre-et-Loire	27,6%	26,9%	26,9%	24,6%	24,3%	24,3%
Loir-et-Cher	10,3%	10,2%	10,2%	9,5%	9,5%	9,6%
Loiret	13,4%	13,4%	13,2%	13,6%	13,8%	13,6%
Centre	12,6%	12,3%	12,3%	12,1%	11,9%	11,9%

Source : DGPR-DREAL, AZI et PPRI, 2010 – Insee, RP, 2006 - Insee, RFL, 2009 - ©IGN, BD TOPO®, 2009 - ©IGN, Contours...Iris®, 2008. Traitements : SOeS, 2012.

Les tests réalisés à l'échelle d'une région révèlent globalement peu d'écarts entre la population comptabilisée à partir des données carroyées et la population estimée après calage sur le RP. De plus, les deux types de méthodes d'estimation en zone inondable mènent à des résultats également proches.

### 3.2. Conclusion

Chaque méthode mobilise plusieurs sources auxiliaires et repose sur des traitements géographiques et statistiques. La comparaison des estimations ne permet pas de privilégier une méthode uniquement sur la base des résultats de tests.

L'argument principal en faveur de l'utilisation des données carroyées de population est d'ordre pratique. En effet, si l'on utilise comme données de base les populations carroyées, les travaux géographiques, de croisements de données, sont moins importants.

Les méthodes BD TOPO® nécessitent en revanche des traitements géographiques et statistiques plus complexes directement sur les bases BD TOPO® (et CORINE Land Cover). Mais l'intérêt de la méthode reposant sur la BD TOPO® est d'être applicable à tout le territoire français, y compris les 5 départements d'outre-mer. Les carroyages de population ne sont en effet pas disponibles sur la Guadeloupe, la Guyane et Mayotte. De plus, on ne dispose pas de CORINE Land Cover sur Mayotte.

## Annexes

### Annexe 1 : Problèmes de modélisation

#### Variance des perturbations

Les conditions sur la variance des résidus (voir 2.3.2.1) ne sont généralement pas respectées. Il existe plusieurs manières de traiter les problèmes d'hétéroscédasticité des résidus en modifiant le modèle à estimer.

La première est de diviser chaque variable du modèle par une variable censée être liée aux résidus comme la surface développée totale par Iris. Si on ignore la structure de la variance des résidus, on a recours à la méthode dite des moindres carrés quasi généralisés. C'est le modèle qui (théoriquement) traite le mieux l'hétéroscédasticité des résidus mais est difficilement applicable en prévision.

Enfin, on peut envisager un modèle logarithmique en expliquant le logarithme du nombre de logements par les surfaces développées. Mais la significativité du modèle est moins bonne.

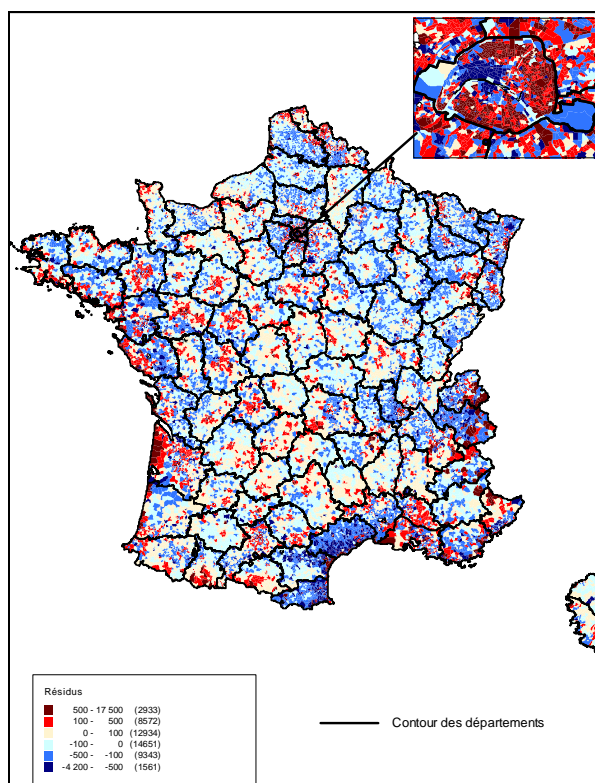
En définitive, aucune méthode ne permet de traiter de manière satisfaisante l'hétéroscédasticité des résidus. En théorie, ces modèles s'avèrent inefficaces si on n'a pas préalablement traité le problème dit d'autocorrélation des résidus.

#### Autocorrélation spatiale des perturbations

Mesurer l'autocorrélation spatiale n'a de sens que sur un ensemble d'Iris contigus ou tout du moins proches. C'est le cas dans les modèles différenciés par départements.

Sauf de manière empirique, en observant le signe des résidus, elle est difficile à tester.

Carte 2: Résidus des modèles départementaux



Note : Les résidus représentés sont issus des 95 modèles départementaux. Ils ne sont comparables qu'à l'intérieur d'un même département.

Source : Modèles BD TOPO®, Calculs SOeS.

Notons cependant une sous estimation des logements sur le littoral atlantique et méditerranéen.

Pour corriger ce phénomène, un modèle avec variable endogène décalée peut être envisagé.

La variable que l'on estime ne serait plus le nombre de logements par Iris mais un nombre de logements lié aux Iris voisins si ces derniers sont très proches géographiquement.

Le modèle s'écrirait alors

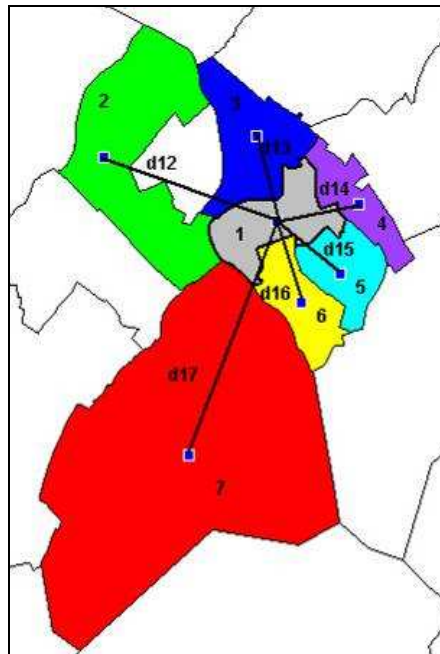
$$\text{NbLog}_i = \rho \cdot \sum_{i \neq j} w_{ij} \text{NbLog}_j + \sum_{k=1}^n a_k \text{SD}_{i,k} + \varepsilon_i$$

où  $\text{NbLog}_i$  est le nombre de logements dans l'Iris  $i$  et  $\text{SD}_{i,k}$  est la surface développée des polygones de l'Iris  $i$  dans la modalité  $k$  d'occupation du sol.

Le poids  $W_{ij}$  est une fonction décroissante de la distance  $d_{ij}$  entre les centroïdes des Iris  $i$  et  $j$  (calculé à partir de leurs coordonnées Lambert) rapportée à leur moyenne sur les Iris contigus.

Ainsi, la somme de tous les poids d'un Iris vaut 1.

On considère que ce poids est nul pour des Iris non contigus à l'Iris  $i$ .  
Dans notre cas,



$$W_{ij} = \frac{e^{-d_{ij}^2}}{\sum_{i \text{ contigu à } k} e^{-d_{ik}^2}}$$

Sur ce schéma, les Iris 2, 3, 4, 5, 6 et 7 ont une frontière commune avec l'Iris 1.

En reprenant les mêmes notations, le poids  $W_{12}$  entre 1 et 2 vaut donc :

$$W_{12} = \frac{e^{-d_{12}^2}}{e^{-d_{12}^2} + e^{-d_{13}^2} + e^{-d_{14}^2} + e^{-d_{15}^2} + e^{-d_{16}^2} + e^{-d_{17}^2}}$$

Le paramètre  $\rho$  est le **coefficient de corrélation** entre le nombre de logements au recensement et le nombre de logements estimés dans l'Iris. Il est généralement proche de 1.

Ce modèle peut être implémenté et produire des estimations mais il pose plusieurs problèmes.

Le nombre d'Iris possédant des zones inondables est généralement inférieur au nombre total d'Iris d'où une impossibilité d'appliquer le modèle en prévision sur l'ensemble des Iris.

Il faudrait estimer un modèle différent par Iris pour pouvoir l'appliquer en prévision !

De plus, en faisant intervenir les Iris voisins, on pourrait estimer des logements en zone inondable dans des Iris sans zone inondable, ce qui serait encore délicat dans l'interprétation.

## Annexe 2: Méthodes d'estimation des populations au carreau

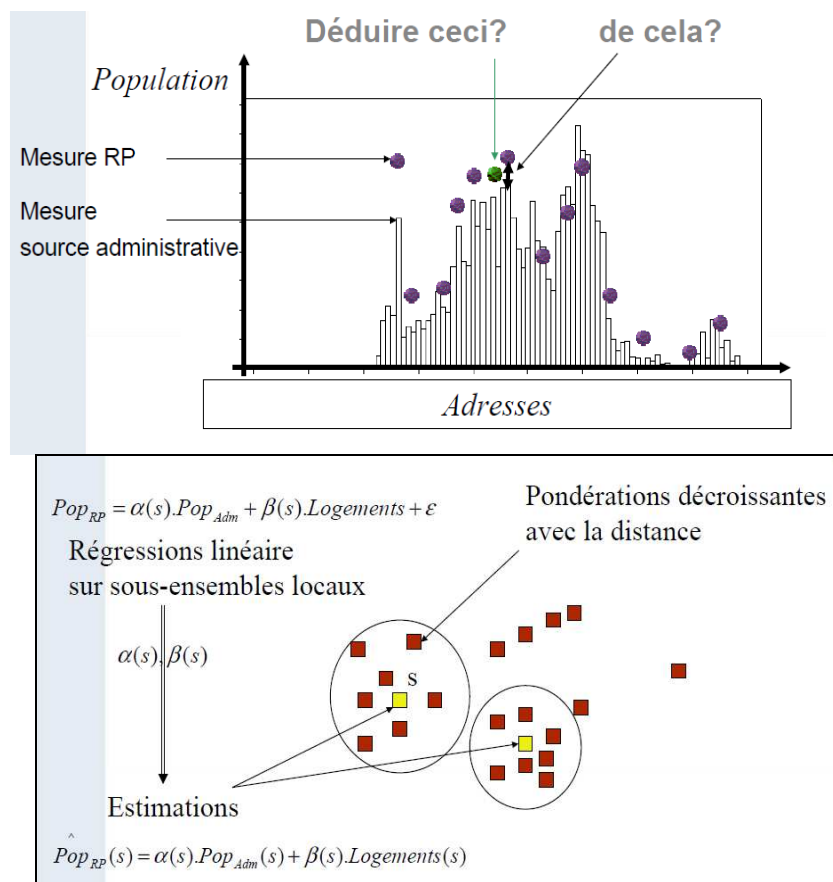
### 1. Communes de plus de 10 000 habitants

Les communes de plus de 10 000 habitants réalisent chaque année une enquête de recensement auprès d'une partie de leur population. La collecte annuelle porte sur un échantillon d'adresses tirées au hasard et représentant environ 8 % des logements de la commune, donc environ 8 % de sa population. Toutes les personnes vivant dans les logements des adresses ainsi tirées au sort sont recensées.

Mais ce mode de collecte ne permet pas, même sur les cinq années du cycle, d'avoir des données exhaustives. Un calage des poids avec le nombre de logements présents dans le répertoire d'immeubles localisés (RIL) est ainsi effectué à l'Iris. En dessous, ce calage n'est plus possible.

Afin d'obtenir une population sur une zone plus fine que l'Iris, il faut « remplir les trous de collecte » à partir de données administratives comme le RIL.

- Pour chacune des adresses à la population inconnue, on détermine un modèle de régression pondéré par la distance à partir des 50 adresses les plus proches géographiquement et qui ont été recensées (dont la population est connue).



Source : Journées de méthodologie statistique 2012- J.L.Lipatz- Insee/DG/DET.

- À partir des coefficients du modèle, un nombre de personnes est déterminé en prévision pour l'adresse non recensée.
- Il est fréquent que la population connue dans un logement ne soit pas tout à fait égale à celle qu'on aurait obtenue en appliquant le modèle. Un calage est réalisé pour être cohérent avec les données connues. La population estimée doit être égale à la somme pondérée des populations des adresses collectées.
- Les populations vivant en communautés sont simplement rajoutées à celles qui sont estimées.
- Enfin, les données individuelles sont agrégées pour chaque carreau.

## 2. Communes de moins de 10 000 habitants

Les communes de moins de 10 000 habitants sont recensées une fois tous les cinq ans, par roulement. À cet effet, elles ont été réparties en cinq groupes constitués selon des règles précises, qui assurent le même effectif et la même structure démographiques à chaque groupe. Chaque année, la totalité de la population et des logements de toutes les communes d'un même groupe est recensée, si bien que, au bout de cinq ans, l'ensemble des communes de moins de 10 000 habitants est recensé. Des méthodes d'interpolation et d'extrapolation des données (selon l'année de collecte) permettent de diffuser une population au district pour l'année médiane.

Afin de localiser la population, récupérer manuellement le nom des adresses saisies lors de la collecte serait trop coûteux et difficile à transcrire en coordonnées géographiques. Pour cela, l'Insee a développé la localisation des adresses à partir des données fiscales.

Après recodage de l'identifiant de parcelle de la taxe d'habitation au format des fichiers cadastraux, le processus se déroule en plusieurs phases successives :

### 1) Appariement des données de la TH avec les données RP en utilisant :

- la commune de résidence ;
- le sexe ;
- la date de naissance ;
- la commune (éventuellement seulement le département) de naissance.

Dans un certain nombre de cas, il y aura plusieurs possibilités.

### 2) Calcul d'un score individuel

À partir de la qualité de la reconnaissance du lieu de naissance de l'individu et des caractéristiques du logement (résidence principale ou secondaire, maison ou appartement...), un score est calculé.

### 3) Calcul par agrégation d'un score au niveau immeuble RP

### 4) Choix de l'appariement le plus probable, invalidation des appariements improbables

Les scores les plus élevés sont privilégiés mais il reste encore des cas de non appariement.

### 5) Interpolation des appariements invalidés et des échecs d'appariement

Les agents recenseurs ne collectent pas l'information de manière aléatoire. Deux enregistrements RP voisins dans le fichier devraient l'être aussi géographiquement sur le terrain lorsqu'ils ont été identifiés dans la même rue.

## Différentes étapes qui mènent au dénombrement de la population à l'intérieur des carreaux

