

---

## LA REpondÉRATION DES ENQUÊTES ANNUELLES DE RECENSEMENT POUR UNE DIFFUSION COMPLÉMENTAIRE DU RP

*Sébastien HALLÉPÉE, Pierre-Arnaud PENDOLI, Olivia SAUTORY*

*Insee, Département de la Démographie*

[sebastien.hallepee@insee.fr](mailto:sebastien.hallepee@insee.fr)  
[pierre-arnaud.pendoli@insee.fr](mailto:pierre-arnaud.pendoli@insee.fr)  
[olivia.sautory@insee.fr](mailto:olivia.sautory@insee.fr)

**Mots-clés** : sondage, calage, recensement, sources administratives

---

### Résumé

Le dernier recensement de la population (RP) exhaustif date de 1999. Depuis 2004, la France a rénové son opération de recensement qui repose désormais sur un cycle de cinq ans : chaque année, tout ou partie d'un groupe est recensé au cours des enquêtes annuelles de recensement (EAR). Un RP repose ainsi sur l'agrégation des données de cinq EAR.

Les groupes de rotation, sur lesquels sont construites les EAR, sont parfois très déséquilibrés en termes de nombre de logements et de population. On observe ainsi dans certaines régions d'importantes fluctuations dans les séries de populations calculées à partir des EAR, freinant jusqu'à présent une utilisation plus poussée de cette source.

L'exploitation des données d'une seule EAR offre pourtant plusieurs avantages, notamment la fraîcheur des données, l'unicité de la période d'observation et une plus grande réactivité en cas de modifications du questionnaire. Par exemple, une modification du questionnaire de la feuille de logement a lieu en 2018. Une nouvelle analyse ménage-famille pourra conduire à des études dès 2019 à partir de l'EAR, contre 2023 si on devait utiliser le RP.

Nous présentons une nouvelle méthodologie de calage des EAR visant à améliorer la précision de ces données. Celle-ci repose notamment sur l'utilisation de variables auxiliaires issues du Fichier démographique des logements et des individus (Fidéli) basé sur des sources fiscales.

La nouvelle pondération réduit les écarts entre les estimations issues de l'EAR 2015 et celles du RP 2015. Elle permet aussi de lisser les fluctuations des séries de population et de nombre de logements dues aux déséquilibres des groupes de rotation.

### Abstract

The last comprehensive Population Census was held in 1999. Since 2004, French census has been renewed and is based on a 5-year cycle : each year, all or part of a group is collected during the annual census surveys. A Population Census is based on the aggregation of 5 annual census surveys data.

The rotation groups, on which the annual census surveys are built, are sometimes imbalanced in terms of number of dwellings and inhabitants. In some regions, there has also been considerable

fluctuations in the series of population calculated from the annual census surveys. This problem has so far prevented a more elaborate use of these surveys.

However, using the data of the annual census survey provides several advantages, in particular the timeliness of data, the uniqueness of the observation period and an enhanced responsiveness in case of a modification of the questionnaire. For instance, a modification of the housing questionnaire happened in 2018. By using the annual census surveys data, studying complex family relationships would be possible from 2019. On the other hand, by using the Population Census, that is a complete cycle of five annual census surveys, these studies would be postponed to 2023.

We will present a new calibration methodology of the annual census surveys whose aim is to improve the estimations accuracy. This new method is based on the use of auxiliary variables from the new tax system used at INSEE and called "Demographic file of housings and individuals".

The new calibrated weights reduce the gap between the estimations from 2015 annual survey and the estimations from 2015 Population Census. Besides, the new calibration methodology enables to smooth the fluctuations in the series of population that are caused by the imbalance of the rotation groups.

## 1. Introduction

Le recensement de la population est basé sur un cycle de cinq ans depuis 2004 [1]. Ce changement a permis d'une part de répartir les coûts de collecte dans le temps et d'autre part de pouvoir diffuser de nouvelles populations et de nouveaux résultats statistiques basés sur le recensement chaque année et pour chaque commune depuis 2008.

Pour ce faire, on réalise chaque année la collecte du recensement d'un cinquième des communes de moins de 10 000 habitants (dites « petites communes ») de manière exhaustive et, sur toutes les communes de 10 000 habitants et plus (dites « grandes communes »), d'un échantillon d'adresses représentant 8 % des logements de la commune [2]. Les résultats du RP sont établis en cumulant les informations collectées sur un cycle de cinq années successives.

Le protocole de collecte est adapté aux catégories de population à enquêter. Les individus faisant partie de ménages ordinaires représentent la très grande majorité (97 %) de la population. Les individus résidant dans des communautés sont recensés selon un protocole spécifique. En particulier, la collecte des communautés est la plupart du temps concentrée sur une seule des cinq années du cycle pour la plupart des grandes communes. De même, les individus vivant dans des habitations mobiles et les sans-abris ne sont pas recensés de la même façon. Enquêtés exhaustivement tous les cinq ans, ils l'ont été par exemple en 2016 pour toutes les grandes communes.

L'opération du recensement comporte des spécificités dans les départements d'outre-mer où la mise à jour de la base de sondage des adresses à enquêter dans les grandes communes fait appel à une enquête cartographique menée sur une partie du territoire alors qu'en métropole, les mises à jour du répertoire d'adresses sont notamment initiées par le suivi des permis de construire et de démolir et les échanges avec les communes.

Ces particularités impliquent que la méthode d'estimation doit être adaptée selon la taille de la commune (plus ou moins de 10 000 habitants), la catégorie de population (ménages ordinaires, communautés, habitations mobiles et sans abris) et le territoire (métropole ou outre-mer).

L'enquête annuelle de recensement a été considérée pendant longtemps surtout comme un élément du cumul de cinq années successives de recensement. Néanmoins, l'EAR constitue depuis 2004 un produit de diffusion à part entière. Il est actuellement peu mis en avant à l'extérieur de l'Insee et du service statistique public et son utilisation est assortie de recommandations d'utilisation restrictives.

En effet, le produit est largement perfectible et conduit à des séries de résultats très volatils, même pour des indicateurs assez agrégés comme la population des régions. L'exploitation des données d'une seule EAR offre pourtant plusieurs avantages, notamment la fraîcheur des données, l'unicité de la période d'observation et une plus grande réactivité en cas de modifications du questionnaire. Par exemple, une modification du questionnaire de la feuille de logement a lieu en 2018. Une nouvelle analyse ménage-famille pourra conduire à des études dès 2019 à partir de l'EAR, contre 2023 si on devait utiliser le RP.

L'EAR a déjà servi de base à des opérations de diffusion complémentaires au RP en cumul. Au moment de la rénovation du recensement au début des années 2000, aucun cycle complet de recensement n'avait encore été réalisé. L'EAR 2004 a donc servi à mettre en place les prémices de la diffusion du RP. En 2015, certaines questions du bulletin individuel du recensement ont évolué pour répondre aux demandes formulées par le Cnis. Afin de disposer rapidement d'informations sur les thématiques concernées par le changement du questionnaire, il était nécessaire d'utiliser l'EAR indépendamment du cumul RP. L'EAR 2015 a ainsi été exploitée seule pour mener une étude sur l'utilisation du vélo comme mode de transport domicile – travail [3]. L'exploitation de l'EAR 2015 a également permis d'étudier pour la première fois les Pacs et les unions libres suite à l'introduction de ces statuts conjugaux dans le bulletin individuel [4]<sup>1</sup>. Ces études ont été menées au niveau national et dans plusieurs régions et ont permis de disposer d'estimations robustes aux niveaux des régions, des départements et des communes de plus de 100 000 habitants. Pour ces études, les travaux sur l'amélioration des pondérations de l'EAR n'étaient pas encore aboutis. Un examen de la qualité des estimations a montré qu'il était possible d'exploiter l'EAR avec les anciennes pondérations en se limitant aux indicateurs de structure (moyennes et proportions). En revanche, à l'exception d'un chiffre national, aucune estimation de total n'a été communiquée à cette occasion, car les indicateurs en volumes étaient trop instables. C'est en particulier pour l'estimation des totaux à partir d'une EAR que sont attendues les plus fortes améliorations de qualité avec la mise en place de la nouvelle méthode de pondération qui est développée dans la troisième partie de ce document.

## 2. Équilibrage imparfait des groupes de rotation

### 2.1. Plans de sondage équilibrés pour les groupes de rotation

Le tirage des cinq groupes de rotation du recensement en continu a été réalisé au début des années 2000 selon un plan de sondage spécifique à chaque champ, les champs étant définis selon plusieurs axes :

- Axe géographique : métropole *versus* outre-mer.
- Axe communal : petites communes *versus* grandes communes.
- Axe « catégorie de logement » : logements ordinaires *versus* autres logements (i.e. communautés, logements situés dans l'enceinte d'une communauté, habitations mobiles et sans abris).

À de rares exceptions près<sup>2</sup>, les groupes de rotation définissant les années de collecte des petites communes et des adresses ou des communautés des grandes communes ne sont pas modifiés au cours du temps.

<sup>1</sup>Insee Première n°1682 « Les Pacs à l'Ouest, les mariages à l'Est : une répartition des types d'unions différente selon les territoires », Isabelle Robert-Bobée, Vincent Vallès, 2018.

<sup>2</sup>Les modifications de géographie peuvent remettre en question ces groupes de rotation lorsque, par exemple dans le cas des fusions de communes, certaines communes sont amenées à changer de groupe.

### 2.1.1. Pour les logements ordinaires de métropole

- En petites communes

La commune constitue l'unité statistique. L'attribution d'une année de collecte à chaque petite commune s'est faite au moyen d'un tirage stratifié par région.

Au sein de chaque région, les petites communes sont réparties aléatoirement en cinq échantillons, par tirage équilibré à probabilités égales (1/5) sur des variables de type logement et des variables socio-démographiques données par le RP 1999 : le nombre total de logements, le nombre de logements en immeuble collectif, la population selon cinq classes d'âge<sup>3</sup>, la population par sexe et la population associée à chaque département.

La population rattachée aux logements ordinaires des petites communes de métropole rassemble près de 50 % de la population totale France entière.

La figure 1 ci-dessous représente les groupes de rotation de petites communes du département du Bas-Rhin.

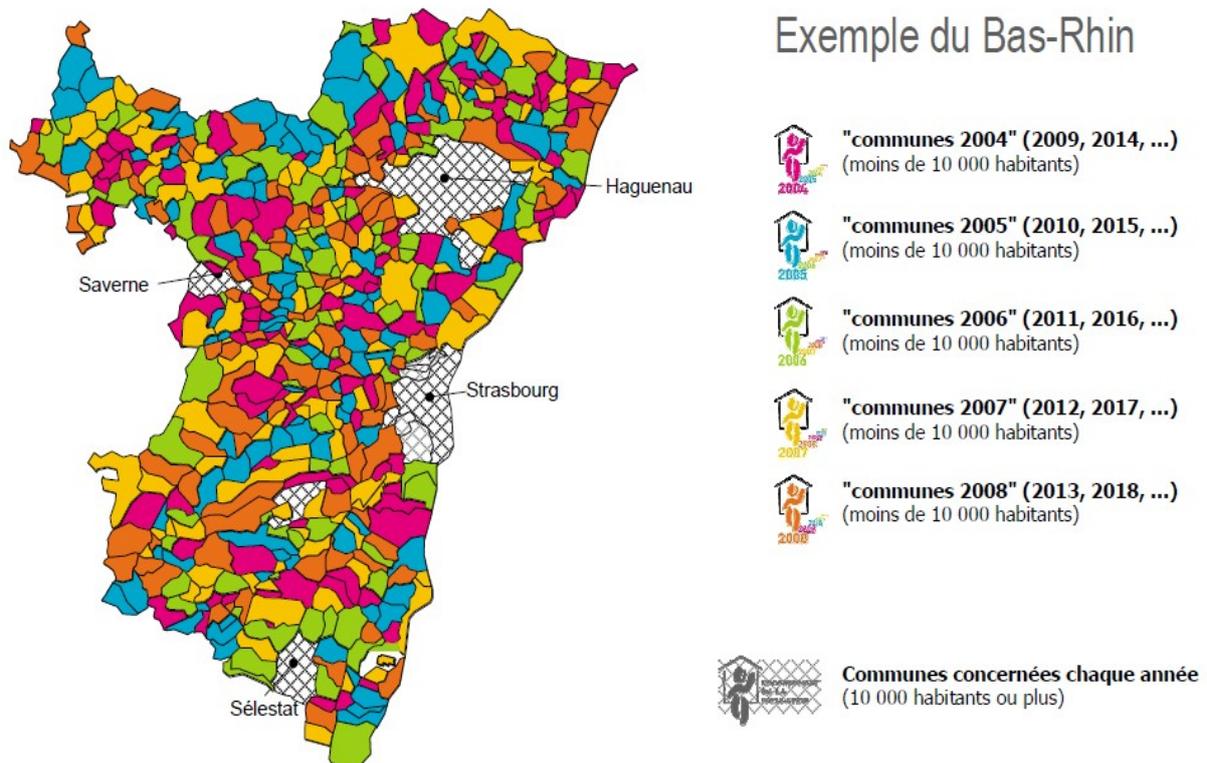


Figure 1 : Groupes de rotation de petites communes dans le département du Bas-Rhin

- En grandes communes

Chaque grande commune fait l'objet d'un plan de sondage indépendant. L'adresse constitue l'unité statistique.

Les adresses sont réparties en cinq groupes, par tirage aléatoire équilibré à probabilités égales (1/5) sur les mêmes variables que celles utilisées pour les petites communes de métropole, hors la population par département.

La population rattachée aux logements ordinaires des grandes communes de métropole rassemble environ 45 % de la population totale France entière.

<sup>3</sup> Moins de 20 ans ; 20 – 39 ans ; 40 – 59 ans ; 60 – 74 ans ; 75 ans et plus.

La figure 2 ci-dessous représente schématiquement la répartition des adresses d'une grande commune entre les groupes de rotation.



Figure 2 : Représentation schématique des groupes de rotation d'adresses d'une grande commune de métropole

#### 2.1.2. Pour les logements ordinaires des départements d'outre-mer (DOM)<sup>4</sup> et de Saint-Martin<sup>5</sup>

- En petites communes

La commune constitue l'unité statistique. La même méthodologie qu'en métropole a été utilisée pour l'échantillonnage des groupes de rotation. Toutefois le faible nombre de petites communes dans ces départements ne permet pas de les répartir en cinq groupes aussi strictement équilibrés sur les mêmes variables qu'en métropole. À titre d'exemple, le DOM comportant le plus de petites communes est la Martinique, soit seulement 22 petites communes. La seule variable d'équilibrage utilisée est la population au RP1999.

- En grandes communes

Chaque grande commune fait l'objet d'un plan de sondage indépendant. L'unité statistique est un ensemble d'adresses contiguës appelé « îlot ».

Le premier degré de tirage consiste à répartir aléatoirement les îlots en cinq échantillons, par tirage aléatoire équilibré à probabilités égales (1/5) sur les mêmes variables que celles utilisées pour les petites communes de métropole, hors la population par département. Le second degré de tirage consiste à sélectionner un échantillon d'adresses représentant 40 % des logements des îlots sur lesquels porte le recensement l'année considérée. L'échantillon d'adresses résulte d'un tirage aléatoire dont le taux de sondage du second degré vaut 40 % et qui est équilibré sur le nombre de logements et le nombre d'adresses des îlots du groupe de rotation.

<sup>4</sup> Il s'agit de la Guadeloupe, de la Martinique, de la Guyane et de La Réunion. Mayotte, considérée comme un DOM depuis le début du processus de départementalisation en 2009, n'a pas encore intégré le processus des EAR ; des recensements exhaustifs y ont été réalisés en 2007, 2012 et 2017.

<sup>5</sup> La grande commune de Saint-Martin est une collectivité d'outre-mer qui était encore une grande commune de Guadeloupe au moment de la rénovation du recensement.

La figure 3 ci-dessous représente une partie du territoire de la commune de Saint-Denis de La Réunion partitionnée en îlots. La figure 4 représente le résultat de la répartition aléatoire de ces îlots entre les groupes de rotation.

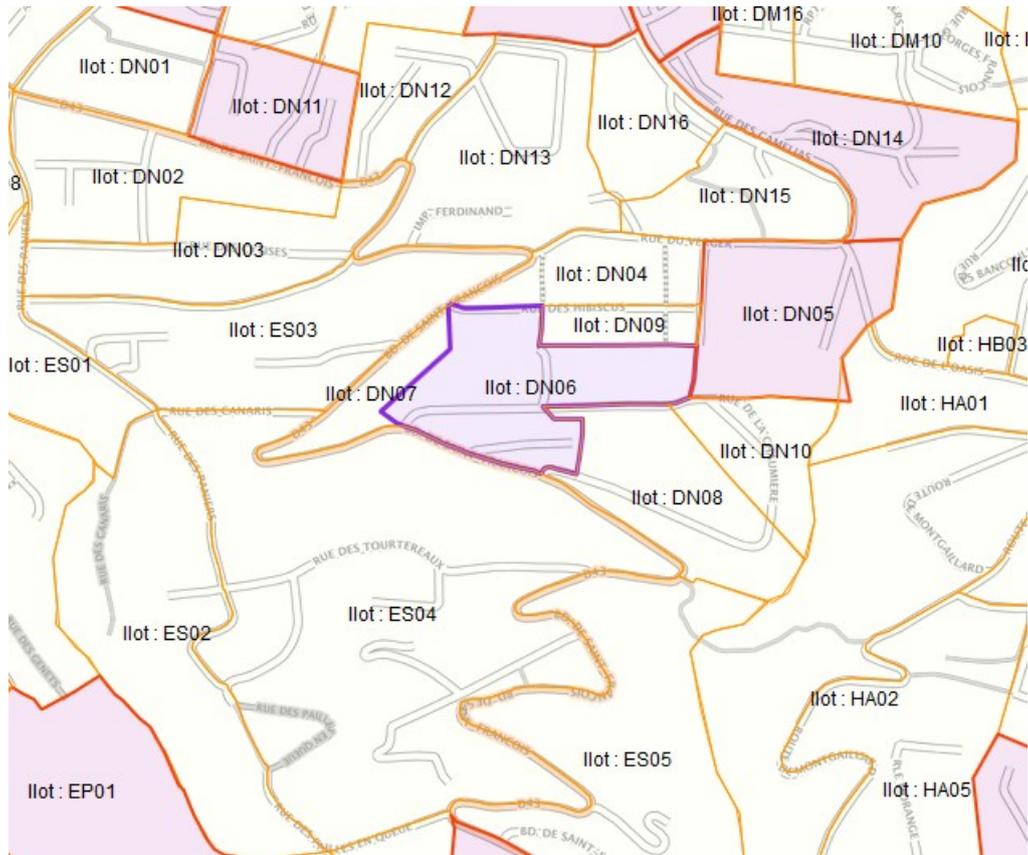


Figure 3 : Partition du territoire de la commune de Saint-Denis de La Réunion en îlots

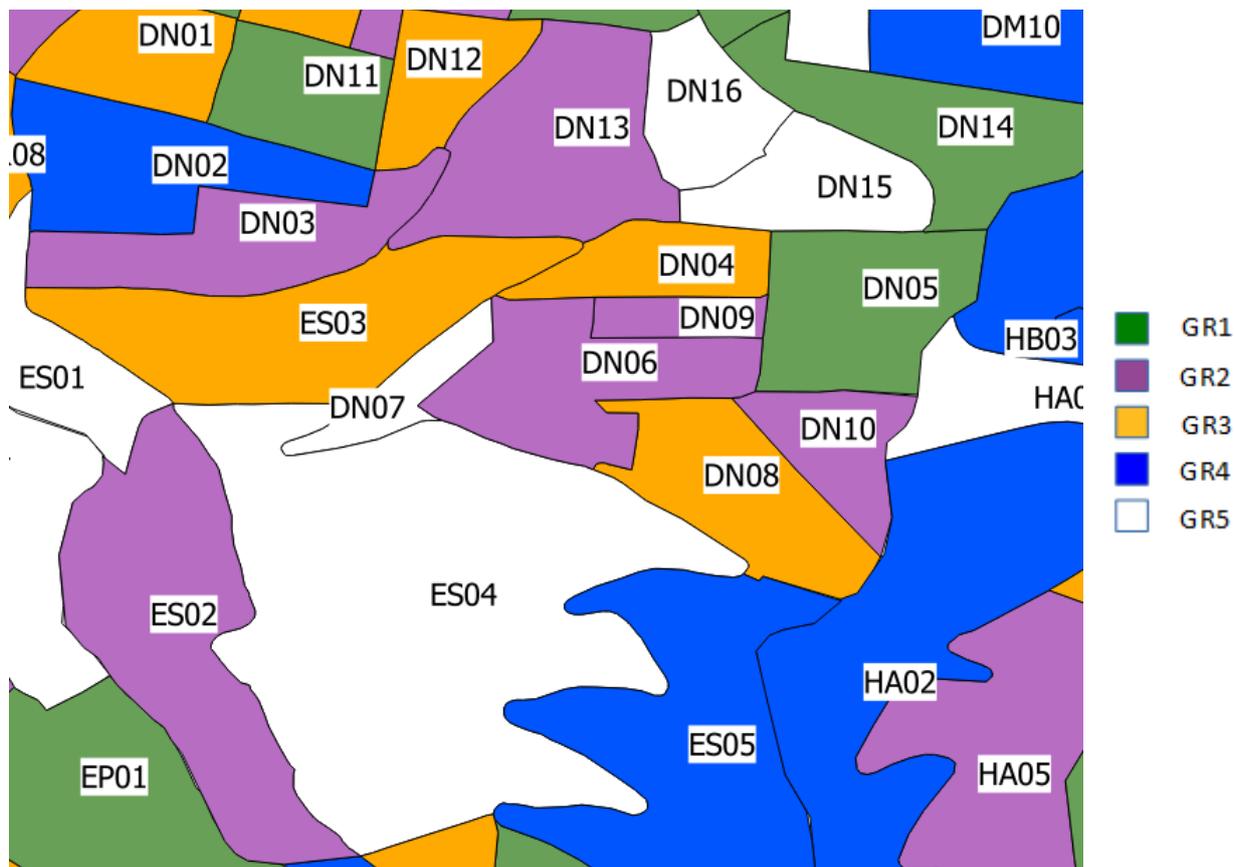


Figure 4 : Répartition des îlots entre les groupes de rotation (GR)

### 2.1.3. Pour les communautés et les logements se situant dans leur enceinte

La communauté constitue l'unité statistique.

- En petites communes

La commune constitue l'unité statistique. La communauté appartient au même groupe de rotation que la commune où elle est située.

- En grandes communes

Dans la majorité des cas, les communautés d'une grande commune sont recensées une seule fois sur un cycle de cinq ans, le groupe de rotation correspondant ayant été choisi de manière déterministe. Pour certaines grandes communes, en général les plus peuplées, les communautés ont été réparties de manière déterministe entre plusieurs groupes de rotation.

La population rattachée aux communautés et aux ménages qui y résident représente environ près de 2,5 % de la population totale France entière.

### 2.1.4. Pour les autres champs

Ces champs sont recensés exhaustivement une seule fois sur un cycle de cinq ans : les habitations mobiles et les sans-abris, les collectivités d'outre-mer (COM) de Saint-Pierre-et-Miquelon, de Saint-Martin et de Saint-Barthélemy.

### 2.1.5. Bilan des plans de sondage

Les unités d'échantillonnage et les plans de sondage diffèrent selon les catégories de population. Ainsi, même si l'EAR est une enquête de grande taille, la taille de l'échantillon et le taux de sondage varient selon les champs, comme le montre le tableau 1 ci-dessous.

| Catégorie de population             | Ménages ordinaires |           |         |         | Communauté |
|-------------------------------------|--------------------|-----------|---------|---------|------------|
|                                     | Métropole          |           | DOM     |         |            |
| Niveau géographique                 | Petites            | Grandes   | Petites | Grandes |            |
| Type de commune                     |                    |           |         |         |            |
| Nombre d'individus recensés à l'EAR | 6 480 000          | 2 320 000 | 63 000  | 130 000 | 330 000    |
| Nature de l'unité statistique       | Commune            | Adresse   | Commune | Adresse | Communauté |
| Nombre total d'unités statistiques  | 35 000             | 5 000 000 | 70      | 500 000 | 30 000     |
| Nombre d'unités échantillonnées     | 7 000              | 400 000   | 14      | 40 000  | 6 000      |
| Taux de sondage                     | 20%                | 8%        | 20%     | 8%      | 20%        |

Tableau 1. Tailles des échantillons et taux de sondage selon les champs de l'EAR

Les choix du plan de sondage et du taux de sondage impactant la précision des estimations associées aux paramètres d'intérêt, à la lecture de ces chiffres nous pouvons tirer les conclusions suivantes.

- En grandes communes, l'utilisation d'un plan de sondage en deux phases en métropole et d'un plan de sondage à deux degrés dans les DOM, avec un taux de sondage final des adresses de seulement 8 % vont impacter négativement la précision des estimateurs comparativement au plan de sondage à un seul degré associé à un taux de sondage de 20 % qui est utilisé pour les petites communes.
- En petites communes en revanche, la précision sera pénalisée par une taille d'échantillon relativement faible eu égard à l'échantillon d'adresses en grandes communes. En effet, si l'échantillon annuel de petites communes comporte environ 7 000 unités, l'échantillon d'adresses des grandes communes de métropole en comporte environ 400 000. Or, la variance des estimateurs diminue rapidement avec la taille de l'échantillon. Ainsi, comme nous allons le voir aux paragraphes suivants, c'est au niveau du champ des petites communes que les groupes de rotation sont les plus déséquilibrés, rendant plus volatiles les estimations annuelles des paramètres d'intérêt.

## 2.2. Déséquilibre empirique des groupes de rotation

Si l'équilibrage des groupes de rotation vis-à-vis des variables auxiliaires était exact, on devrait pouvoir estimer parfaitement la population au 1er janvier 1999 de chaque département.

Toutefois, lorsque l'on tente d'estimer pour chaque département la population des ménages au RP 1999 à partir des groupes de rotation, il apparaît que ces estimations ne sont pas exactes. D'après la figure 5, la majorité des estimations de populations départementales par groupe de rotation s'écartent de plus de 4 % par rapport au vrai total au 1er janvier 1999.

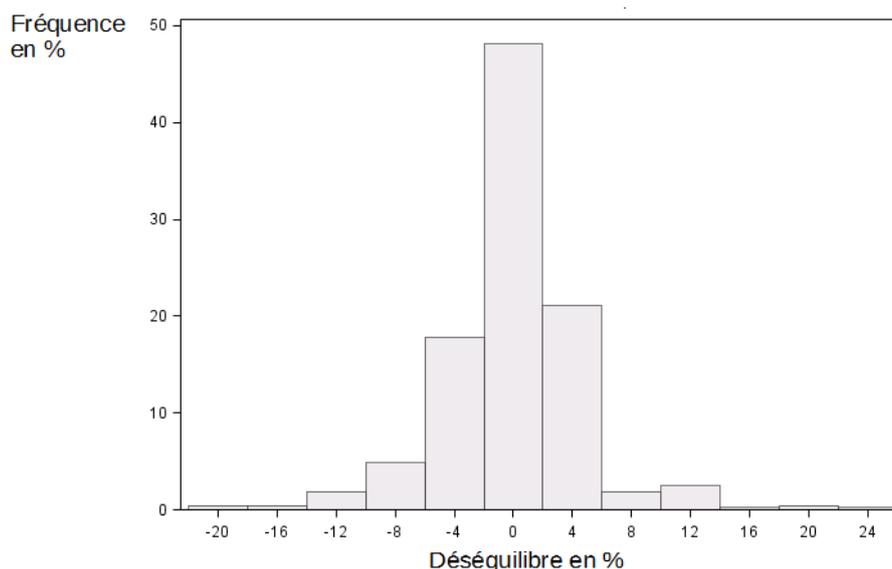


Figure 5 : Déséquilibres (en %) des GR vis-à-vis de la population par département au 01/01/1999.

Pour 9 départements sur 10, les petites communes contribuent à plus de 80 % du déséquilibre des groupes de rotation vis-à-vis de la population totale du département. Cela s'explique par le fait qu'au niveau départemental, chaque groupe de rotation peut contenir un nombre relativement faible de petites communes, en tout cas plus faible que le nombre d'adresses de chaque groupe de rotation en grandes communes. Calculées à partir d'échantillons de tailles plus faibles qu'en grandes communes, les estimations de population en petites communes se caractérisent par une variance plus forte au niveau départemental.

### 2.3. Volatilité des estimations de population à partir de l'EAR

Comme le montre la figure 6 pour la région Bourgogne-Franche-Comté, ces déséquilibres des groupes de rotation en termes de population sont à l'origine d'importantes fluctuations des estimations du nombre d'habitants par région entre les EAR. Dans l'état actuel des choses, il est donc impossible d'estimer le niveau de la population et son évolution à partir de l'EAR. Pour l'instant, seules les estimations du RP, qui cumulent les résultats de cinq EAR successives, peuvent être utilisées à cet effet. La rénovation de la méthode de pondération de l'EAR répondra notamment à l'objectif de pouvoir estimer des indicateurs démographiques centraux, tels que la population, aux niveaux des grandes communes, des départements et des régions.

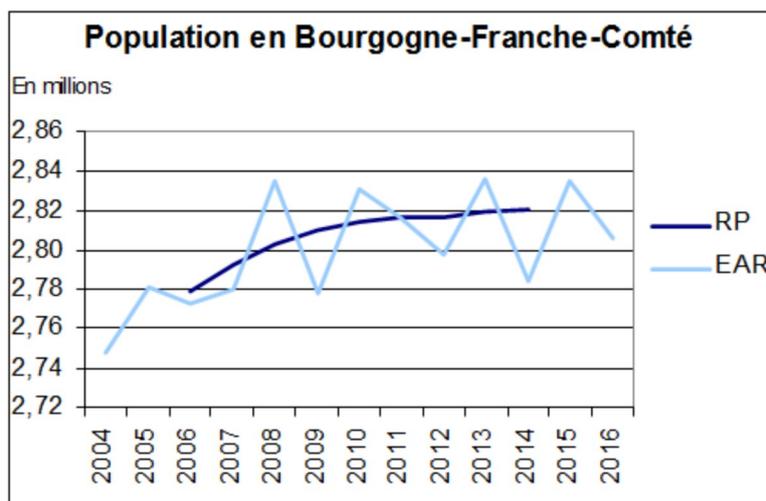


Figure 6 : Évolution des estimations de la population de la Bourgogne-Franche-Comté entre 2004 et 2016, à partir du RP d'une part et de l'EAR d'autre part.

### 3. Nouvelle méthode de calage des EAR

#### 3.1. Identification des champs de l'EAR

Les plans de sondage de l'EAR diffèrent selon les champs. Pour corriger les déséquilibres associés au plan de sondage initial, il est naturel que les unités de calage coïncident avec les unités d'échantillonnage.

##### 3.1.1. Échantillons de petites communes de métropole

Pour les petites communes, la création des groupes de rotation de communes est le seul degré d'échantillonnage du recensement. Les communes du groupe de rotation annuel sont recensées exhaustivement lors de l'EAR. Pour résorber les déséquilibres constatés précédemment, nous proposons une méthode de calage des groupes de rotation régionaux des petites communes sur les variables d'équilibrage actualisées.

Dans le cadre des travaux de repondération de l'EAR 2016 par exemple, les petites communes de chaque région administrative<sup>6</sup> recensées en 2016 forment un échantillon de communes à caler. On a donc autant d'échantillons de petites communes qu'il existe de régions.

##### 3.1.2. Échantillons de petites communes des DOM

Comme pour la métropole, dans chaque DOM les petites communes du groupe de rotation annuel, qui sont recensées exhaustivement, forment un échantillon de communes à caler.

##### 3.1.3. Échantillons d'adresses en grandes communes (métropole et DOM)

Dans chaque grande commune, on sélectionne pour chaque EAR un échantillon d'adresses à recenser. Dans le cadre des travaux de repondération de l'EAR 2016, les adresses d'habitation recensées en 2016 forment un échantillon d'adresses à caler pour chaque grande commune. On a donc autant d'échantillons d'adresses qu'il existe de grandes communes.

<sup>6</sup> Il s'agit des régions administratives issues de la réforme territoriale de 2015.

#### 3.1.4. Échantillons de communautés et de logements situés dans leur enceinte

Chaque année, on recense exhaustivement toutes les communautés du groupe de rotation annuel et les ménages qui y résident. Dans le cadre des travaux de repondération de l'EAR 2016, les communautés recensées en 2016 forment un échantillon de communautés à caler. On ne distinguera pas dans ces traitements les individus résidant en communautés de ceux résidant dans un logement situé dans l'enceinte de la communauté.

#### 3.1.5. Champs pour lesquels les poids de sondage sont conservés

Pour les champs qui sont recensés exhaustivement une seule fois sur un cycle de cinq ans (habitations mobiles et sans abris, Saint-Pierre-et-Miquelon, Mayotte et Saint-Barthélemy), il n'y a pas d'étape d'échantillonnage pour le recensement. Pour les individus qui sont recensés sur ces champs, on conserve donc un poids égal à 1.

On conserve également les poids de sondage pour la grande commune de Saint-Martin qui constitue à elle seule une collectivité d'outre-mer et où un échantillon d'adresses est tiré chaque année.

### 3.2. Une ancienne méthode de repondération perfectible

Auparavant, pour chaque champ, l'échantillon de l'EAR était calé sur une seule variable, ce qui revenait à utiliser l'estimateur par le ratio. Par exemple, en métropole comme dans les DOM, les petites communes du groupe de rotation recensé étaient calées, au niveau régional, sur le nombre de résidences principales (au RP 1999) de l'ensemble des petites communes de la région associée. Pour le champ des adresses d'habitation des grandes communes de métropole, on distinguait deux cas :

- Les grandes adresses du groupe de rotation recensé étaient calées, au niveau régional, sur le nombre de logements présents en grandes adresses dans l'ensemble des grandes communes de la région d'après le répertoire des immeubles localisés.
- Les petites adresses du groupe de rotation recensé étaient calées, au niveau communal, sur le nombre de logements présents en petites adresses dans la commune associée d'après le répertoire des immeubles localisés.

Cette ancienne méthode présentait ainsi deux inconvénients que la nouvelle méthode de repondération doit corriger :

- l'échantillon associé à un champ donné du recensement n'était calé que sur une seule variable ;
- pour certains champs, la variable de calage était issu du RP 1999, c'est-à-dire un millésime ancien de recensement dont les informations sont de moins en moins reliées à la situation correspondant au millésime de recensement que l'on souhaite corriger.

### 3.3. Détermination de la source auxiliaire

Une fois déterminés les échantillons correspondant aux différents champs du recensement, il est nécessaire d'identifier une source auxiliaire permettant d'améliorer les estimations de l'EAR. Cette base de calage doit répondre aux critères suivants :

- Des variables de type socio-démographique telles que celles présentes dans le recensement doivent être disponibles et mises à jour régulièrement de manière exhaustive. Vis-à-vis de ce critère, l'ancienne méthode de repondération n'était pas satisfaisante.
- Les totaux de ces variables auxiliaires doivent pouvoir être calculés à un niveau agrégé (i.e. les régions pour l'échantillon de petites communes ; les communes pour les échantillons d'adresses des grandes communes).

- Les variables auxiliaires doivent être connues pour chaque unité de l'échantillon (i.e. pour chaque petite commune d'une part et pour chaque adresse échantillonnée en grande commune d'autre part).

Pour le calage de l'EAR de l'année N, il serait possible de recourir au dernier RP disponible, i.e. le RP N-3 : pour l'EAR 2016, il s'agit du RP 2013 qui agrège les cinq EAR de 2011 à 2015. Mais cela impliquerait de caler l'échantillon de l'EAR sur des totaux ayant trois ans d'ancienneté. De plus, les données du RP 2013 comportent une relation avec la composition de l'échantillon de l'EAR 2016. En effet, les petites communes de l'EAR 2016 ont également été recensées en 2011 normalement. Les données de structure du RP 2013 pour ces communes sont donc toutes associées à une seule des cinq EAR du cumul du RP 2013, ce qui ne représente pas correctement l'intégralité des années d'observation prises en compte au RP 2013.

En revanche le Fichier démographique des logements et des individus (Fidéli) rassemble des données annuelles et exhaustives issues des sources fiscales (cadastre, taxe d'habitation, impôt sur le revenu...) similaires aux variables utilisées dans les plans de sondage des groupes de rotation du RP, telles que le sexe et l'âge des individus. Dans la nouvelle méthode de pondération, les variables auxiliaires sont donc recherchées dans ce fichier. Pour caler l'EAR de l'année N, elles sont issues du dernier millésime de Fidéli disponible, c'est-à-dire celui relatif à l'année N-1. Le calage va donc être efficace si les données démographiques N-1 de Fidéli sont bien corrélées aux données démographiques qu'on aurait collectées par un recensement exhaustif en N.

### 3.4. Calage de l'échantillon de petites communes de métropole

En petites communes, le calcul des variables auxiliaires est simple puisque le code officiel de la commune est toujours présent dans Fidéli. Pour le calage de l'EAR 2016, la méthode consiste à suivre les étapes suivantes :

- Pour chaque petite commune recensée en 2016, on calcule les variables auxiliaires suivantes à partir des données de Fidéli 2015 : nombre total de logements ; nombre total de logements en immeubles collectifs ; nombre total de personnes selon cinq classes d'âge ; nombre total de femmes ; nombre total d'hommes ; nombre total d'habitants.
- Pour chaque région administrative, on calcule les marges de calage associées à chaque variable auxiliaire à partir des données de Fidéli. Concernant le nombre total d'habitants, on calcule une marge pour chaque département de la région, ce qui permettra d'envisager une diffusion jusqu'à ce niveau géographique.
- Un calage de petites communes est réalisé pour chaque région à l'aide de la macro SAS %Calmar. Afin de nous prémunir contre une trop grande dispersion des poids calés, nous utilisons la méthode Logit qui permet un contrôle des déformations maximales des poids. Les paramètres de la macro %Calmar sont choisis de façon à ce que les rapports de poids soient compris entre 0,25 et 2, ce qui garantit que les nouveaux poids des communes soient compris dans l'intervalle  $[1,25 ; 10]^7$ , évitant ainsi d'avoir des unités trop influentes.

### 3.5. Calage de l'échantillon de petites communes des DOM

Les petites communes étant très peu nombreuses dans les DOM, les groupes de rotation sont très déséquilibrés, au regard du nombre d'habitants notamment. Pour chaque DOM, les petites communes du groupe de rotation annuel sont calées uniquement sur le nombre total d'habitants, ce qui revient à utiliser l'estimateur par le ratio. En notant  $X_i$  la variable auxiliaire donnant la population de la commune  $i$  et sachant que le poids de sondage de chaque commune dans son

<sup>7</sup> Chaque petite commune recensée a en effet un poids de sondage égal à 5.

groupe de rotation est  $d_i = 5$ , le poids calé d'une commune pour un département donné s'exprime selon la formule (1) suivante :

$$w_i = d_i \frac{\sum_{\substack{k \in \text{DOM} \\ k \in \text{PC}}} X_k}{\sum_{\substack{k \in \text{DOM} \\ k \in \text{PC} \\ k \in \text{GR}}} d_k X_k} \Rightarrow w_i = \frac{\sum_{\substack{k \in \text{DOM} \\ k \in \text{PC}}} X_k}{\sum_{\substack{k \in \text{DOM} \\ k \in \text{PC} \\ k \in \text{GR}}} X_k} \quad (1)$$

De la même façon qu'en métropole, il est possible de calculer la variable auxiliaire  $X$  pour chaque petite commune des DOM à partir du millésime le plus récent de Fidéli. Toutefois, des résultats empiriques montrent que le nombre d'habitants présente une qualité dégradée dans Fidéli pour les petites communes des DOM comparativement aux données du RP. C'est la raison pour laquelle cette variable auxiliaire et les marges de calage départementales associées sont calculées d'après les données du RP N-3, c'est-à-dire le dernier millésime de RP disponible.

### 3.6. Calage de l'échantillon d'adresses en grandes communes de métropole

En grandes communes, la constitution des variables auxiliaires pour chaque unité de l'échantillon n'est pas immédiate. En effet, il n'existe pas d'identifiant commun entre Fidéli et l'échantillon d'adresses du recensement. Le tirage de l'échantillon annuel d'adresses pour le recensement est réalisé parmi les adresses de la Base de Sondage d'Adresses (BSA) issue du Répertoire d'Immeubles Localisés (RIL) mis à jour chaque année, et qui appartiennent au groupe de rotation de l'année. Pour être en mesure de créer les variables de calage à partir de Fidéli pour chaque adresse de l'échantillon, un appariement est réalisé avec la BSA. Pour le calage de l'EAR 2016 par exemple, la méthode consiste à suivre les étapes suivantes.

#### 3.6.1. Appariement de la BSA avec Fidéli

Dans les adresses de la BSA, différents types d'adresse co-existent : des adresses simples, des adresses avec suffixes (bis, ter, etc.), des adresses dégroupées et des adresses non normalisées (cf. figure 7). Les adresses dégroupées sont des adresses découpées en plusieurs parties identifiables sur le terrain, parfois via leur bâtiment (A, B, C...) pour réduire la taille des unités à enquêter et ainsi améliorer la précision des estimations. Les adresses non normalisées n'ont pas de numéro sur leur voie. La diversité et la complexité des types d'adresses nécessitent de recourir à un algorithme permettant d'apparier les adresses de la BSA à Fidéli tout en évitant les faux appariements : environ 90 % des adresses de la BSA sont apparées. L'appariement est réalisé sur la base de variables d'adressage que l'on trouve à la fois dans Fidéli et dans la BSA : code commune, identifiant de la voie (code rivoti), numéro dans la voie, suffixe (bis, ter...), référence cadastrale.

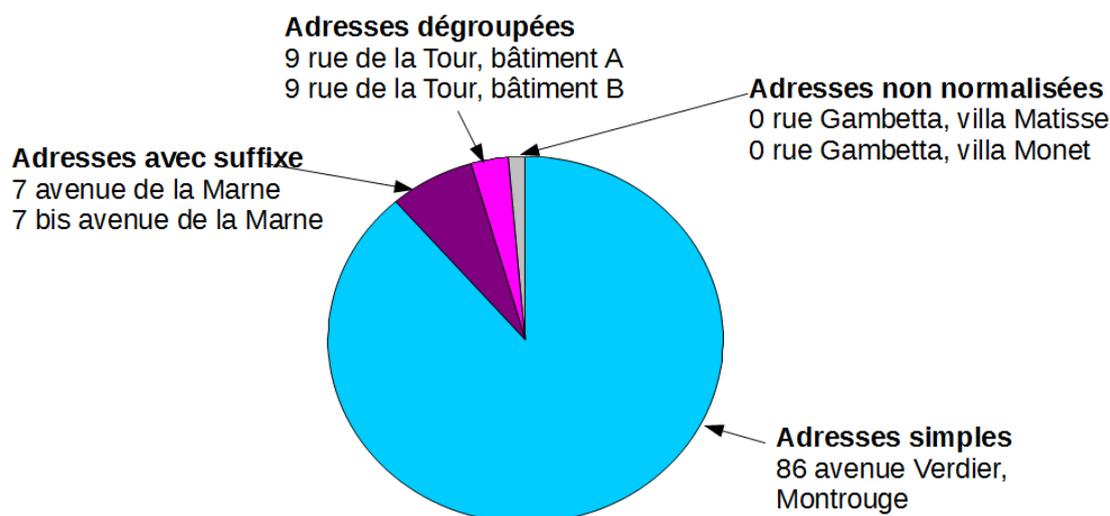


Figure 7 : Les catégories d'adresses de la BSA

Des difficultés d'appariement entre la BSA et Fidéli sont rencontrées pour chaque catégorie d'adresses.

- Pour les adresses simples : dans Fidéli un logement peut avoir pour adresse fiscale de traitement le 86 avenue Verdier mais pour adresse foncière le 84 avenue Verdier.
- Pour les adresses avec suffixe : dans Fideli , on peut trouver deux logements au 7 avenue de la Marne mais rien au 7 bis.
- Pour les adresses dégroupées : il peut n'y avoir aucune distinction selon le bâtiment dans Fidéli. Tous les bâtiments peuvent être localisés sur la même parcelle cadastrale dans Fidéli.
- Pour les adresses non normalisées : tous les bâtiments peuvent également être localisés sur la même parcelle cadastrale dans Fidéli.

Ces difficultés potentielles nécessitent de contrôler avec précision l'appariement afin d'éviter au maximum d'apparier des adresses à tort entre la BSA et Fidéli.

### 3.6.2. Calcul des variables auxiliaires composites

Il s'agit des variables auxiliaires usuelles : nombre total de personnes selon cinq classes d'âge ; nombre total de femmes ; nombre total d'hommes ; nombre total d'habitants ; nombre total de logements.

Comme on le voit dans la figure 8 ci-dessous, les variables auxiliaires ne sont pas calculées à partir de la seule source Fidéli. Il s'agit en fait de variables composites calculées de la manière suivante :

- Dans tous les cas, le nombre total de logements que l'on utilise en tant que variable auxiliaire est le nombre de logements de la BSA corrigé de la collecte. En effet, il s'agit de la seule variable auxiliaire dont on dispose de manière exhaustive dans la BSA et qui est mise à jour annuellement dans le RIL. Il n'est donc pas utile d'utiliser le nombre de logements issu de Fidéli.
- Pour la fraction (89 %) de la BSA appariée à Fidéli, les autres variables auxiliaires sont calculées pour chaque adresse à partir de Fidéli .
- Pour la fraction qui n'a pas été appariée, les variables auxiliaires sont calculées à partir de la dernière collecte de recensement connue en excluant l'EAR que l'on cherche à pondérer (9 %) ou sont imputées (2 %). Dans ce dernier cas, pour chaque variable auxiliaire, on calcule sur les adresses de la même commune appariées avec Fidéli ou précédemment enquêtées au recensement la valeur moyenne du total de la variable par logement. Pour chaque adresse

sujette à imputation, on affecte cette valeur moyenne multipliée par le nombre de logements de l'adresse dans la BSA corrigé de la collecte.

Par exemple, pour une adresse de la BSA qui n'a pas été appariée avec Fidéli et qui n'a jamais été recensée, la valeur imputée pour la variable auxiliaire « nombre d'hommes » est donnée par la formule (2) suivante :

$$\text{nb d'hommes imputé} = \frac{\text{nb total d'hommes de la commune}}{\text{nb total de logements de la commune}} \times \text{nb logements de l'adresse} \quad (2)$$

### 3.6.3. Calcul des marges de calage

Pour chaque grande commune, les marges de calage sont les totaux au niveau communal des variables auxiliaires composites calculées sur la BSA.

### 3.6.4. Mise en œuvre du calage

L'échantillon à caler est l'ensemble des adresses recensées lors d'une EAR donnée. À chaque adresse sont rattachées les variables auxiliaires composites issues de la BSA enrichie par les traitements décrits précédemment. Pour éviter de réaliser près de 1 000 calages<sup>8</sup>, on réalise finalement 13 calages, c'est-à-dire un par région, avec des marges communales<sup>9</sup>. Les calages sont réalisés à l'aide de la macro SAS %Calmar dont on précise certains paramètres.

- On utilise la méthode Logit avec des bornes de rapports de poids identiques pour chaque région : 0,3 et 2,5.
- Le paramètre POIDS correspond aux poids de sondage des adresses de l'échantillon.
- On renseigne le paramètre PONDQK en entrée du calage. Pour chaque adresse de l'échantillon, la valeur de la variable PONDQK est donnée par la formule (3) suivante :

$$\text{pondQK} = \frac{1}{\text{nb logements de l'adresse dans la BSA corrigé de la dernière collecte du recensement}} \quad (3)$$

Il s'agit d'une variable de pondération spécifique des adresses de l'échantillon, qui n'est pas liée aux poids de sondage. L'utilisation de ce paramètre est nécessaire pour que le calage converge pour certaines régions, et évite que les plus grandes adresses subissent des déformations de poids importantes. On associe ainsi mieux pour chaque adresse la dimension des bases de données de diffusion qui sont aux niveaux des ménages et des individus.

<sup>8</sup> La France métropolitaine compte près de 1 000 grandes communes.

<sup>9</sup> Sept variables de calage sont associées à chaque commune. Pour une région de 50 grandes communes, on aura donc 350 variables de calage.

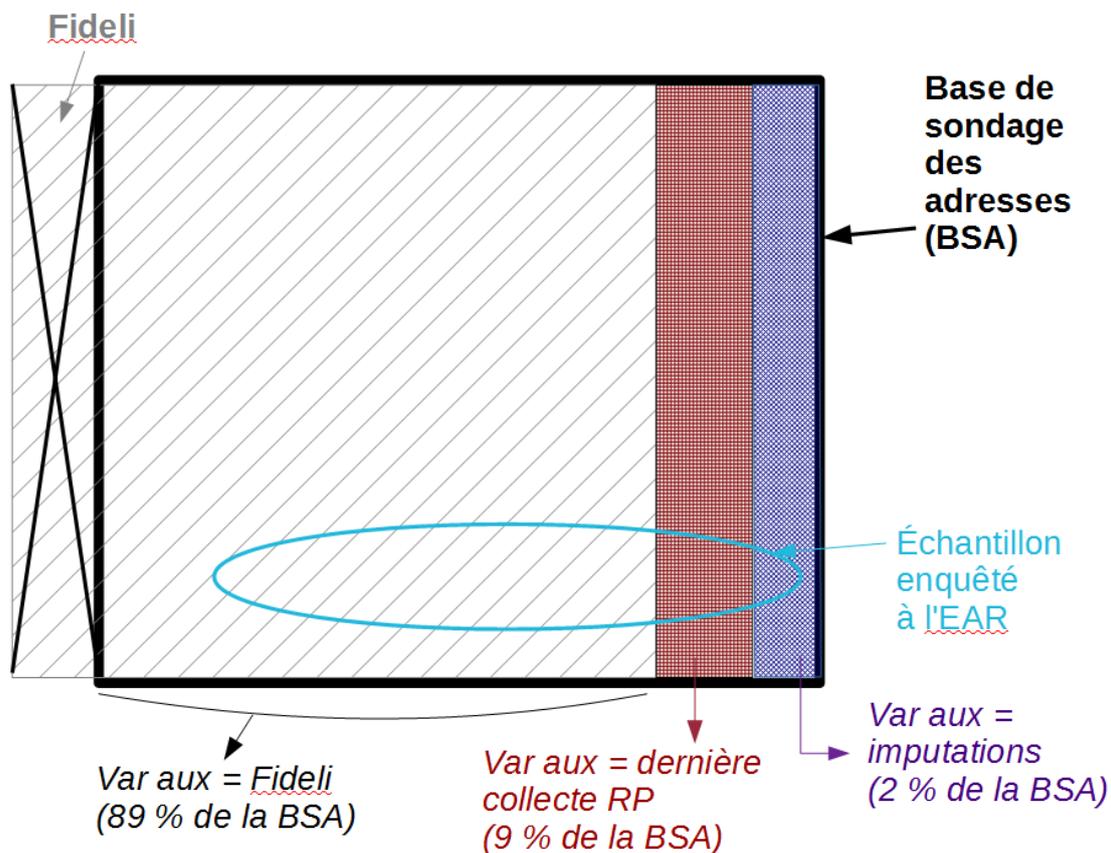


Figure 8 : Schématisation de l'appariement entre la BSA et Fidéli et du calcul des variables auxiliaires composites

### 3.7. Calage pour le champ des grandes communes des Antilles-Guyane

La même problématique qu'en métropole se pose pour l'échantillon annuel d'adresses des grandes communes des DOM au sujet du calcul des variables de calage. Pour être en mesure de créer les variables auxiliaires à partir de Fidéli pour chaque adresse de l'échantillon, un appariement est réalisé avec la BSA.

#### 3.7.1. Mauvaise qualité de l'appariement de la BSA avec Fidéli

Comme on le voit dans le tableau 2 ci-dessous, une minorité d'adresses de la BSA est appariée avec Fidéli en Guyane, la situation étant encore plus dégradée en Guadeloupe et en Martinique. La mauvaise qualité de l'appariement aux Antilles et en Guyane s'explique par la part élevée que représentent les adresses non normalisées, c'est-à-dire sans numéro de voie, parmi les adresses de ces départements.

Afin d'identifier dans Fidéli différentes adresses non normalisées d'une même voie de la BSA, on peut recourir à la parcelle cadastrale qui correspond à un découpage géographique plus fin au niveau du bâti. Cependant, une majorité d'adresses non normalisées ne possèdent pas de parcelle cadastrale unique en Guadeloupe et en Guyane (cf. tableau 2). Cela signifie que dans la majorité des cas, les entités adressées non numérotées d'une même voie sont rattachées à des références cadastrales identiques. Il est alors impossible d'apparier ces entités adressées car aucun identifiant unique ne peut être constitué pour ces adresses entre la BSA et Fidéli.

| Zone             | % d'adresses de la BSA appariées avec Fidéli | % d'adresses non normalisées dans la BSA | % d'adresses non normalisées dont la parcelle cadastrale est en doublon |
|------------------|--|--|---|
| 971 – Guadeloupe | 8,3%   | 61,7%                                    | 51,2 %  |
| 972 – Martinique | 12,7%  | 55,9%                                    | 41,6 %  |
| 973 – Guyane     | 31,5%  | 42,7%                                    | 65,3 %  |
| 974 – Réunion    | 67,5%  | 5,4%                                     | 32,1 %  |
| Métropole        | 89,0%  | 1,7%                                     | s.o.  |

Tableau 2. Résultat de l'appariement de la BSA avec Fidéli

Pour les Antilles et la Guyane, il n'est donc pas possible de calculer de manière convenable les variables auxiliaires pour chaque adresse de la BSA à partir de Fidéli. Sur ce champ, le calage de l'échantillon d'adresses de l'EAR sur les totaux de Fidéli est exclu.

### 3.7.2. Un calage au 1<sup>er</sup> degré tirage plutôt qu'au second

Pour chacune des grandes communes des DOM, on rappelle que l'échantillon de l'EAR suit un plan de sondage à deux degrés :

- échantillon de 1<sup>er</sup> degré : les îlots du groupe de rotation annuel ;
- échantillon de 2<sup>nd</sup> degré : les adresses sélectionnées parmi les adresses des îlots du groupe de rotation annuel.

Comme pour les autres champs du recensement, notamment les groupes de rotation de petites communes, les déséquilibres des groupes de rotation d'îlots pour les grandes communes des DOM s'accroissent au cours du temps. Pour le champ des grandes communes des Antilles et de la Guyane on propose donc une méthode de calage des groupes de rotation d'îlots.

### 3.7.3. Calage du groupe de rotation annuel d'îlots

Comme pour les petites communes, d'une part, et les adresses de la BSA en grandes communes, d'autre part, il est nécessaire de calculer pour tous les îlots de chaque grande commune des DOM les variables auxiliaires usuelles à partir de Fidéli. Néanmoins on ne dispose pas du code îlot pour les logements et les individus situés dans les DOM dans Fidéli. Dans le cadre des travaux de repondération des EAR, la méthode consiste à suivre les étapes suivantes :

- Tout d'abord, il s'agit de trouver un niveau géographique infra-îlot présent à la fois dans Fidéli et dans la BSA des DOM qui permette d'imputer un îlot dans Fidéli à chaque observation. On utilise pour cela les coordonnées géographiques (X, Y). À partir des coordonnées (X, Y) présentes dans Fidéli et des contours d'îlots présents dans l'application Rorcal<sup>10</sup>, il est possible d'imputer à l'aide d'une requête géométrique un code îlot aux logements et aux individus présents dans Fidéli.
- À partir des fichiers Logements et Individus de Fidéli enrichis du code îlot, on calcule pour tous les îlots des grandes communes des DOM les variables auxiliaires usuelles : nombre total de logements ; nombre total de logements en immeuble collectif ; nombre total de personnes selon cinq classes d'âge ; nombre total de femmes ; nombre total d'hommes.

<sup>10</sup> Rorcal est l'application de l'Insee permettant la gestion du Répertoire d'Immeubles Localisés (RIL) de chacune des grandes communes de France (métropole et DOM). C'est à partir des fichiers RIL mis à jour annuellement que l'on constitue les BSA des grandes communes.

- Pour chaque commune des DOM, on calcule les marges de calage associées à chaque variable auxiliaire à partir des données de Fidéli.
- On doit réaliser le calage des îlots du groupe de rotation associé à l'EAR pour chaque grande commune. Pour simplifier l'implémentation, on choisit une solution alternative mais équivalente consistant à caler simultanément les îlots du groupe de rotation annuel de chaque DOM, en calculant des variables auxiliaires communales.
- Le calage de l'échantillon d'îlots est mis en œuvre pour chaque DOM à l'aide de la macro SAS %Calmar. Afin de nous prémunir contre une trop grande dispersion des poids calés, nous utilisons la méthode Logit. Les paramètres de la macro SAS %Calmar sont choisis de façon à ce que les rapports de poids soient compris entre 0,2 et 3, ce qui garantit que les poids calés des îlots du groupe de rotation annuel soient compris dans l'intervalle [1 ; 15]<sup>11</sup>. Au niveau des logements et des individus recensés, leurs poids de sondage conditionnels de 2<sup>d</sup> degré valant 2,5, leurs poids finaux sont alors compris dans l'intervalle [2,5 ; 37,5]<sup>12</sup>.

### 3.8. Calage pour le champ des grandes communes de la Réunion

#### 3.8.1. Méthode de calage simultanée pour un sondage à deux degrés

Contrairement aux grandes communes des Antilles et de la Guyane, l'appariement des adresses de la BSA pour les grandes communes de la Réunion est de relativement bonne qualité puisque environ les deux tiers des adresses de la BSA sont appariées avec Fidéli. Ce résultat peut s'expliquer par le fait que les adresses non normalisées représentent une part des adresses beaucoup plus faible à la Réunion que dans les autres DOM (cf. tableau 2). Le calage de l'échantillon de 2<sup>d</sup> degré, c'est-à-dire de l'échantillon d'adresses, sur la source Fidéli, est donc envisageable pour les grandes communes de la Réunion.

Néanmoins le calage de l'échantillon de 1<sup>er</sup> degré, c'est-à-dire des îlots du groupe de rotation annuel, sur cette même source Fidéli est également possible puisque la méthode décrite pour les grandes communes des Antilles et de la Guyane peut s'appliquer également aux grandes communes de la Réunion. Pour concilier ces deux approches, nous optons pour un calage simultané dans le cas d'un sondage à deux degrés, dont nous décrivons ci-dessous l'application.

- Cette méthode prend en compte l'aléa d'échantillonnage du 1<sup>er</sup> degré.

Soit  $X$  le vecteur de variables auxiliaires pour lesquelles on connaît :

- les valeurs  $x_m$  pour chaque îlot  $m$  appartenant au groupe de rotation annuel  $S_1$  (il s'agit de l'échantillon de 1<sup>er</sup> degré) d'une grande commune donnée
- et le vecteur des totaux calculés sur l'ensemble  $U_1$  des îlots de la commune :

$$X = \sum_{m \in U_1} x_m$$

Soit  $\pi_m = \frac{1}{5}$  la probabilité d'inclusion de l'îlot  $m$  dans son groupe de rotation.

- De plus cette méthode prend en compte l'aléa d'échantillonnage du 2<sup>d</sup> degré.

Soit  $V$  le vecteur de variables auxiliaires pour lesquelles on connaît :

<sup>11</sup> Le poids de sondage de chaque îlot dans son groupe de rotation est en effet égal à 5.

<sup>12</sup> Le taux de sondage des adresses au 2<sup>d</sup> degré de tirage est en effet égal à 0,4. En effet, on souhaite recenser chaque année environ 40 % des logements et des individus au sein des îlots du groupe de rotation annuel.

- les valeurs  $v_k$  pour chaque adresse  $k$  appartenant à l'échantillon annuel d'adresses  $S_2$  (il s'agit de l'échantillon de 2<sup>d</sup> degré) d'une grande commune donnée
- et le vecteur des totaux calculés sur l'ensemble  $U_2$  des adresses de la commune :

$$V = \sum_{k \in U_2} v_k$$

Soit  $S_{2m}$  l'échantillon de 2<sup>d</sup> degré (i.e. d'adresses) tiré dans l'îlot  $m$  .

- Pour toute adresse  $k \in S_{2m}$  on note  $d_{k,m}$  son poids de sondage conditionnel de 2<sup>d</sup> degré.
- On peut alors estimer les totaux des variables  $v$  au niveau de chaque îlot à partir des échantillons d'adresses qui y ont été sélectionnés au 2<sup>d</sup> degré de tirage. Ces estimations de totaux au niveau îlot sont notées  $v_m$  :

$$v_m = \sum_{k \in S_{2m}} d_{k,m} v_{k,m}$$

Pour l'échantillon d'adresses d'une grande commune de la Réunion, l'équation du calage simultané s'écrit selon la formule (4) :

$$\sum_{m \in S_1} \frac{F(x'_m \lambda + v'_m \gamma)}{\pi_m} (x_m, v_m) = (X, V) \quad (4)$$

où  $F$  est une fonction de distance et les vecteurs  $\lambda$  et  $\gamma$  sont les multiplicateurs de Lagrange.

Les quantités  $\pi_m$  et  $x_m$  sont calculées au niveau des îlots, comme c'est le cas également pour le champ des grandes communes des Antilles et de la Guyane. En revanche les autres quantités :  $d_{k,m}$  et  $v_k$  doivent être calculées au niveau de l'échantillon d'adresses.

### 3.8.2. Calcul des variables auxiliaires du 2<sup>d</sup> degré de tirage

Comme on l'a vu dans le tableau 2, environ les deux tiers des adresses de la BSA des grandes communes de la Réunion sont appariées avec Fidéli. Il n'est donc pas possible de calculer les variables auxiliaires usuelles<sup>13</sup> à partir de Fidéli pour toutes les adresses de l'échantillon annuel. Le tirage de 2<sup>d</sup> degré est pris en compte de la manière suivante dans le calage :

- **Pour la fraction de l'échantillon d'adresses appariée à Fidéli**
  - Le vecteur  $v_k$  comprend les valeurs des variables auxiliaires usuelles (nombre d'hommes...) calculées pour l'adresse  $k$  à partir de Fidéli.
  - Le vecteur  $v_m$  comprend les estimations pondérées<sup>14</sup> du total de chaque variable auxiliaire dans l'îlot  $m$  calculées sur la fraction de l'échantillon appariée à Fidéli.
  - Le vecteur  $V$  comprend les totaux communaux de chaque variable auxiliaire calculés sur la fraction du fichier de Fidéli appariée à la BSA.
- **Pour la fraction de l'échantillon d'adresses non appariée à Fidéli**
  - Le vecteur  $\bar{v}_k$  comprend le nombre de logements de l'adresse  $k$  issu de la BSA.

<sup>13</sup> Nombre d'hommes, nombres de femmes, nombre de personnes pour les cinq classes d'âge, nombre de logements, nombre de logements en immeuble collectif.

<sup>14</sup> Le poids utilisé est le poids de sondage conditionnel de l'adresse pour le tirage de 2<sup>d</sup> degré  $d_{k,m}$  .

- Le vecteur  $\bar{v}_m$  comprend les estimations pondérées du nombre total de logements dans l'îlot  $m$  calculées sur la fraction de l'échantillon non appariée à Fidéli.
- Le vecteur  $\bar{V}$  comprend le nombre total de logements de chaque commune calculé sur la fraction de la BSA appariée à Fidéli.

Finalement l'équation du calage simultané à deux degrés se réécrit selon la formule (5) suivante :

$$\sum_{m \in S_1} \frac{F(x'_m \lambda + v'_m \gamma + \bar{v}'_m \alpha)}{\pi_m} (x_m, v_m, \bar{v}_m) = (X, V, \bar{V}) \quad (5)$$

### 3.8.3. Mise en œuvre du calage des îlots

Toutes les informations auxiliaires, y compris celles propres aux adresses, sont agrégées au niveau de l'îlot. La méthode présentée consiste donc à caler l'échantillon d'îlots correspondant au groupe de rotation annuel.

- Comme pour les Antilles et la Guyane, à partir des fichiers Logements et Individus de Fidéli enrichis du code îlot, on calcule pour tous les îlots des grandes communes de la Réunion les variables auxiliaires usuelles (vecteurs  $x_m$ ).
- À partir de l'échantillon d'adresses de l'EAR :
  - On calcule les estimations des variables auxiliaires usuelles au niveau îlot sur le champ des adresses appariées à Fidéli (vecteurs  $v_m$ ).
  - On calcule les estimations du nombre de logements au niveau îlot à partir de la BSA, sur le champ des adresses non appariées à Fidéli (vecteurs  $\bar{v}_m$ ).
- Les marges de calage sont calculées pour chaque grande commune de la Réunion :
  - Les marges du vecteur  $X$  sont calculées à partir de l'ensemble des îlots de la commune.
  - Les marges du vecteur  $V$  sont calculées à partir de la fraction des fichiers de Fidéli appariée à la BSA.
  - Les marges du vecteur  $\bar{V}$  sont calculées à partir de la fraction de la BSA qui n'a pas été appariée à Fidéli.
- Le calage de l'échantillon d'îlots est mis en œuvre à l'aide de la macro SAS %Calmar, en utilisant les mêmes paramètres que pour le champ des grandes communes des Antilles et de la Guyane.

## 3.9. Calage de l'échantillon de communautés

### 3.9.1. Spécificité de la source auxiliaire

On ne dispose pas d'information actuellement suffisante pour identifier les communautés dans Fidéli. Les seules données dont on dispose l'année N sur le champ des communautés sont les suivantes :

- d'une part les populations légales du RP N-3 dont il faut noter qu'il s'agit d'estimations ;
- d'autre part des caractéristiques structurelles qui sont contenues dans le RIL à partir duquel on constitue un répertoire des communautés. La variable essentielle est la catégorie de la communauté, qui permet de distinguer les EHPAD, les structures Adoma, les établissements pénitentiaires, les casernes ou encore les communautés religieuses. Cette catégorie de communautés caractérise très bien le profil d'habitation de la structure et semble donc une variable auxiliaire intéressante dans le cadre de la repondération.

À partir des populations légales du RP N-3 et de l'image du répertoire des communautés au 1<sup>er</sup> janvier N, on calcule pour chaque communauté les variables auxiliaires suivantes :

- populations par catégorie ;
- populations par région administratives ;
- populations pour certains croisements catégorie x région administrative.

### 3.9.2. Définition des marges de calage et de l'échantillon à caler

Entre le champ des communautés au 1<sup>er</sup> janvier N-3 et le 1<sup>er</sup> janvier N, on observe deux phénomènes.

- Tout d'abord il existe un phénomène d'attrition du champ des communautés au cours du temps, entre le 1<sup>er</sup> janvier N-3 et la collecte de l'EAR N. À partir de la population des communautés du RP N-3 qui existent toujours au 1<sup>er</sup> janvier N, on calcule les marges de calage, c'est-à-dire les populations légales RP N-3 associées aux catégories, aux régions et aux croisements catégorie x région.
- De plus l'échantillon de communautés de l'EAR N ne correspond pas exactement à la population de référence des communautés au 1<sup>er</sup> janvier N-3. En effet, parmi les communautés recensées lors de l'EAR N, de nouvelles communautés ont été ouvertes depuis le 1<sup>er</sup> janvier N-1, qui ne rentrent donc pas en compte dans le champ de la population de référence des communautés du RP N-3. Seule la partie de l'échantillon de communautés de l'EAR N présente dans l'image millésimée N-1 peut donc être calée sur des agrégats issus du RP N-3. Les communautés de l'EAR N qui ont été ouvertes après le 1<sup>er</sup> janvier N-1 ne participent pas au calage et leurs poids initiaux, tous égaux à 5, sont conservés.

### 3.9.3. Mise en œuvre du calage

Le calage de l'échantillon de communautés et des ménages qui y résident, au niveau France entière<sup>15</sup>, est mis en œuvre à l'aide de la macro SAS %Calmar. Pour les communautés de l'EAR N qui ont été ouvertes après le 1<sup>er</sup> janvier N-1, on conserve les poids de sondage. Afin de nous prémunir contre une trop grande dispersion des poids calés, nous utilisons la méthode Logit. Les paramètres de la macro SAS %Calmar sont choisis de façon à ce que les rapports de poids soient compris entre 0,2 et 3,5, ce qui garantit que les nouveaux poids soient compris dans l'intervalle [1 ; 17,5]<sup>16</sup>.

## 4. Impact sur la précision des estimations issues de l'EAR

### 4.1. Réduction de l'écart entre estimations EAR et estimations RP

Pour un paramètre d'intérêt donné, l'utilisation des nouveaux poids permet de réduire globalement les écarts entre l'estimation issue de l'EAR et celle issue du RP. En utilisant les anciennes pondérations, l'écart entre l'estimation EAR 2015 et l'estimation RP2015 de la population départementale est supérieure à +/- 2 % pour un quart des départements, comme le montre la figure 9 ci-dessous. En utilisant les nouvelles pondérations, seuls 9 départements se caractérisent par de tels écarts.

<sup>15</sup> Hors les COM : Saint-Pierre-et-Miquelon, Saint-Barthélemy et Saint-Martin, où l'on conserve les poids de sondage.

<sup>16</sup> Chaque communauté recensée a en effet un poids de sondage égal à 5.

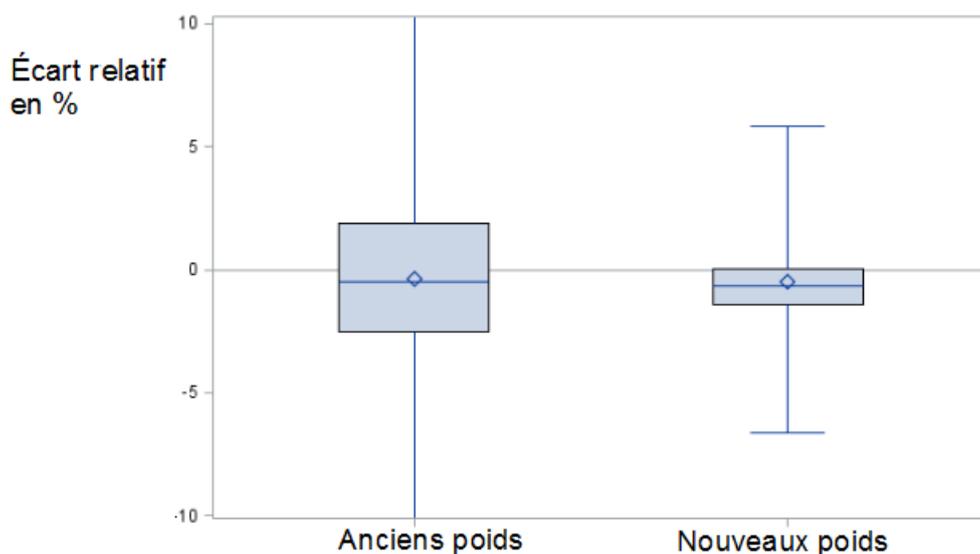


Figure 9 : Distribution des écarts relatifs (en %) entre les estimations de la population départementale obtenues par l'EAR 2015 d'une part et par le RP2015 d'autre part.

Cet « alignement » des estimations de l'EAR sur celles du RP obtenu grâce à l'utilisation des nouvelles pondérations peut également être observé pour l'estimation du total de populations plus spécifiques comme le nombre total d'actifs occupés à temps partiel par département. En revanche, de telles améliorations ne sont pas visibles pour l'estimation de pourcentages. En effet, les estimations issues de l'EAR 2015 utilisant les anciennes pondérations sont déjà proches des estimations du RP. C'est d'ailleurs pour cette raison que nous avons recommandé de ne communiquer que sur ces proportions dans les études menées sur l'EAR 2015 et l'EAR 2016 qui se basaient encore sur les anciennes pondérations. Par exemple, quel que soit le jeu de poids utilisé, pour 95 % des départements, l'estimation de la proportion de temps partiel parmi les actifs occupés d'après l'EAR ne diffère du pourcentage issu du RP que de 0,9 point au maximum.

#### 4.2. Réduction de la volatilité des estimations EAR au cours du temps

La difficulté de l'utilisation des EAR pour la réalisation d'études statistiques est liée à la volatilité des estimations d'un paramètre d'intérêt donné au cours du temps, ce qui est dû aux déséquilibres des groupes de rotation. Au niveau régional, le recours aux nouvelles pondérations permet de réduire cette volatilité, introduisant ainsi la possibilité de dégager des évolutions tendancielle de ces paramètres d'intérêt. À l'aide des nouvelles pondérations, on peut par exemple observer une stabilité de la population de la région Bourgogne-Franche-Comté depuis le début de la période où on dispose des nouvelles estimations, comme le montre la figure 10 ci-dessous. Une telle évolution n'aurait pas pu être observée à l'aide des anciennes pondérations qui ne permettaient de dégager aucune tendance dans la série de population de cette région.

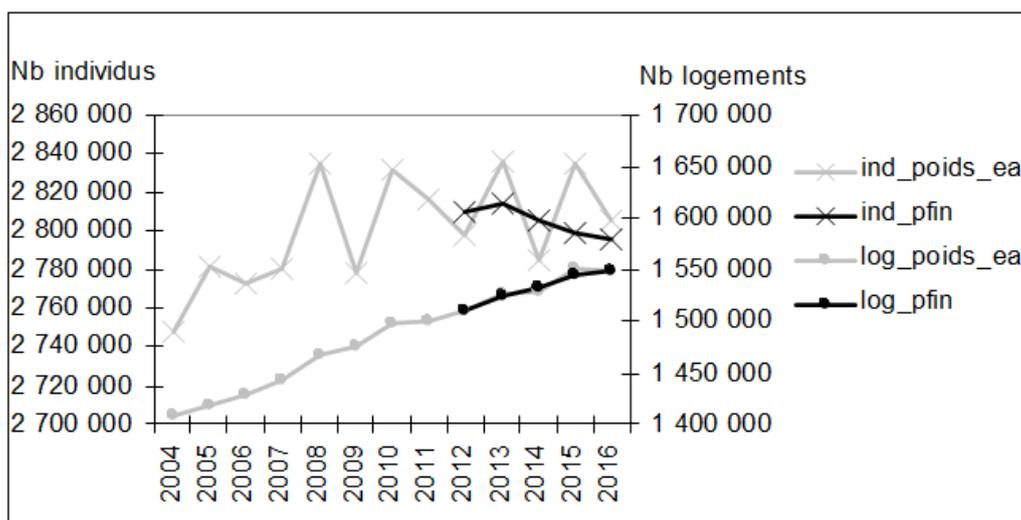


Figure 10 : Évolution des estimations EAR du nombre de logements et du nombre d'individus en Bourgogne-Franche-Comté entre 2004 et 2016, en utilisant les anciennes pondérations (\_poids\_ea) ou les nouvelles pondérations (\_pfin).

Les données de Fidéli sont disponibles à partir du millésime 2011. Il est donc en théorie possible de calculer les nouvelles pondérations à partir de l'EAR 2012. Toutefois les prototypes Fidéli (resp. millésimes 2011 et 2012) ne sont pas de qualité suffisante pour disposer d'un appariement correct avec les BSA (resp. millésimes 2012 et 2013). Ainsi, pour les EAR 2012 et 2013, les nouvelles pondérations ont été calculées uniquement pour le champ des petites communes. En grandes communes, on conserve les anciennes pondérations.

Au niveau communal également (pour les grandes communes), on observe que les nouvelles pondérations permettent de réduire la dispersion des estimations du nombre de logements et d'habitants au cours du temps, comme le montre par exemple la figure 11 pour la commune du Mée-sur-Seine.

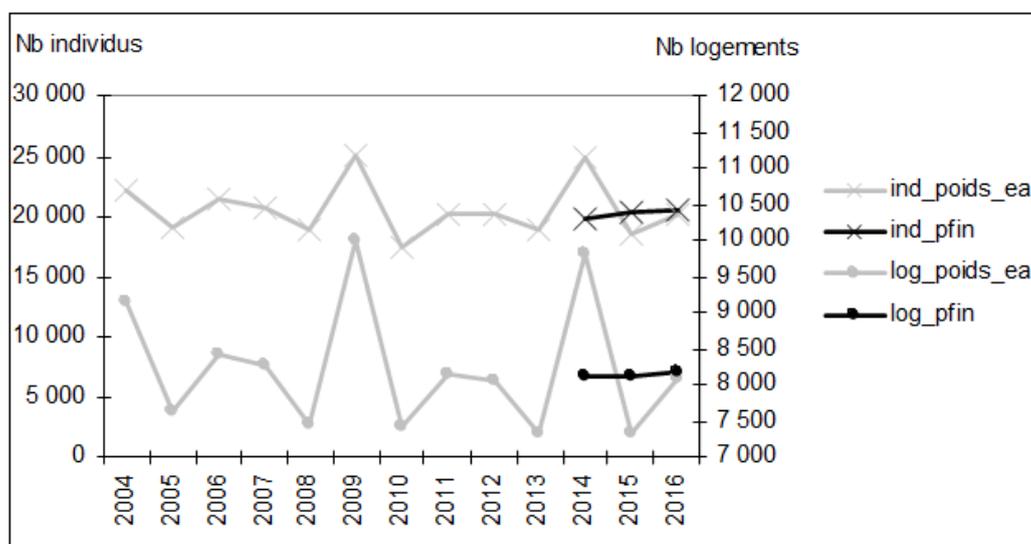


Figure 11 : Évolution des estimations EAR du nombre de logements et du nombre d'individus de la commune du Mée-sur-Seine entre 2004 et 2016, en utilisant les anciennes pondérations (\_poids\_ea) ou les nouvelles pondérations (\_pfin).

### 4.3. Réduction de la variance des estimations

Pour compléter l'approche empirique développée précédemment, l'impact de l'utilisation de la nouvelle méthode a aussi été mesuré au travers du gain associé à la précision d'échantillonnage.

Comme on pouvait s'y attendre, ce gain est particulièrement important sur le champ des petites communes et pour les indicateurs statistiques qui sont les proxy directs des variables qui ont servi au calage. Le gain de précision est plus réduit sur des populations rares pour lesquelles la corrélation avec les variables de calage est plus faible. Le gain de précision est également plus élevé au niveau départemental qu'aux niveaux régional et national, étant donné que la taille de l'échantillon importante garantissait déjà une précision suffisante sur les niveaux les plus agrégés.

## Bibliographie

[1] Bertrand P., Chauvet G., Christian B., Grosbras J.M. (Insee), « Les plans de sondage du nouveau recensement », VIII<sup>èmes</sup> Journées de méthodologie statistique, 16-17 décembre 2002.

[2] Godinot A. (Insee), « Pour comprendre le recensement de la population », 2005.

[3] Tallet F., Vallès V. « Partir de bon matin, à bicyclette... », *Insee Première* n°1629, janvier 2017.

[4] Robert-Bobée I., Vallès V., « Les Pacs à l'Ouest, les mariages à l'Est : une répartition des types d'unions différente selon les territoires », *Insee Première* n°1682, janvier 2018.