

## **L'amélioration de la précision des enquêtes annuelles de recensement par la mobilisation de sources fiscales**

**Sébastien Hallépée, Pierre-Arnaud Pendoli  
et Olivia Sautory**





Institut national de la statistique et des études économiques

---

F2020-03

# **L'amélioration de la précision des enquêtes annuelles de recensement par la mobilisation de sources fiscales**

Sébastien Hallépée, Pierre-Arnaud Pendoli et Olivia Sautory\*

novembre 2020

---

Direction des Statistiques Démographiques et Sociale - Timbre F001  
88, avenue Verdier - CS 70058 - 92541 MONTRouGE CEDEX - France  
Tél. : 33 (1) 87 69 62 82 - E-mail : [dg75-f001@insee.fr](mailto:dg75-f001@insee.fr) - Site Web Insee : <http://www.insee.fr>

*Ces documents de travail ne reflètent pas la position de l'Insee et n'engagent que leurs auteurs.*

*Working papers do not reflect the position of INSEE but only their author's views.*

---

\* Insee, DSDS, département de la Démographie, division Méthodes et traitements des recensements

## **L'amélioration de la précision des enquêtes annuelles de recensement par la mobilisation de sources fiscales**

### **Résumé**

Le dernier recensement de la population (RP) date de 1999. Depuis 2004, la France a rénové son opération de recensement qui repose désormais sur un cycle de cinq ans : chaque année, seule une partie de la population est recensée au cours d'une enquête annuelle 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 déséquilibrés en termes de nombre de logements et de population. Si ce phénomène n'est pas de nature à affecter la qualité des estimations issues du recensement de la population (dans la mesure où celui-ci mobilise conjointement les cinq groupes de rotation), on observe dans certaines régions d'importantes fluctuations dans les séries de populations calculées à partir d'une seule 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 l'unicité de la période d'observation et une disponibilité plus rapide des résultats en cas de modifications du questionnaire. Par exemple, une modification du questionnaire de la feuille de logement permettant une meilleure identification des liens entre les personnes résidant dans le logement a eu lieu en 2018. Les résultats de cette nouvelle analyse ménage-famille ont conduit à la publication d'une première étude dès le début de l'année 2020 à partir de l'EAR, contre 2023 s'il avait fallu attendre les résultats du RP.

Jusqu'à l'EAR 2018, les pondérations de ces enquêtes étaient obtenues à partir du poids de sondage et d'un calage sur un nombre limité de variables. Depuis l'EAR 2019, une nouvelle méthodologie de calage des EAR est mise en œuvre. Celle-ci vise à améliorer la précision des données estimées à partir d'une seule EAR. Elle repose notamment sur l'utilisation de variables auxiliaires issues du Fichier démographique des logements et des individus (Fidéli) constitué à partir de sources fiscales.

La nouvelle pondération permet de lisser les fluctuations des séries de population et de nombre de logements dues aux déséquilibres des groupes de rotation. Ce phénomène reflète notamment les gains de précision<sup>1</sup> permis par l'enrichissement des variables auxiliaires utilisées dans le redressement. Comme la pondération utilisée jusqu'à l'EAR 2018, cette nouvelle pondération ne vise pas à corriger la date de validité des estimations dans les communes de 10 000 habitants ou plus, qui reste liée à la date de référence de la base de sondage utilisée pour le tirage de l'échantillon d'adresses à enquêter (juillet de l'année précédant l'EAR).

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

---

<sup>1</sup> Les auteurs remercient Martin Chevalier (DMTR) pour sa contribution à l'évaluation des gains en matière de précision des estimateurs.

## Improving the accuracy of annual census surveys using tax sources

### Abstract

The last population census was conducted in 1999. Since 2004, France has renewed its census operation, which is now based on a five-year cycle: each year, only part of the population is enumerated in the annual census survey (ACS). A population census is thus based on the aggregation of data from five ACSs.

The rotation groups, on which the ACSs are built, are sometimes unbalanced in terms of number of dwellings and population. While this does not affect the quality of the population census estimates (since the population census rely on all five rotation groups), in some regions there are significant fluctuations in the population series calculated from a single ACS, which has hitherto hindered the further use of this source.

However, the use of data from a single ACS offers several advantages, including the uniqueness of the observation period and increased timeliness in the event of changes in the questionnaire. For example, a modification of the housing sheet questionnaire allowing for a better identification of the links between persons residing in the dwelling took place in 2018. The results of this new household-family analysis led to the publication of a first study in early 2020 from the EAR, as opposed to 2023 if it had been necessary to wait for the results of the census.

Until the ACS 2018, the weights for these surveys were obtained from the sample weight and a calibration on a limited number of variables. Beginning with the 2019 EAR, a new methodology for calibrating the ACS is being implemented. This methodology aims to improve the accuracy of data estimated from a single ACS. It is based, among other things, on the use of auxiliary variables from the "Fichier démographique des logements et des individus" (Fidéli), which is compiled from tax sources.

The new weighting makes it possible to smooth fluctuations in the population and housing number series due to imbalances in rotation groups. This reflects in particular the gains in precision made possible by the enrichment of the auxiliary variables used in the adjustment. Like the weighting used up to ACS 2018, this new weighting does not aim at correcting the validity date of estimates in municipalities with 10,000 or more inhabitants, which remains linked to the reference date of the sampling frame used for the selection of the sample of addresses to be surveyed (July of the year preceding the ACS).

**Keywords:** census, administrative data, sampling, calibration on margins

## Table des matières

L'amélioration de la précision des enquêtes annuelles de recensement par la mobilisation de sources fiscales.....	1
Résumé.....	1
Introduction.....	4
I. Plans de sondage des enquêtes annuelles de recensement.....	6
I.1 Logements ordinaires de France métropolitaine.....	6
I.1.1 En petites communes.....	6
I.1.2 En grandes communes.....	7
I.2 Logements ordinaires des départements d'outre-mer (DOM).....	8
I.2.1 En petites communes.....	8
I.2.2 En grandes communes.....	8
I.3 Population en communauté et logements dans l'enceinte d'une communauté.....	10
I.4 Habitations mobiles, sans-abris et bateliers.....	10
I.5 Les collectivités d'outre-mer (COM).....	10
I.6 Bilan des plans de sondage.....	10
II. Déséquilibre des groupes de rotation et volatilité des estimations.....	11
II.1 Déséquilibre des groupes de rotation.....	11
II.2 Une pondération actuelle perfectible.....	13
II.3 Volatilité des estimations de population à partir de l'EAR.....	14
III. Nouvelle méthode de calage des EAR.....	15
III.1 Détermination de la nouvelle source auxiliaire.....	15
III.2 Méthode de repondération.....	15
III.2.1 Logements ordinaires des petites communes de métropole.....	15
III.2.2 Logements ordinaires des petites communes des DOM.....	16
III.2.3 Logements ordinaires des grandes communes de métropole.....	16
III.2.4 Logements ordinaires des grandes communes des DOM.....	20
III.2.5 Communautés.....	21
III.3 Résumé des méthodes de repondération.....	22
IV. Impact sur la précision des estimations issues de l'EAR.....	23
IV.1 Distribution des nouvelles pondérations.....	23
IV.2 Réduction de la variance.....	23
IV.3 Réduction de la volatilité empirique des estimations de population à partir des EAR.....	26
V. Comparabilité des données de l'EAR et du RP.....	30
V.1 Des données qui n'ont pas la même temporalité.....	30
V.1.1 Comparaison des données EAR et RP en volume.....	30
V.1.2 Comparaison des données EAR et RP en taux.....	31
V.2 Comparaison des données EAR et RP aux niveaux régional et départemental.....	32
V.4 Comparaison des données EAR et RP en évolution.....	33
Conclusion.....	35
Bibliographie.....	36
Annexe : méthode de calage simultané envisagée pour les grandes communes de La Réunion.....	37

## Introduction

Le recensement de la population (RP) est basé sur un cycle de cinq ans depuis 2004 ([1] et [2]). 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 fin 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 »), la collecte d'un échantillon d'adresses représentant 8 % des logements de la commune. 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 personnes faisant partie de ménages « ordinaires », c'est-à-dire celles vivant dans des logements classiques, représentent la très grande majorité de la population (97 %). Les personnes résidant dans des communautés sont recensées selon un protocole spécifique. En particulier, la collecte des communautés est en général concentrée sur une seule des cinq années du cycle pour la plupart des grandes communes. De même, les personnes vivant dans des habitations mobiles et les sans-abris (HMSA) ne sont pas recensés de la même façon. Enquêtés exhaustivement tous les cinq ans au sein de chaque commune, 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 (DOM) 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 principalement réalisées par le suivi des permis de construire et de démolir et les échanges avec les communes.

Ces particularités impliquent que la méthode de calcul des populations 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 (EAR) a été considérée pendant longtemps surtout comme un élément intermédiaire utilisé en cumul de cinq années successives pour produire le 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 système statistique public et sa diffusion est assortie de recommandations d'utilisation restrictives. En effet, le produit est largement perfectible et conduit à des séries de résultats très volatiles, même pour des indicateurs portant sur des territoires assez vastes comme la population des régions. L'exploitation des données d'une seule EAR offre pourtant plusieurs avantages (par rapport au recensement de la population), notamment la fraîcheur des données, l'unicité de la période d'observation et des résultats plus rapides en cas de modifications du questionnaire. Par exemple, une modification du questionnaire de la feuille de logement permettant une meilleure identification des liens entre les personnes résidant dans le logement a eu lieu en 2018. Les résultats de cette nouvelle analyse ménage-famille ont conduit à la publication d'une première étude dès le début de l'année 2020 à partir de l'EAR, contre 2023 s'il avait fallu attendre les résultats du RP.

L'EAR a déjà servi de base à des opérations de diffusion complémentaires au RP. 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]. 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 présentés dans ce document de travail 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) sur des niveaux géographiques agrégés. 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.

L'objectif des travaux qui sont présentés dans ce document est d'améliorer la précision des estimateurs calculables à partir d'une seule EAR, en particulier les indicateurs en niveau.

Ce document de travail présente la méthode de pondération de l'EAR qui est mise en place à partir de l'EAR 2019. Pour mieux comprendre la méthode, une première partie rappelle le plan de sondage de l'EAR. La deuxième partie du document revient en détail sur le sujet du déséquilibre entre groupes de rotation du recensement, qui constitue une source majeure d'imprécision. La troisième partie présente les méthodes de calage mises en œuvre sur les différentes catégories de population et tailles et types de communes à partir de l'EAR 2019. La quatrième partie enfin évalue les gains en termes de précision. La cinquième partie fournit des éléments permettant d'expliquer certains écarts observés entre les résultats d'une EAR et ceux issus du RP, en particulier la date de référence de la base de sondage dans les communes de 10 000 habitants ou plus.



# I. Plans de sondage des enquêtes annuelles de recensement

Les EAR recensent les logements et individus de France métropolitaine, Guadeloupe, Martinique, Guyane, Réunion ainsi que des collectivités d'outre-mer Saint-Pierre-et-Miquelon, Saint-Barthélemy et Saint-Martin. Mayotte n'a pas encore intégré le processus des EAR ; des recensements exhaustifs y ont été réalisés en 2007, 2012 et 2017.

Pour répartir la collecte sur cinq enquêtes annuelles, cinq groupes de rotation (GR) ont été définis lors de la mise en place du nouveau recensement en 2004. Le tirage des groupes de rotation 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 de moins de 10 000 habitants *versus* grandes communes de plus de 10 000 habitants.
- 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).

De manière générale, 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>2</sup>.

Le tableau 1 indique la répartition de la population suivant leur lieu de résidence et leur catégorie de logement et de commune pour la France métropolitaine.

Tableau 1 : répartition de la population

<b>France métropolitaine</b>	<b>97,1 %</b>
Individus vivant en logement ordinaire	94,5 %
<i>grandes communes</i>	45,3 %
<i>petites communes</i>	49,3 %
Individus vivant en logement dans l'enceinte d'une communauté <sup>3</sup>	0,3 %
Individus vivant en communauté	2,0 %
Habitations mobiles et sans-abris	0,2 %
<b>DOM (Guadeloupe, Martinique, Guyane, La Réunion)</b>	<b>2,9 %</b>
<b>COM (Saint-Pierre-et-Miquelon, Saint-Barthélemy, Saint-Martin)</b>	<b>0,1 %</b>

Champ : individus de France métropolitaine, DOM (hors Mayotte) et COM.

Source : RP 2015.

## I.1 Logements ordinaires de France métropolitaine

### I.1.1 En petites communes

La population rattachée aux logements ordinaires des petites communes de métropole rassemble près de la moitié de la population française (tableau 1).

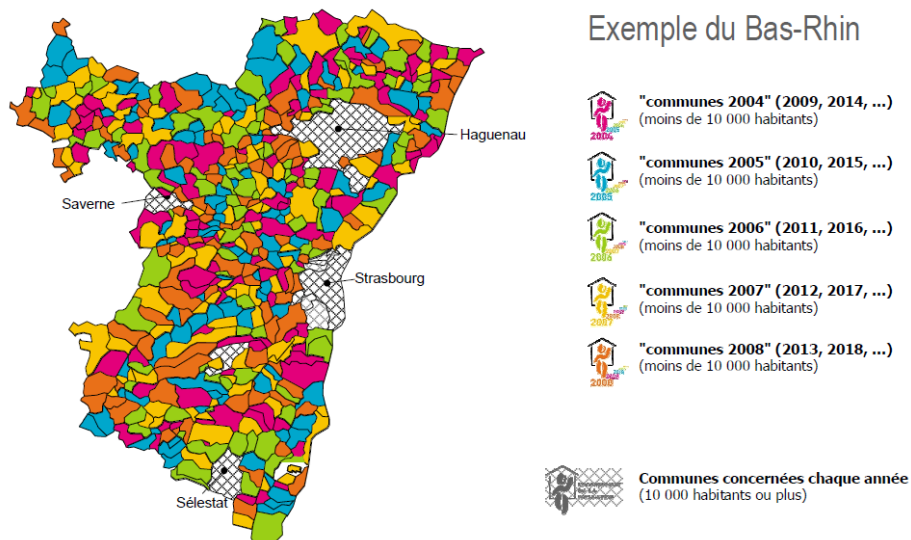
La commune constitue l'unité statistique. Au sein de chaque région, les petites communes sont réparties aléatoirement en cinq groupes de rotation (figure 1), par tirage équilibré à probabilités égales sur des variables de type de logement et des variables socio-démographiques issues du RP 1999 : le nombre de logements, le nombre de logements en immeuble collectif, la population selon cinq classes d'âge (moins de 20 ans ; 20-39 ans ; 40-59 ans ; 60-74 ans ; 75 ans et plus), la population par sexe et la population associée à chaque département.

Lors d'une EAR, les logements ordinaires du groupe de rotation de l'année sont recensés exhaustivement.

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

<sup>3</sup> Ou encore en « logements de fonction ».

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



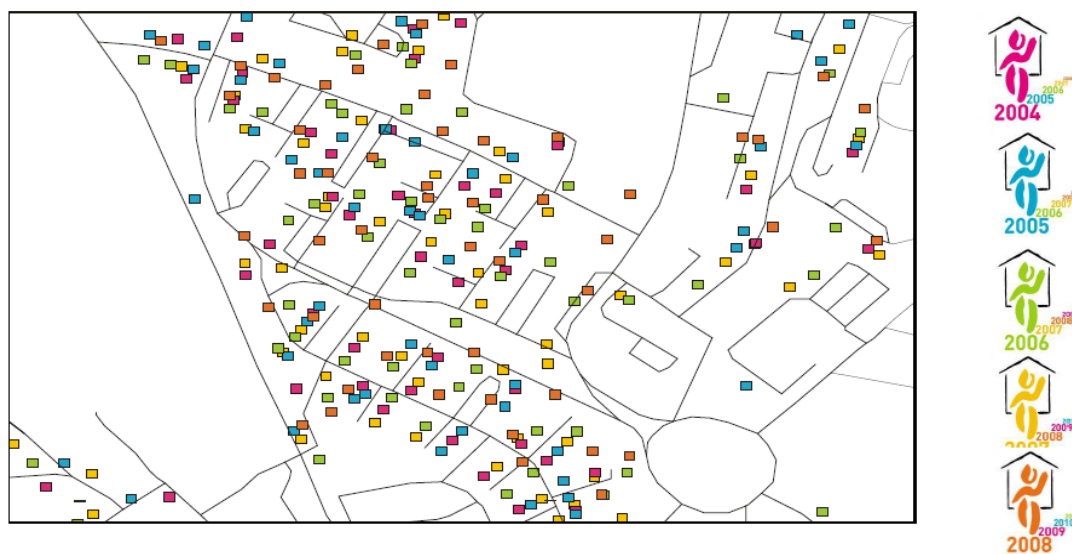
### I.1.2 En grandes communes

La population rattachée aux logements ordinaires des grandes communes de métropole rassemble 45 % de la population totale France entière (tableau 1).

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

Les *adresses d'habitation* (99,8 % des adresses) sont réparties en cinq groupes de rotation, par tirage aléatoire équilibré à probabilités égales sur les mêmes variables que celles utilisées pour les petites communes de métropole (excepté la population par département) (figure 2).

Figure 2 : représentation schématique des groupes de rotation d'adresses d'une grande commune de France métropolitaine



On distingue trois types d'adresses d'habitation : les grandes adresses (rassemblant au moins 60 logements, seuil pouvant être supérieur pour certaines communes de sorte que l'ensemble des grandes adresses ne rassemblent pas plus de 10 % des logements de la commune), les petites adresses nouvelles et les petites adresses connues. Lors d'une EAR, les grandes adresses et les petites adresses nouvelles du groupe de rotation sont recensées exhaustivement. Les petites adresses connues sont en revanche échantillonnées,

afin que la collecte porte sur environ 40 % des logements du groupe de rotation. Les logements recensés chaque année représentent in fine environ 8 % des logements ordinaires de la commune.

Les *adresses d'hôtels, campings et résidences hôtelières* (0,2 % des adresses) sont réparties entre les cinq groupes de rotation par tirage aléatoire équilibré à probabilités égales (1/5) sur le nombre d'adresses. Chaque année, l'ensemble des adresses du groupe de rotation est recensé.

## **I.2 Logements ordinaires des départements d'outre-mer (DOM)**

### **I.2.1 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. A titre d'exemple, le DOM comportant le plus de petites communes est la Martinique, avec seulement 22 petites communes. La seule variable d'équilibrage utilisée est la population au RP1999. Comme en métropole, chaque année, on recense tous les logements ordinaires des petites communes du groupe de rotation.

### **I.2.2 En grandes communes**

Chaque grande commune fait l'objet d'un plan de sondage indépendant. L'adresse constitue l'unité statistique. Comme dans les grandes communes de métropole, les adresses des grandes communes des DOM sont réparties en cinq groupes de rotation. À la différence des grandes communes de métropole cependant, les groupes de rotation sont constitués par un processus à deux niveaux :

- les adresses sont rassemblées en ensembles d'adresses contiguës appelés « îlots » (figure 3) ;
- les groupes de rotation d'adresses sont constitués par le biais d'un tirage aléatoire d'îlots à probabilités égales et équilibré sur les mêmes variables que celles utilisées pour les grandes communes de métropole (figure 4).

Comme dans les grandes communes de métropole, l'échantillon d'adresses à enquêter pour une EAR donnée est obtenu par tirage d'environ 40 % des adresses du groupe de rotation, représentant environ 8 % des logements de la commune<sup>4</sup>.

---

<sup>4</sup> Jusqu'à l'EAR 2018 incluse, le tirage de l'échantillon d'adresses à enquêter dans les grandes communes des DOM ne faisait pas intervenir de strate exhaustive (pas de strate de grandes adresses ou de petites adresses nouvelles) ; c'est le cas à partir de l'EAR 2019.

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

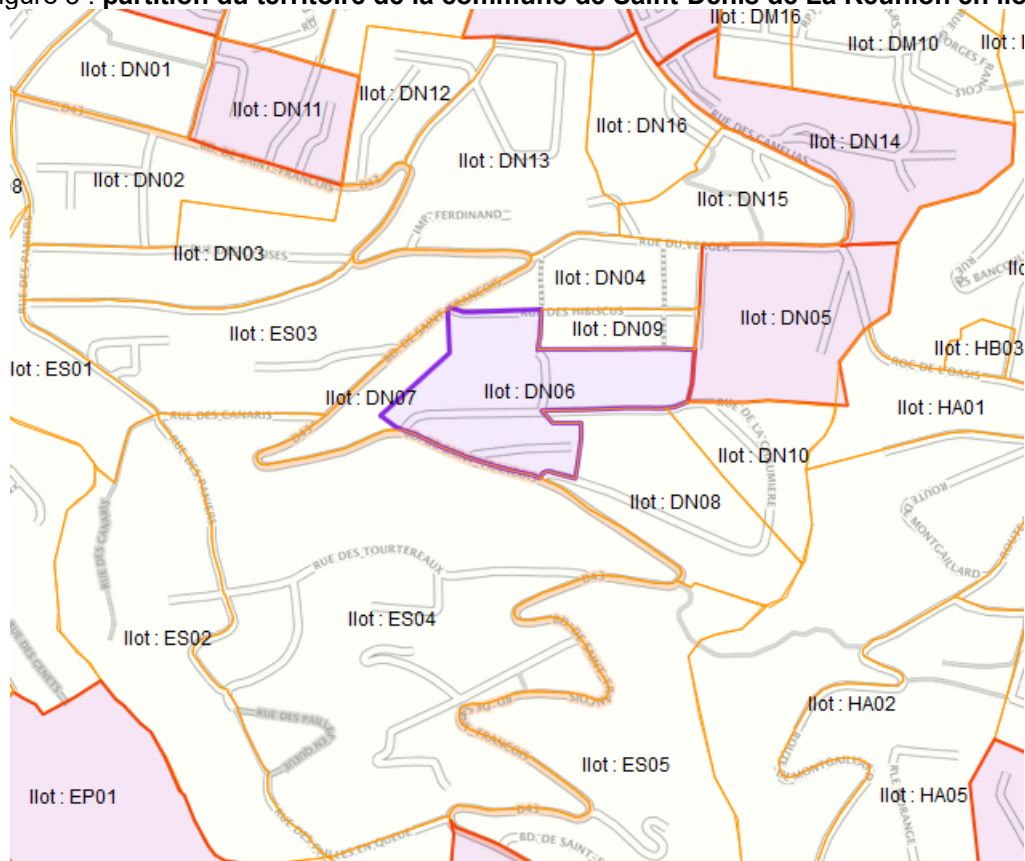
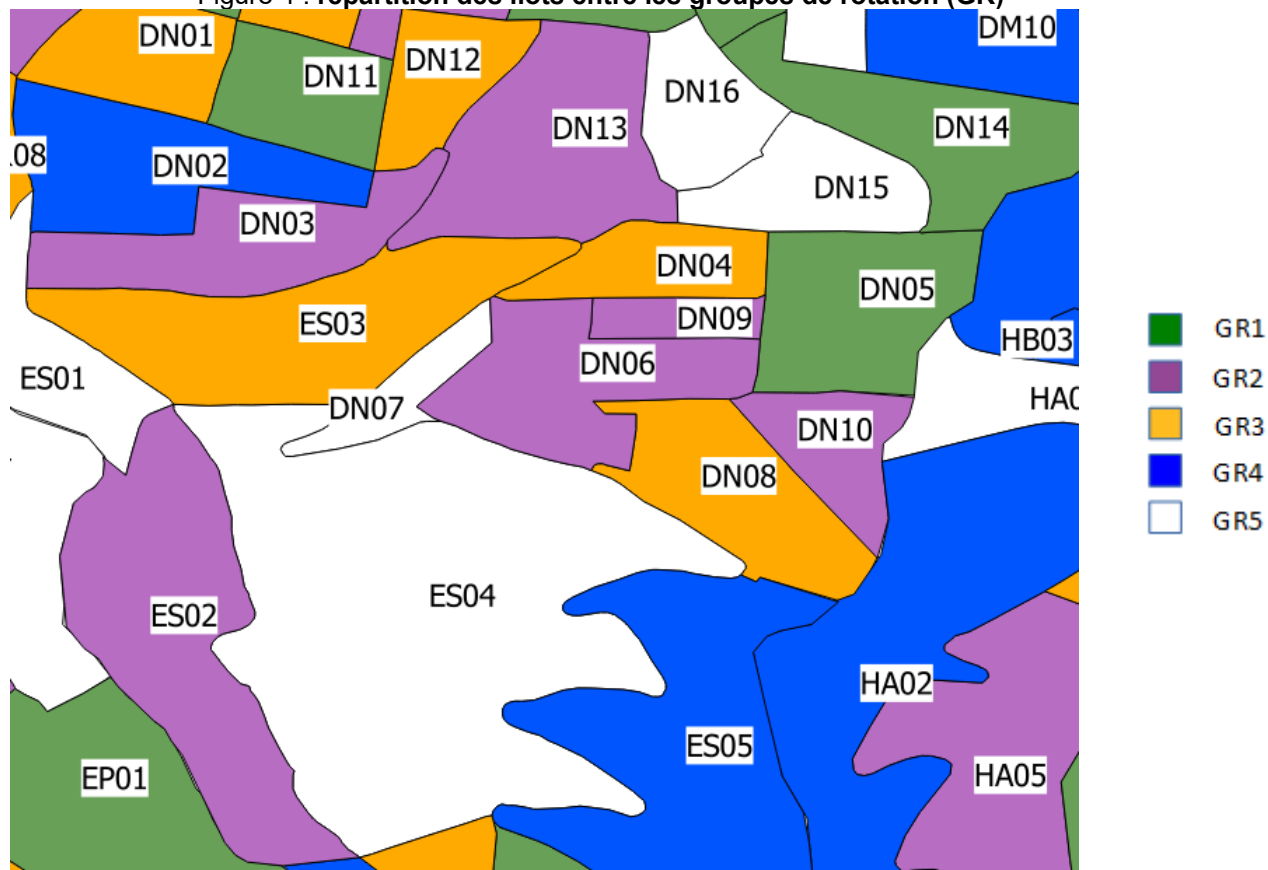


Figure 4 : répartition des îlots entre les groupes de rotation (GR)



### I.3 Population en communauté et logements dans l'enceinte d'une communauté

La population rattachée aux communautés et aux logements dans l'enceinte d'une communauté représente environ 2 % de la population (tableau 1). La communauté constitue l'unité statistique : les éventuels logements situés dans son enceinte sont recensés en même temps que le reste de la population de la communauté. Chaque communauté est recensée l'année correspondant à son groupe de rotation :

- En petites communes, les communautés appartiennent au même groupe de rotation que les logements ordinaires de la commune où elles sont situées.
- En grande commune :
  - dans la majorité des cas les communautés sont recensées en une seule fois sur un cycle de cinq ans, le groupe de rotation correspondant ayant été choisi de manière déterministe ;
  - dans les DOM et dans certaines grandes communes de métropole, en général les plus peuplées, les communautés ont été réparties de manière déterministe entre plusieurs groupes de rotation (2 à 5 selon les cas).

### I.4 Habitations mobiles, sans-abris et bateliers

En petites communes, les habitations mobiles et les sans-abris (HMSA) appartiennent au même groupe de rotation que les logements ordinaires de leur commune : ils sont recensés exhaustivement l'année correspondant au groupe de rotation. En grandes communes, les HMSA sont recensés exhaustivement tous les cinq ans à partir de 2006. Les bateliers sont également recensés une fois tous les cinq ans à partir de 2006.

### I.5 Les collectivités d'outre-mer (COM)

Le recensement à Saint-Martin, collectivité d'outre-mer dont la population excède 10 000 habitants et qui appartenait à la Guadeloupe au moment de la rénovation du recensement, est similaire à celui des autres grandes communes des DOM. Les collectivités d'outre-mer de Saint-Pierre-et-Miquelon et de Saint-Barthélemy, dont la population est inférieure à 10 000 habitants sont, comme les autres petites communes des DOM, recensées une fois tous les cinq ans.

### I.6 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 (tableau 2).

Tableau 2 : **tailles des échantillons et taux de sondage selon les champs de l'EAR**

Catégorie de population	Logements ordinaires				Communauté
	Métropole		DOM		
	Petites	Grandes	Petites	Grandes	
Nombre d'individus recensés à l'EAR	6 480 000	2 320 000	63 000	130 000	330 000
<i>Nature de l'unité statistique</i>	<i>Commune</i>	<i>Adresse</i>	<i>Commune</i>	<i>Adresse</i>	<i>Communauté</i>
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%

Le plan de sondage à l'œuvre pour les logements ordinaires de grandes communes est plus complexe (deux phases) et conduit à un taux de sondage global plus faible que sur les autres champs. Néanmoins, le nombre d'unités échantillonnées, bien supérieur dans le cas des logements ordinaires de grandes

communes (car il s'agit d'adresses et non de communes ou de communautés) peut conduire en pratique à des estimateurs calculés à partir d'une seule EAR plus précis en grande commune que sur les autres champs.

Autrement dit et qualitativement, on s'attend à ce que le nombre sensiblement plus faible d'unités échantillonnées en petites communes (et en communauté) conduise à un moins bon équilibrage des groupes de rotation tirés lors de la rénovation du recensement de la population. Ce moins bon équilibrage est susceptible d'affecter sensiblement la précision des estimateurs calculés à partir d'une seule EAR.

## II. Déséquilibre des groupes de rotation et volatilité des estimations

### II.1 Déséquilibre des groupes de rotation

Si l'équilibrage des groupes de rotation était exact, on devrait pouvoir estimer parfaitement la population des ménages au 1er janvier 1999 de chaque département à l'aide d'un groupe de rotation, cette population étant une des variables auxiliaires intégrées dans l'algorithme d'équilibrage.

On appréhende le déséquilibre des groupes de rotation (GR) à l'initialisation<sup>5</sup> de la manière suivante : notons  $X_d$  la population du département  $d$  au RP 1999. La population du département  $d$  estimée par le GR  $i$  est noté  $\hat{X}_{d;GRi}$  et vaut simplement la population dans le GR du département au RP1999 multipliée par 5.

Les groupes de rotation sont considérés comme parfaitement équilibrés si :  $\forall i \hat{X}_{d;GRi} = X_d$ .

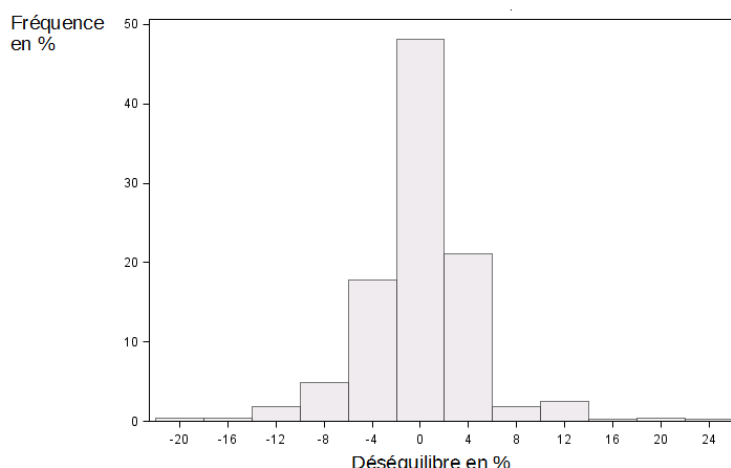
Toutefois, lorsque l'on tente d'estimer pour chaque département la population des ménages au RP 1999 à partir des groupes de rotation, il apparaît que ces estimations ne sont pas exactes : le déséquilibre d'un GR

se mesure par l'écart relatif  $\frac{\hat{X}_{d;GRi} - X_d}{X_d}$ .

En France métropolitaine, la moitié des estimations de populations départementales par groupe de rotation s'écartent d'au moins 2 % du vrai total au 1er janvier 1999 (figure 5). Dans 7 départements, au moins une estimation départementale s'écarte d'au moins 14 % du vrai total au 1<sup>er</sup> janvier 1999. Il s'agit le plus souvent de départements dont le nombre d'habitants est plus petit que la moyenne (Hautes-Alpes, Territoire-de-Belfort, Ariège, Lot, Haute-Marne, Hautes-Pyrénées, Haute-Saône). L'existence de ces écarts dès l'initialisation du recensement rénové de la population est la conséquence d'un équilibrage imparfait des GR de petites communes en termes de population départementale au moment du tirage : celui-ci a en effet été mené par région et les contraintes correspondants à la population par département ont été les premières à être levées pour permettre la convergence de l'algorithme (lors de la phase dite d'atterrissage, cf. [7]).

<sup>5</sup> Les évaluations présentées dans cette partie portent sur les groupes de rotation tels que définis à l'initialisation du recensement rénové de la population en 2004. La composition communale actuelle de ces groupes est affectée par les changements de géographie survenus depuis lors (fusion de communes en particulier). Ce phénomène est de nature à amplifier les déséquilibres initiaux mis en évidence dans cette partie.

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



Note : les pourcentages présentés sur l'axe des abscisses correspondent à un intervalle de valeurs comprises entre plus ou moins 2 points par rapport à cette valeur. Par exemple, 0 % correspond à l'intervalle [-2 %: +2%]

Lecture : 49 % des estimations de la population départementale s'écartent d'au plus 2 % par rapport au vrai total.

Champ : France métropolitaine. 480 estimations de la population des ménages des départements, correspondant à 5 estimations pour chacun des 96 départements.

Source : RP1999.

Pour chaque département d, on calcule le déséquilibre pour les petites (respectivement grandes) communes de la manière suivante :

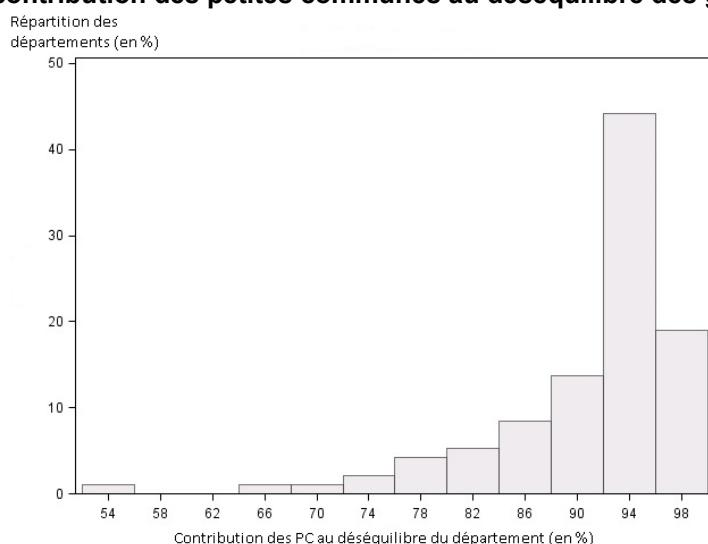
$$I_{PC;d} = \sum_{i=1}^5 |pop\hat{PC}_{d;GRi} - popPC_d|, \quad pop\hat{PC}_{d;GRi} = 5 * population\ en\ petites\ communes\ appartenant\ au\ GRi$$

$$I_{GC;d} = \sum_{i=1}^5 |pop\hat{GC}_{d;GRi} - popGC_d|, \quad pop\hat{GC}_{d;GRi} = 5 * population\ résidant\ aux\ adresses\ appartenant\ aux\ GRi\ des\ GC$$

La contribution des petites communes au déséquilibre des groupes de rotation pour le département d se calcule ainsi :  $\frac{I_{PC;d}}{I_{PC;d} + I_{GC;d}}$ . 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 (figure 6). 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 en grandes communes. Calculées à partir d'échantillons plus petits qu'en grandes communes, les estimations de population en petites communes se caractérisent par une variance plus forte au niveau départemental.

Figure 6 : contribution des petites communes au déséquilibre des groupes de rotation



Lecture : pour un département sur cinq, les PC contribuent à plus de 96 % du déséquilibre des groupes de rotation vis-à-vis de la population totale du département.

Champ : 95 départements de France métropolitaine. Paris est exclu (pas de petites communes).

Source : RP1999.

## II.2 Une pondération actuelle perfectible

Dans la méthode de pondération utilisée jusqu'à l'EAR 2018, les logements et les habitants faisaient l'objet d'un calage dans l'EAR sur le champ de France métropolitaine et des DOM, hors HMSA.

En effet :

- les individus en HMSA, présents une EAR sur cinq, ont un poids de 1 ;
- les logements et individus des collectivités d'outre-mer de Saint-Pierre-et-Miquelon et Saint-Barthélemy, présents eux aussi une EAR sur cinq, ont également un poids de 1 ;
- les logements et individus situés aux adresses échantillonnées chaque année à Saint-Martin ont le poids de sondage de ces adresses.

L'ancienne pondération, produite jusqu'à l'EAR 2018, diffère selon les catégories de population et le champ géographique. Elle repose sur peu de variables auxiliaires, essentiellement de niveau logement, souvent datées (cf. tableau 3).

Tableau 3 : ancienne pondération de l'EAR

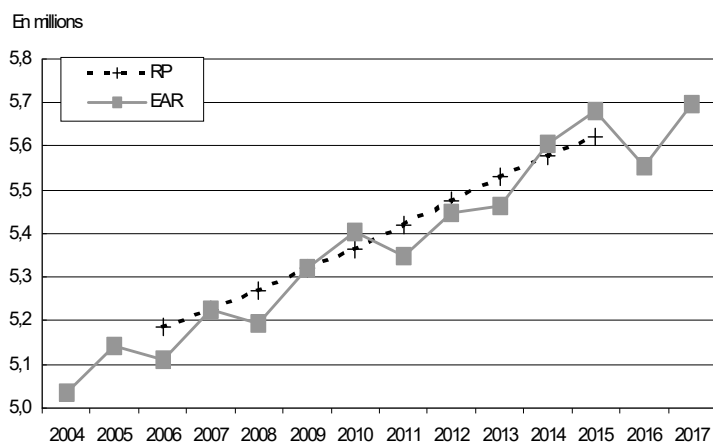
Champ	Marges	Calcul de l'ancien poids
<b>Petites communes</b>		
Résidences principales (y.c. logements dans l'enceinte d'une communauté) Individus vivant en résidences principales Individus vivant en communauté	Région	$\frac{\text{total résidences principales au RP}_{(N-3)}}{\text{nb résidences principales du GR au RP}_{(N-3)}}$
Résidences non principales	Région	$\frac{\text{total résidences non principales au RP}_{(N-3)}}{\text{nb résidences non principales du GR au RP}_{(N-3)}}$
<b>Grandes communes – France métr.</b>		
Logements ordinaires du RIL et leurs individus		
Grande adresse (GA)	Région	$\frac{\text{total logements en GA sur la BSA corrigée de la collecte}}{\text{nb logements en GA échantillon}}$
Petite adresse (PA)	Commune	$5 \times \alpha \frac{\text{total logements en PA sur la BSA corrigée de la collecte}}{\text{nb logements pondéré en PA échantillon}}$ où $\alpha = 1$ si petite adresse nouvelle, $\alpha = \frac{1}{\text{taux de sondage des PAC}}$ si petite adresse connue
Hôtels hors RIL habitation et leurs individus		1
<b>Grandes communes – DOM</b>		
Logements ordinaires et leurs individus	Commune	$\frac{\text{nblog RP99}}{\text{nblog RP99 du GR}} \times \frac{\text{nblog GR enquête carto}}{\text{nblog échantillon}}$
<b>Grandes communes – Communautés</b>		
Logements dans l'enceinte d'une communauté		5
Individus	Région	Les poids initiaux sont initialisés par catégorie de communautés au niveau national en calculant le rapport du nombre de communautés dans le répertoire des communautés sur le nombre de communautés dans l'échantillon, et ce par catégorie de communautés. On opère ensuite un calage, par région, sur les marges (sexe et tranches d'âge) du RP (N-3).



## II.3 Volatilité des estimations de population à partir de l'EAR

L'ancienne pondération ne permet pas de corriger complètement les déséquilibres des groupes de rotation. Comme le montre la figure 7 pour la région Occitanie, 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.

Figure 7 : évolution des estimations de la population d'Occitanie entre 2004 et 2016, à partir du RP d'une part et de l'EAR de l'autre part



Champ : population des ménages en logement ordinaire.

Source : EAR de 2004 à 2017 et RP de 2006 à 2015.

Dans l'état actuel des choses, il est donc impossible d'estimer de manière fiable le niveau et l'évolution de la population à partir d'une seule 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 vise notamment à pouvoir estimer les principaux indicateurs démographiques, tels que la population, aux niveaux des plus grandes communes, des départements et des régions.

### III. Nouvelle méthode de calage des EAR

#### III.1 Détermination de la nouvelle source auxiliaire

Afin d'améliorer les estimations de l'EAR, il est nécessaire d'identifier une source auxiliaire répondant aux critères suivants :

- Des variables socio-démographiques corrélées aux variables d'intérêt du recensement doivent être disponibles et mises à jour régulièrement de manière exhaustive. Vis-à-vis de ce critère, l'ancienne méthode de repondération n'est pas satisfaisante.
- Les variables auxiliaires doivent être connues pour chaque unité de l'échantillon : pour chaque petite commune d'une part et pour chaque adresse échantillonnée en grande commune d'autre part.
- Les totaux de ces variables auxiliaires doivent pouvoir être calculés à un niveau agrégé : les régions ou les départements pour l'échantillon de petites communes ; les communes pour les échantillons d'adresses des grandes communes.

Dans cette perspective, c'est sur le Fichier démographique des logements et des individus (Fidéli) que repose principalement la nouvelle méthode de repondération des EAR. Ce fichier 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. Selon le niveau auquel elles sont mobilisées (communes ou adresses), des algorithmes d'appariements ont été développés (cf. encadré 1). Pour caler l'EAR de l'année N, les données sont issues du dernier millésime de Fidéli disponible, c'est-à-dire celui relatif à l'année N-1. Du point de vue de la précision des estimateurs, le calage sera 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.

La source Fidéli ne peut cependant pas être mobilisée sur certains champs (individus appartenant à une communauté, petites communes des DOM) : dans ce cas, c'est le dernier cycle RP disponible qui est utilisé comme information auxiliaire pour la repondération de l'EAR, c'est-à-dire le RP N-3.

#### III.2 Méthode de repondération

Dans cette partie, les nouvelles méthodes de repondération sont présentées pour chaque sous-population. Dans l'élaboration de la nouvelle pondération, les unités de calage coïncident avec les unités d'échantillonnage. Pour les communautés par exemple, les personnes recensées en communauté, que ce soit en logements ou dans la communauté elle-même, font l'objet de la même procédure de calage.

##### III.2.1 Logements ordinaires des petites communes de métropole

Pour le calage des logements ordinaires et de leurs habitants recensés en petites communes de France métropolitaine à l'EAR 2017, la méthode de repondération consiste à suivre les étapes suivantes :

- **Calcul des variables auxiliaires**

Pour chaque petite commune, on calcule à partir des données de Fidéli 2016 les variables auxiliaires suivantes : nombre de logements, nombre de logements en immeubles collectifs, nombre d'habitants, nombre d'hommes, nombre de femmes<sup>6</sup>, nombre de personnes selon cinq classes d'âge.

Par rapport à l'ancienne pondération (section II.2), la nouvelle pondération inclut beaucoup plus de variables auxiliaires, et celles-ci sont plus fraîches.

<sup>6</sup> Dans Fidéli, le sexe n'est généralement pas renseigné pour les mineurs. Il fait l'objet d'une imputation s'appuyant sur les données démographiques.

- **Calcul des marges**

Pour chaque région, les marges de calage sont calculées comme les totaux des variables auxiliaires (nombre de logements dans la région, etc.). 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. Deux exceptions à ces principes généraux ont été introduites :

- En raison du très faible nombre de petites communes en Hauts-de-Seine, Seine-Saint-Denis et Val-de-Marne, le calage des petites communes en Île-de-France ne fait pas intervenir de total de population départementale.
- Pour des raisons similaires, les régions Corse et Provence-Alpes-Côte d'Azur sont considérées comme un ensemble<sup>7</sup>.

- **Mise en œuvre du calage**

Un calage des petites communes recensées en 2017 est réalisé pour chaque région à l'aide de la macro SAS%Calmar [5]. Afin d'éviter une trop grande dispersion des poids calés, la méthode Logit est utilisée, 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], évitant ainsi d'avoir des unités trop influentes (les poids de sondage initiaux sont de 5). *In fine*, les logements et les individus ont les poids de la petite commune dans laquelle ils sont situés.

### **III.2.2 Logements ordinaires des 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.

Le poids calé d'une commune est calculé ainsi : 
$$\frac{\text{total d'habitants en PC du DOM}}{\text{nombre d'habitants en PC du GR}}$$

Il est donc identique pour toutes les petites communes d'un même DOM.

De la même façon qu'en métropole, il est possible de calculer la variable auxiliaire du nombre d'habitants 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.

### **III.2.3 Logements ordinaires des grandes communes de métropole**

Dans la base de sondage des adresses (BSA) constituée à partir du répertoire d'immeubles localisées (RIL), deux catégories d'adresses co-existent : les adresses d'hôtels, campings et résidences hôtelières d'une part et les adresses d'habitation d'autre part. Le tirage aléatoire d'adresses de logements ordinaires à recenser chaque année (section I.1.2) ne concerne que les adresses d'habitation.

#### *III.2.3.1 Adresses d'hôtels, campings et résidences hôtelières*

Les adresses d'hôtels, campings et résidences hôtelières représentent 0,2 % des adresses de la BSA. Chaque année on recense l'ensemble de ces adresses du groupe de rotation. Différentes modifications ont été réalisées depuis 2004 sur ces types d'adresse. La rénovation de la pondération de l'EAR a été l'occasion pour celles-ci d'ajuster leur pondération au mode de recensement actuel de ces adresses.

<sup>7</sup> C'est donc l'ensemble des petites communes de Corse et de Provence-Alpes-Côte d'Azur qui constituent l'échantillon à caler sur les marges supra-régionales. Concernant le nombre total d'habitants, on calcule une marge pour chaque département faisant partie des deux régions.

Ainsi, pour chaque région :

- le poids des adresses d'hôtels est calculé ainsi  $\frac{\text{total d'adresses d'hôtels BSA}}{\text{nb adresses d'hôtels du GR}}$

- le poids des adresses de campings et résidences hôtelières est calculé ainsi  $\frac{\text{total logements d'adresses de campings et résidences hôtelières BSA}}{\text{nb logements d'adresses de campings et résidences hôtelières GR}}$

Les individus et logements situés à des adresses d'hôtels ont le même poids que leurs adresses. Ce poids est identique au sein de chaque région. Il en est de même pour les individus et logements situés à des adresses de campings et résidences hôtelières.

### *III.2.3.2 Adresses d'habitation*

Pour chaque grande commune, le tirage de l'échantillon d'adresses pour l'EAR est réalisé parmi les adresses de la BSA appartenant au groupe de rotation de l'année. Les variables auxiliaires nécessaires au calage doivent être donc être calculées au niveau de l'adresse. Il s'agit des variables suivantes : nombre de logements, nombre d'hommes, nombre de femmes, nombre de personnes selon cinq classes d'âge.

Plus précisément, pour le calage de l'EAR 2017, la méthode de repondération consiste à suivre les étapes suivantes :

- **Calcul des variables auxiliaires**

#### ***Nombre de logements***

La BSA N dans laquelle est tiré l'échantillon recensé à l'EAR N est constituée à partir du RIL établi en juillet N-1. Cette base exhaustive fournit ainsi pour chaque adresse de chaque grande commune le nombre de logements. Celui-ci peut être corrigé lors de la collecte de janvier-février de l'année N pour les adresses recensées cette année. La variable auxiliaire utilisée pour le calage de l'EAR N est le nombre de logements de la BSA corrigé de la collecte.

#### ***Variables auxiliaires relatives aux individus***

Les variables auxiliaires relatives aux individus (nombre de personnes par sexe et par classes d'âge) sont recherchées prioritairement dans Fidéli. Cependant, il n'existe pas d'identifiant commun entre Fidéli et l'échantillon d'adresses du recensement. Un appariement est donc réalisé entre Fidéli et la BSA sur la base d'éléments d'adressage et du plan cadastral (encadré 1).

### Encadré 1 : appariement de la BSA de France métropolitaine avec Fidéli

Dans 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.

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 appariées. 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 rivioli), numéro dans la voie, suffixe (bis, ter...), référence cadastrale.

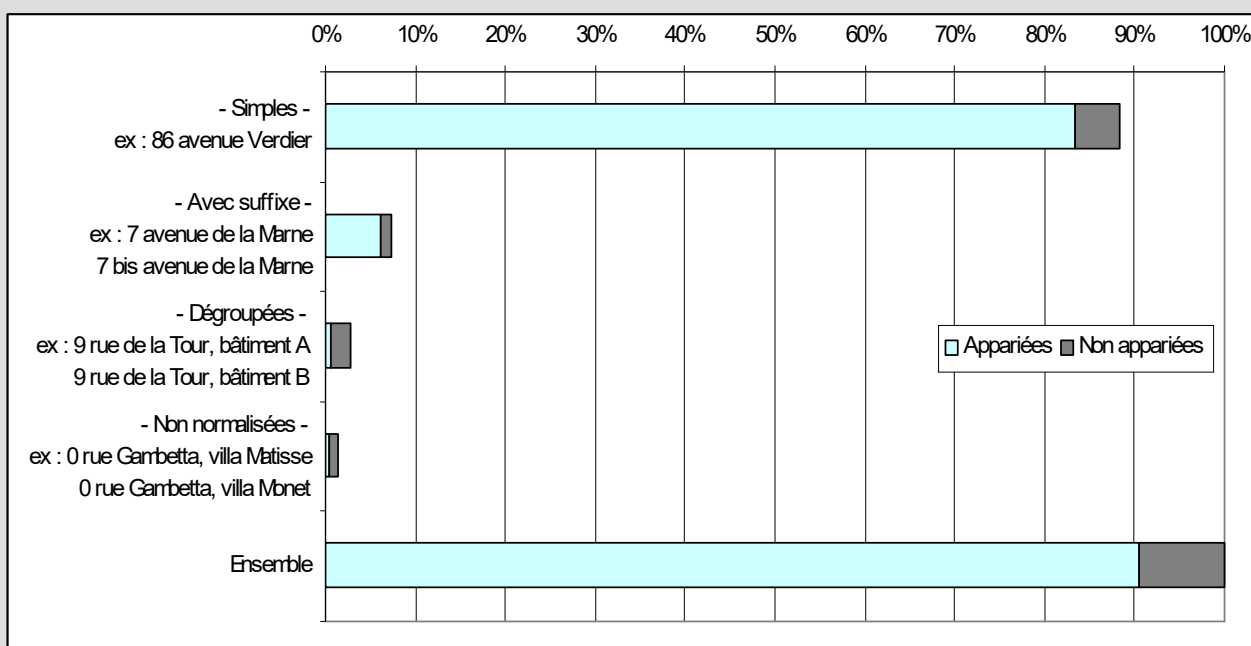
Des difficultés d'appariement entre la BSA et Fidéli sont rencontrées pour chaque catégorie d'adresses.

- 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 Fidéli, 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.

Les taux d'appariement sont les plus élevés pour les adresses simples, qui constituent la majorité des adresses de la BSA (figure A). En Corse cependant, la moitié des adresses sont non normalisées : le taux d'appariement global est de ce fait très faible par rapport aux autres régions, autour de 22 %.

Figure A : taux d'appariement selon les catégories d'adresses de la BSA 2017 à Fidéli 2016



Champ : adresses « appariables » de la BSA (adresses d'habitation rivolisées, et sans hôtel, résidence hôtelières ou camping au numéro), soit 99 % des adresses de la BSA.

Lecture : 88 % des adresses de la BSA sont dites « simples » dont 83 % appariées et 5 % non appariées à Fidéli.

Source : BSA 2017 et Fidéli 2016.

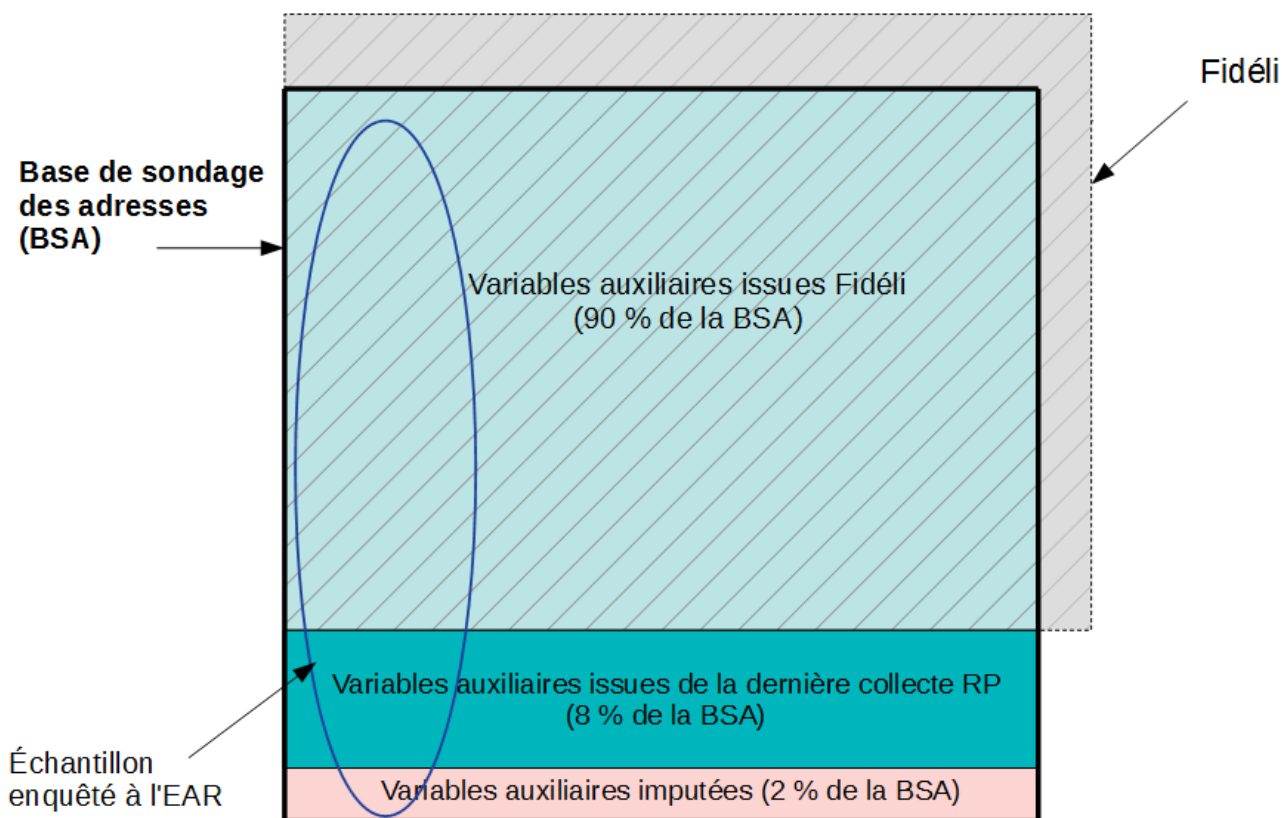
90 % des adresses trouvent un écho dans Fidéli. Les variables auxiliaires relatives aux individus sont donc calculées de manière « composite » (figure 7) :

- Pour la fraction (90 %) de la BSA 2017 appariée à Fidéli 2016, les variables auxiliaires sont calculées pour chaque adresse à partir de **Fidéli**.
- Pour la fraction qui n'a pas été appariée à Fidéli, les variables auxiliaires sont calculées à partir de la **dernière collecte de recensement connue** (8 %) (en excluant l'EAR que l'on cherche à pondérer) si l'adresse a déjà été recensée
- Pour le reste des adresses (ni appariées avec Fidéli, ni recensées lors d'une précédente EAR), les variables auxiliaires relatives aux individus sont **imputées** (2 %) de la manière suivante :
  - dans chaque commune, on calcule pour chaque variable auxiliaire sa valeur moyenne par logement sur le champ des adresses appariées à Fidéli ou collectée lors d'une précédente EAR.

Exemple :  $nb\ hommes\ moyen\ par\ logement = \frac{nb\ hommes\ de\ la\ commune}{nb\ de\ logements\ de\ la\ commune}$

  - pour chaque adresse, la variable auxiliaire imputée est égale à cette valeur moyenne multipliée par le nombre de logements de l'adresse dans la BSA corrigé de la collecte :  
 $nb\ hommes\ imputé = nb\ hommes\ moyen \times nb\ logements\ de\ l'adresse$

Figure 7 : source des variables auxiliaires composites



- **Calcul des marges**

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

- **Mise en œuvre du calage**

A l'EAR 2017, on dénombre 939 grandes communes en France métropolitaine, soit 939 échantillons d'adresses à caler. À chaque adresse sont rattachées les variables auxiliaires composites issues de la BSA enrichie par les traitements décrits précédemment.

Les calages sont réalisés à l'aide de la macro SAS %Calmar en utilisant la méthode Logit de sorte que les rapports de poids soient compris dans l'intervalle [0,3 ; 2,5]. Les individus et logements situés à des adresses d'habitation ont le même poids que leurs adresses.

### III.2.4 Logements ordinaires des grandes communes des DOM

Pour chacune des grandes communes des DOM, on rappelle que l'échantillon de l'EAR suit un plan de sondage à deux degrés (cf. section I.1.2) :

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

Deux niveaux de calage sont possibles : îlot et adresse. Les variables auxiliaires souhaitées sont les suivantes : nombre de logements, nombre de logements en immeubles collectifs, nombre d'habitants, par sexe et par tranches d'âge.

Tout comme pour les grandes communes de France métropolitaine, un appariement avec Fidéli est réalisé au niveau de l'adresse pour rechercher ces variables auxiliaires. Cependant, seul le taux d'appariement à La Réunion permet d'envisager un calage des adresses (cf. II.2.4.1 *infra*).

Deux méthodes de calage ont ainsi été envisagées : un calage à l'îlot pour la Guadeloupe, la Martinique et la Guyane d'une part, un calage simultané [8] à l'îlot et à l'adresse pour La Réunion d'autre part. Des travaux d'évaluation ont néanmoins révélé que la méthode de calage simultané, sensiblement plus complexe que la méthode de calage à l'îlot, n'était pas associée à des gains en matière de stabilité des séries. C'est ainsi la méthode de calage à l'îlot qui est appliquée pour les quatre DOM (cf. II.2.4.2 *infra*). La méthode de calage simultané à l'îlot et à l'adresse, envisagée initialement pour La Réunion, est présentée pour mémoire en annexe de ce document.

#### III.2.4.1 Appariement de la BSA des DOM avec Fidéli

Une minorité d'adresses de la BSA est appariée avec Fidéli en Guyane, la situation étant encore plus dégradée en Guadeloupe et en Martinique (Tableau 4). 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 (Tableau 4). 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.

Tableau 4 : résultat de l'appariement de la BSA avec Fidéli

DOM	% 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
Guadeloupe	8%	62%	51%
Martinique	13%	56%	42%
Guyane	32%	43%	65%
La Réunion	68%	5%	32%

Source : BSA et Fidéli

#### II.2.4.2 Calage au premier degré de sondage

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. La méthode proposée repose de ce fait sur un calage des groupes de rotation d'îlots : il est donc nécessaire de calculer pour tous les îlots de chaque grande commune des DOM les variables auxiliaires usuelles à partir de Fidéli.

- **Calcul des variables auxiliaires**

À partir des fichiers de Fidéli enrichis du code îlot<sup>8</sup>, on calcule pour tous les îlots des grandes communes des DOM les variables auxiliaires.

- **Calcul des marges**

Pour chaque commune des DOM, on calcule les marges de calage comme les totaux des variables auxiliaires.

- **Mise en œuvre du calage**

On doit réaliser le calage des îlots du groupe de rotation associé à l'EAR pour chaque grande commune.

Le calage de l'échantillon d'îlots est mis en œuvre à l'aide de la macro SAS %Calmar. Afin d'éviter une trop grande dispersion des poids calés, la méthode Logit est utilisée, qui permet un contrôle des déformations maximales des poids. 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] (le poids de sondage de chaque îlot dans son groupe de rotation est en effet égal à 5). Au niveau des logements et des individus recensés, les poids de sondage conditionnels de 2<sup>d</sup> degré valant 2,5, les poids finaux sont alors compris dans l'intervalle [2,5 ; 37,5]<sup>9</sup>.

#### III.2.5 Communautés

- **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 des caractéristiques structurelles qui sont contenues dans le RIL à partir duquel est constitué le répertoire des communautés. La variable essentielle est la catégorie de la communauté, qui permet de distinguer les établissements d'hébergement pour personnes âgées dépendantes (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 ;
- d'autre part les populations légales du RP N-3.

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 administrative ;
- populations pour certains croisements catégorie x région administrative<sup>10</sup>.

<sup>8</sup> Ce code îlot est obtenu en projetant les données géolocalisées de Fidéli dans les contours d'îlots.

<sup>9</sup> Avant 2019, le taux de sondage des adresses au 2<sup>d</sup> degré de tirage est toujours égal 0,40 (en l'absence de strate exhaustive de grandes adresses ou d'adresses nouvelles). À partir de 2019, les petites adresses connues peuvent avoir une probabilité d'inclusion inférieure à 0,40 et donc théoriquement un poids calé supérieur à 37,5 (à condition que le calage des îlots conduise à une déformation maximale des poids au premier degré).

<sup>10</sup> Soit 60 variables de calage issues des croisements de 5 catégories de communautés : maisons de retraite/EHPAD, services de moyen ou de long séjour d'établissements publics ou privés de santé, communautés religieuses, gendarmeries, établissements hébergeant des élèves ou des étudiants (hors cités universitaires) et 12 régions de métropole (Corse exclue).



- **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.

- **Mise en œuvre du calage**

Le calage de l'échantillon de communautés et de logements dans l'enceinte d'une communauté est mis en œuvre à l'aide de la macro SAS %Calmar. Afin d'éviter une trop grande dispersion des poids calés, la méthode Logit est utilisée, qui permet un contrôle des déformations maximales des poids. 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] (chaque communauté recensée a en effet un poids de sondage égal à 5).

Les individus vivant en communautés ont un poids égal à celui de la communauté à laquelle ils appartiennent<sup>11</sup>. Les logements dans l'enceinte d'une communauté et les individus y habitant ont un poids égal à celui de la communauté dans laquelle ils sont situés.

### III.3 Résumé des méthodes de repondération

Le tableau 5 synthétise les principales caractéristiques des méthodes de repondération qui sont mises en œuvre à partir de l'EAR 2019 selon les champs.

Tableau 5 : **synthèse des méthodes de repondération**

Champ	Unité de calage	Marges	Nombre d'échantillons	Principale source auxiliaire
<b>Logements ordinaires et leurs habitants</b>				
<b>France métropolitaine</b>				
<b>Grandes communes</b>				
Adresses d'habitation	Adresse	Commune	≈ 1000	BSA (RIL) + Fidéli
Adresses d'hôtels	Adresse	Région	13	BSA (RIL)
Adresses de campings et résidences hôtelières	Adresse	Région	13	BSA (RIL)
<b>Petites communes</b>	Commune	Région	13	Fidéli
<b>DOM</b>				
<b>Grandes communes</b>	Adresse	Commune	≈ 50	BSA (enquête cartographique)
<b>Petites communes</b>	Commune	Région	4	RP
<b>Communautés et logements dans l'enceinte d'une communauté</b>	Communauté	France	1	RP + répertoire des communautés

<sup>11</sup> Les individus de 3 à 17 ans résidant en internat ont un poids égal à 0, ceux-ci ne faisant pas partie de la population municipale mais de la [population comptée à part](#).

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

Dans cette partie, l'apport de la nouvelle pondération est évalué sur les champs pour lesquels la nouvelle méthode diffère sensiblement de l'ancienne, à savoir les **logements ordinaires et leurs résidents de France métropolitaine**. Pour mémoire, les individus en ménages ordinaires de métropole représentent 94,5 % de la population française (tableau 1).

Sur ce champ, la nouvelle pondération mobilise les données de Fidéli (cf. sections III.2.1 et III.2.3). 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 n'ont pas été calculées pour les grandes communes de France métropolitaine. Les analyses disponibles sont donc réalisées à partir des données repondérées de l'EAR 2014.

### IV.1 Distribution des nouvelles pondérations

Le tableau 6 présente la distribution des pondérations parmi les individus. Celle au sein des logements est quasiment identique, puisque les individus ont le poids de leur résidence principale.

Les poids sont logiquement plus dispersés avec la nouvelle méthode de pondération (tableau 5). Cela s'explique par l'introduction supplémentaire du nombre de variables auxiliaires. Par exemple, pour les logements ordinaires et leurs habitants de petites communes de métropole, 90 % des nouveaux poids s'étendent entre 3,3 et 6,8 contre 4,7 et 5,3 pour les anciens poids.

Tableau 6 : **distribution des pondérations**

	Pondération	Min	P5	Q1	Méd.	Q3	P95	Max
<b>Petites communes</b>	Ancienne	4,7	4,7	4,9	5,1	5,1	5,3	5,5
	Nouvelle	2,5	3,3	4,5	5,0	5,4	6,8	8,5
<b>Grandes communes</b>	Ancienne	1,0	5,0	5,4	14,5	16,1	18,9	30,4
	Nouvelle	1,5	4,7	5,7	14,3	16,3	20,2	62,5
<b>Ensemble</b>	Ancienne	1,0	4,7	5,0	5,1	5,3	16,4	30,4
	Nouvelle	1,5	3,5	4,7	5,2	6,5	17,0	62,5

Champ : individus en logement ordinaire de France métropolitaine, exceptés les bébés nés pendant ou après la date de référence de la collecte dont le poids est fixé à 0.

Source : EAR 2017.

### IV.2 Réduction de la variance

L'impact de l'utilisation de la nouvelle méthode peut être mesuré au travers du gain de précision des estimateurs. L'objectif est de comparer la variance des estimateurs selon la méthode d'estimation utilisée, à savoir celles correspondant aux anciennes et aux nouvelles pondérations respectivement. Cette comparaison s'appuie sur l'estimateur de variance d'un estimateur calé proposé dans [6] : pour une variable d'intérêt  $Y$  donnée,

$$\hat{V}(\hat{T}_{calé}(Y)) = \hat{V}(\hat{T}(\varepsilon))$$

où  $\hat{T}_{calé}(Y)$  est l'estimateur calé du total de  $Y$  et  $\hat{T}(\varepsilon)$  l'estimateur non-calé du total du résidu  $\varepsilon$  obtenu par la régression de  $Y$  sur les variables de calage et pondérée par le poids calé  $w$ .

En l'occurrence ici, l'objectif est d'évaluer le rapport :

$$\frac{\hat{V}(\hat{T}_{nouv}(Y))}{\hat{V}(\hat{T}_{anc}(Y))} = \frac{\hat{V}(\hat{T}(\varepsilon_{nouv}))}{\hat{V}(\hat{T}(\varepsilon_{anc}))}$$

où  $\hat{T}_{nouv}(Y)$  et  $\hat{T}_{anc}(Y)$  représentent l'estimateur du total de la variable  $Y$  respectivement avec la nouvelle et l'ancienne méthode d'estimation (correspondant à l'utilisation des nouvelles et des anciennes pondérations) et  $\varepsilon_{nouv}$  et  $\varepsilon_{anc}$  sont les résidus obtenus par la régression de  $Y$  sur les variables de calage correspondant respectivement à la nouvelle et à l'ancienne méthode d'estimation pondérée par le poids calé approprié.

La variance de l'estimateur du total de la variable  $\varepsilon$  peut s'exprimer sous la forme :

$$\hat{V}(\hat{T}(\varepsilon)) = N^2 \left(1 - \frac{n}{N}\right) \frac{s_\varepsilon^2}{n} \times Deff$$

où  $n$  est la taille de l'échantillon,  $N$  la taille de la population,  $s_\varepsilon^2$  la variance empirique dans l'échantillon de  $\varepsilon$  et  $Deff$  l'effet de sondage. L'effet de sondage est le rapport entre la variance de l'estimateur du total de  $\varepsilon$  avec le plan de sondage effectivement adopté et la variance de l'estimateur de  $\varepsilon$  si le plan de sondage était un sondage aléatoire simple.

Ici les investigations méthodologiques évaluées ne portent que sur la méthode d'estimation et pas sur le plan de sondage : autrement dit, dans la comparaison entre  $\hat{V}(\hat{T}(\varepsilon_{nouv}))$  et  $\hat{V}(\hat{T}(\varepsilon_{anc}))$ ,  $n$  et  $N$  sont constants. De plus, on formule l'hypothèse que  $Deff$  est lui-même constant, dans la mesure où on ne s'attend pas à ce que le plan de sondage affecte très différemment les résidus des variables d'intérêt correspondant à l'ancien et au nouveau calage respectivement. De ce fait :

$$\frac{\hat{V}(\hat{T}_{nouv}(Y))}{\hat{V}(\hat{T}_{anc}(Y))} = \frac{\hat{V}(\hat{T}(\varepsilon_{nouv}))}{\hat{V}(\hat{T}(\varepsilon_{anc}))} = \frac{s_{\varepsilon_{nouv}}^2}{s_{\varepsilon_{anc}}^2}$$

où  $s_{\varepsilon_{nouv}}^2$  et  $s_{\varepsilon_{anc}}^2$  sont la variance empirique dans l'échantillon respectivement de  $\varepsilon_{nouv}$  et  $\varepsilon_{anc}$ .

Cette simplification (et notamment l'hypothèse formulée sur  $Deff$ ) facilite grandement l'évaluation des gains de précision à attendre de la nouvelle méthode d'estimation, dans la mesure où elle ne suppose pas une estimation de variance complète (estimation rendue difficile par la complexité des plans de sondage en petites et grandes communes).

En définitive, on définit l'indicateur de gain de précision associé à la nouvelle méthode  $I$  comme le rapport des écart-types empiriques des résidus :

$$I = \sqrt{\frac{s_{\varepsilon_{anc}}^2}{s_{\varepsilon_{nouv}}^2}} = \sqrt{\frac{\hat{V}(\hat{T}_{anc}(Y))}{\hat{V}(\hat{T}_{nouv}(Y))}}$$

et on l'interprète comme le rapport entre l'écart-type de l'estimateur du total de  $Y$  avec la nouvelle méthode d'estimation et l'écart-type de l'estimateur du total de  $Y$  avec l'ancienne méthode d'estimation. Ainsi, pour une variable et un domaine donné,  $I=0,5$  signifie que la nouvelle méthode d'estimation conduit à un estimateur du total de  $Y$  dont l'écart-type est deux fois plus faible que celui obtenu avec l'ancienne méthode d'estimation : son intervalle de confiance est ainsi deux fois plus resserré. Plus l'indicateur  $I$  est faible, plus les gains associés à la nouvelle méthode d'estimation sont importants.

L'indicateur  $I$  est calculé séparément sur le champ des petites et des grandes communes. Comme on pouvait s'y attendre, le gain associé à la nouvelle méthode d'estimation 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 (tableau 7). Le gain de précision est plus réduit sur des populations rares pour lesquelles la corrélation avec les variables de calage est plus faible.

Tableau 7 : indicateurs de gain de précision des estimateurs au niveau national

Variabiles d'intérêt	Grandes communes	Petites communes
Population	0,52	0,28
Hommes	0,56	0,28
Femmes	0,56	0,32
Moins de 20 ans	0,50	0,19
20-39 ans	0,62	0,27
40-59 ans	0,55	0,21
60-74 ans	0,56	0,20
75 ans et plus	0,46	0,31
Sans diplôme	0,66	0,76
CAP BEP	0,74	0,73
BEPC	0,76	0,69
Bac	0,79	0,51
Supérieur au bac	0,78	0,65
Célibataires	0,55	0,34
En concubinage	0,89	0,64
Pacsés	0,92	0,54
Mariés	0,64	0,35
Divorcés	0,81	0,73
Veufs	0,62	0,52
En emploi	0,71	0,29
Chômeurs	0,78	0,78
Anciens actifs	0,50	0,26
Elèves	0,72	0,61
H/F au foyer	0,76	0,78
Autres inactifs	0,79	0,80

Champ : individus de logements ordinaires de France métropolitaine.

Lecture : sur le champ des grandes communes, l'écart-type de l'estimateur de la population avec la nouvelle méthode représente 0,52 de l'écart-type de l'estimateur de la population avec l'ancienne méthode ; sur le champ des petites communes, ce rapport est de 0,28.

Source : EAR 2017.

Au niveau régional, les gains de précision sont également notables (tableau 8) : ils sont toujours plus élevés pour les petites communes et varient selon les régions.

Tableau 8 : indicateurs de gain de précision des estimateurs de population au niveau régional

Régions	Grandes communes	Petites communes
Île-de-France	0,47	0,30
Centre-Val de Loire	0,52	0,24
Bourgogne-Franche-Comté	0,64	0,26
Normandie	0,62	0,22
Hauts-de-France	0,58	0,28
Grand Est	0,56	0,46
Pays de la Loire	0,45	0,12
Bretagne	0,47	0,17
Nouvelle-Aquitaine	0,54	0,23
Occitanie	0,64	0,28
Auvergne-Rhône-Alpes	0,54	0,34
Provence-Alpes-Côte d'Azur et Corse	0,55	0,36

Champ : individus de logements ordinaires.

Lecture : sur le champ des grandes communes de la région Pays de la Loire, l'écart-type de l'estimateur de la population avec la nouvelle méthode représente 0,45 de l'écart-type de l'estimateur de la population avec l'ancienne méthode ; sur le champ des petites communes de la région Pays de la Loire, ce rapport est de 0,12.

Source : EAR 2017.

Ces estimations tendent à montrer que la nouvelle méthode de repondération induit des gains de précisions sensibles, tout particulièrement sur le champ des petites communes. Cette évaluation repose sur un certain nombre d'hypothèses (en particulier la constante de l'effet de sondage) : en dépit de ces hypothèses, l'ampleur des gains en termes d'écart-type laisse peu de doute quant à la réalité de l'amélioration de la précision des estimateurs associés à la nouvelle méthode de calage sur marges.

### IV.3 Réduction de la volatilité empirique des estimations de population à partir des EAR

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. On attend de la réduction de la variance des estimateurs de chaque EAR une réduction de la volatilité des séries statistiques élaborées à partir de plusieurs EAR successives.

Puisque la nouvelle pondération concerne tous les champs à partir de l'EAR 2014, on peut analyser une réduction de volatilité sur la période 2014-2017 aux niveaux national, régional, départemental et communal pour les grandes communes.

Cette réduction est appréhendée par un indicateur calculé ainsi :

$X_{(t,p)}$  = grandeur (population, nombre d'individus en emploi...) observée l'année t pondérée par le poids p

$Evol(X_{(t,p)}) = \frac{X_{(t,p)} - X_{(t-1,p)}}{X_{(t-1,p)}}$  l'évolution de  $X_{(t,p)}$  entre les années t-1 et t calculée avec la pondération p

$E\bar{vol}(X_p) = \frac{1}{3} * \sum_{t=2015}^{2017} Evol(X_{(t,p)})$  = moyenne des évolutions annuelles de X sur la période 2015-2017 calculées

avec la pondération p

$Vol(X_p) = \frac{1}{3} \sum_{t=2015}^{2017} |Evol(X_{(t,p)}) - E\bar{vol}(X_p)|$

somme des écarts en valeur absolue des évolutions annuelles à la moyenne, calculés avec la pondération p. L'indicateur retenu s'écrit :  $I(X) = 100 \times [Vol(X_{pnouv}) - Vol(X_{panc})]$ , le facteur 100 permettant d'obtenir des valeurs plus lisibles<sup>12</sup>.

Un indicateur négatif signifie que la nouvelle pondération réduit la volatilité des évolutions, ce qui est l'un des objectifs poursuivis par sa mise en place. Un indicateur positif signifie, à l'inverse, que la nouvelle pondération accroît la volatilité des évolutions.

On peut également analyser la réduction de volatilité sur la période 2012-2017 pour le champ particulier des petites communes, puisque la nouvelle pondération a été calculée pour les EAR 2012 et 2013. Toutefois, il ne faut pas surinterpréter l'indicateur qui est présenté dans la mesure où il repose sur très peu d'observations (quelques années seulement). Par ailleurs, il se peut que la volatilité d'une série reflète une évolution réelle.

*Au niveau national*, les EAR étaient déjà peu volatiles ; la nouvelle pondération n'apporte pas d'amélioration de ce point de vue, l'indicateur étant très proche de 0 ( $I=0,03$ ) (figure 8a). On observe toutefois un gain sur le champ particulier des ménages ordinaires résidant en petites communes<sup>13</sup> ( $I=-0,16$ ), estimations particulièrement influencées par le déséquilibre des groupes de rotation (figure 8b).

12 Une variante de cet indicateur a également été testée (écart quadratique à la moyenne géométrique des évolutions sur la période 2014-2017) et conduit qualitativement aux mêmes résultats.

13 Sur le champ des petites communes, l'indicateur peut être calculé sur la période 2012-2017.

Figure 8 : évolution de la population en France métropolitaine estimée par les EAR

Figure 8a : ensemble

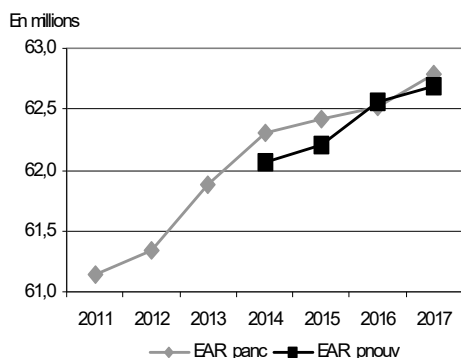
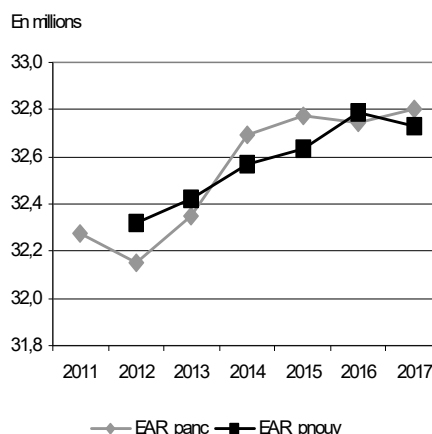


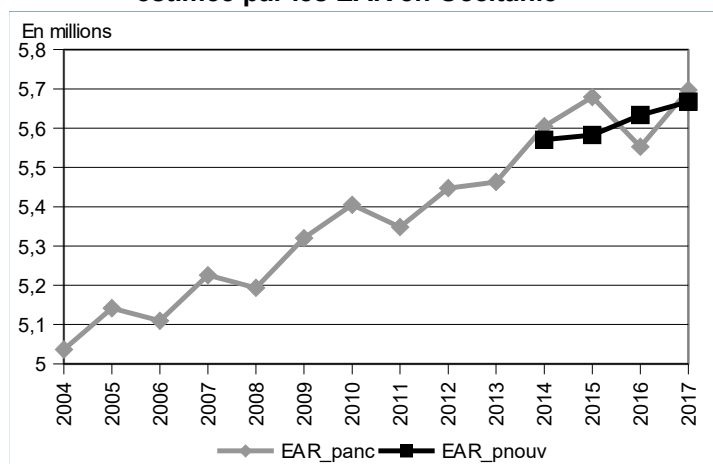
Figure 8b : petites communes



Note : les deux courbes représentent les estimations de populations avec les EAR obtenues avec l'ancienne pondération (courbes EAR\_panc) et la nouvelle pondération (courbe EAR\_pnov).  
 Champ : individus de logements ordinaires en France métropolitaine.  
 Source : EAR 2012 à 2017.

En revanche, *au niveau régional*, le recours aux nouvelles pondérations permet de réduire cette volatilité pour 10 des 13 régions de France métropolitaine, introduisant ainsi la possibilité de dégager des évolutions tendancielle de ces paramètres d'intérêt. Auparavant, la population de la région Occitanie avait un profil très heurté, et évoluant avec une saisonnalité de 5 ans. Avec la nouvelle pondération, son profil est lissé ( $I=-1,6$ ) et on dégage plus facilement une tendance dans la série de population de cette région (figure 9).

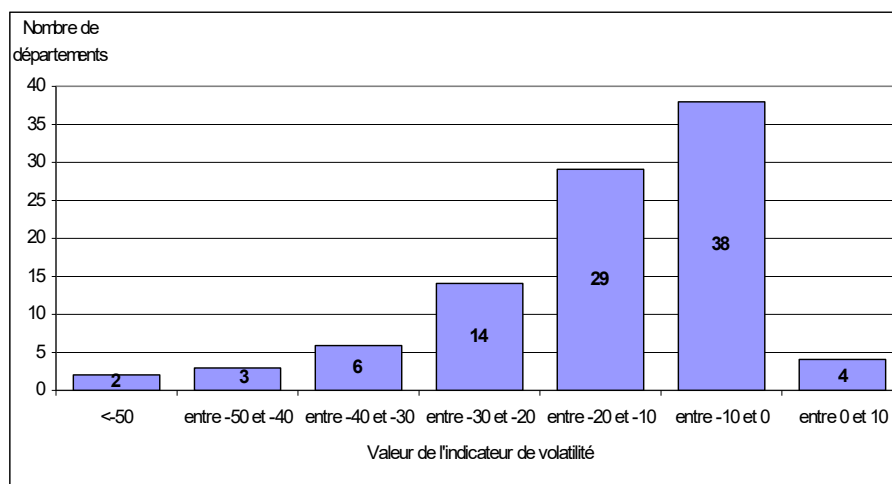
Figure 9 : évolution de la population des ménages ordinaires estimée par les EAR en Occitanie



Champ : individus résidant en ménages ordinaires en Occitanie.  
 Source : EAR 2004 à 2017.

*Au niveau départemental*, pour 92 départements de France métropolitaine, la volatilité de la population estimée par les EAR est réduite avec la nouvelle pondération entre 2014 et 2017. Pour les 4 autres départements (Indre-et-Loire, Maine-et-Loire, Seine-Saint-Denis et Yvelines), elle n'est que très faiblement accentuée (figure 10).

**Figure 10 : répartition des départements  
selon la réduction de la volatilité de leur population apportée par la nouvelle pondération**

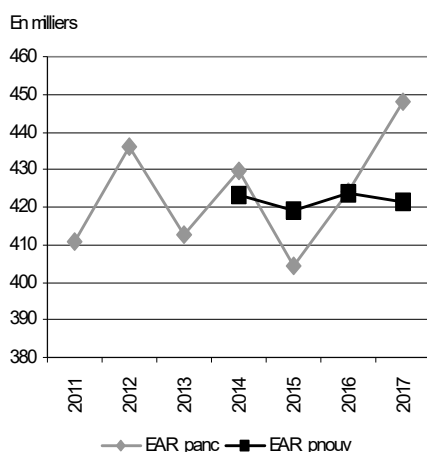


Lecture : 38 départements ont un indicateur de réduction volatilité compris entre -10 et 0.  
 Champ : 96 départements de France métropolitaine (individus de logements ordinaires).  
 Source : EAR 2014 à 2017.

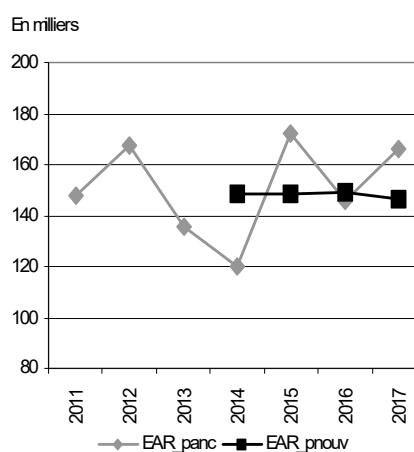
La figure 11 illustre cette réduction pour deux départements : l'Eure-et-Loir (28), dont l'indicateur de réduction de volatilité (-4,2) est proche de la valeur médiane sur la France métropolitaine (-4,0), et l'Ariège, dont l'indicateur de réduction de volatilité figure parmi les plus bas (-18,9).

**Figure 11 : évolution de la population des ménages ordinaires  
estimée par les EAR dans l'Eure-et-Loir et en Ariège**

**Figure 11a : département de l'Eure-et-Loir**



**Figure 11b : département de l'Ariège**



Champ : individus résidant en logements ordinaires.  
 Source : EAR 2011 à 2017.

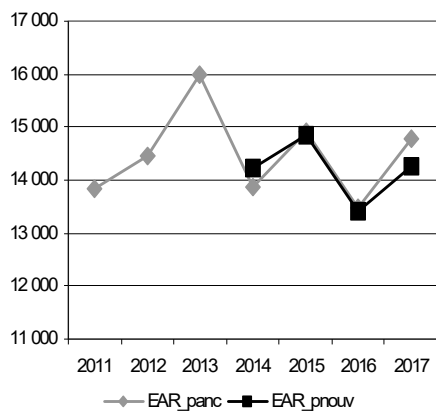
*Pour les grandes communes également, les nouvelles pondérations permettent de réduire la dispersion des estimations du nombre de logements et d'habitants au cours du temps. Ainsi, l'indicateur de volatilité de la population est négatif pour 78 % des grandes communes de France métropolitaine n'ayant pas subi de modifications de contours géographiques sur la période 2014-2017.*

Les figures 12a et 12b illustrent cette réduction pour deux communes : Crépy-en-Valois (60176), dont l'indicateur de réduction de volatilité (-1,5) est égal à la valeur médiane sur la France métropolitaine

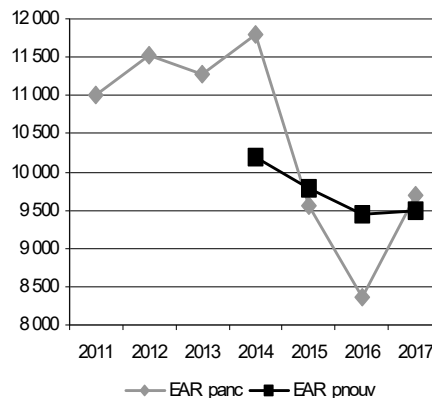
(réduction légère), et le Pradet (83098), dont l'indicateur de réduction de volatilité figure parmi les plus bas (-12,1, réduction très forte).

**Figure 12: évolution de la population des ménages ordinaires estimée par les EAR dans deux grandes communes**

**Figure 12a : commune de Crépy-en-Valois**



**Figure 12b : commune du Pradet**



Champ : individus résidant en logements ordinaires.  
Source : EAR 2004 à 2017.



## V. Comparabilité des données de l'EAR et du RP

L'objectif de cette partie est de confronter les estimations issues d'une EAR (nombre de logements, d'individus...) à celles fournies par le RP du même millésime. Les comparaisons portent sur les millésimes 2014 et 2015 : d'une part, les nouvelles pondérations n'ont été produites pour tous les champs (y compris logements ordinaires en grandes communes) qu'à partir de 2014 ; d'autre part, un changement de questionnaire intervenu à partir de l'EAR 2018 affecte la série de population à partir du RP 2016.

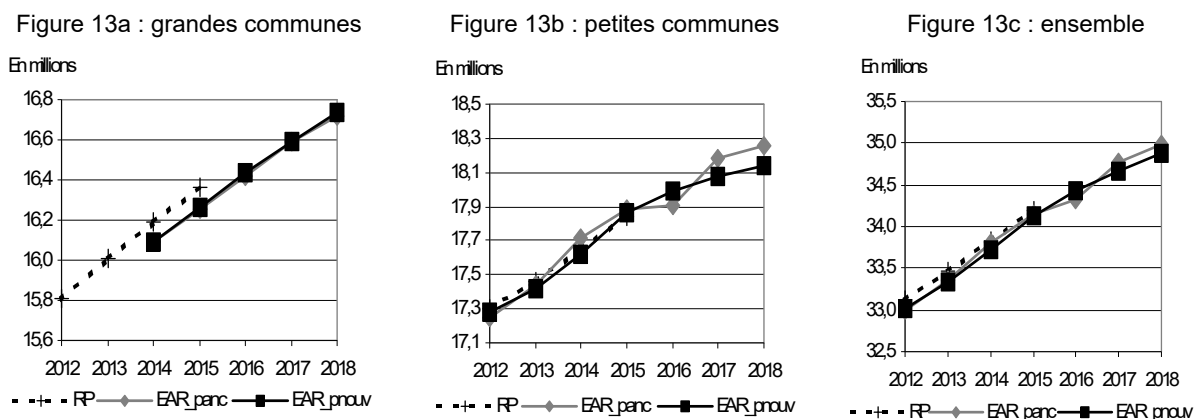
### V.1 Des données qui n'ont pas la même temporalité

#### V.1.1 Comparaison des données EAR et RP en volume

En grandes communes de France métropolitaine, l'EAR sous-estime le nombre de logements par rapport au RP (figure 13a). **Cette sous-estimation est liée au décalage temporel sur le nombre de logements utilisé comme variable auxiliaire dans les deux sources.** En effet, le nombre de logements utilisé pour caler l'EAR N est celui du RIL de l'année N-1 (juillet N-1) corrigé de la collecte N pour la fraction échantillonnée l'année N. Dans le RP N, le nombre de logements utilisé pour caler le RP est la moyenne entre le nombre de logements du RIL N-1 (juillet N-1) et le RIL N (juillet N) corrigé de la collecte N pour la fraction échantillonnée l'année N. En petites communes en revanche, il n'y a priori pas ce décalage de temporalité entre les données de l'EAR et du RP.

En raison de la sous-estimation par l'EAR du nombre de logements en grandes communes, les nouvelles pondérations peuvent difficilement conduire à une réduction des écarts entre l'EAR et le RP au niveau national. En revanche, grâce à un gain en précision (cf. IV.1), elles permettent de diminuer la volatilité des écarts entre les estimations produites par les EAR et celles fournies par le RP (figure 13c).

Figure 13 : estimations du nombre de logements en France métropolitaine selon l'EAR et le RP



Champ : logements ordinaires de France métropolitaine.  
Source : EAR 2012 à 2018 et RP 2012 à 2015.

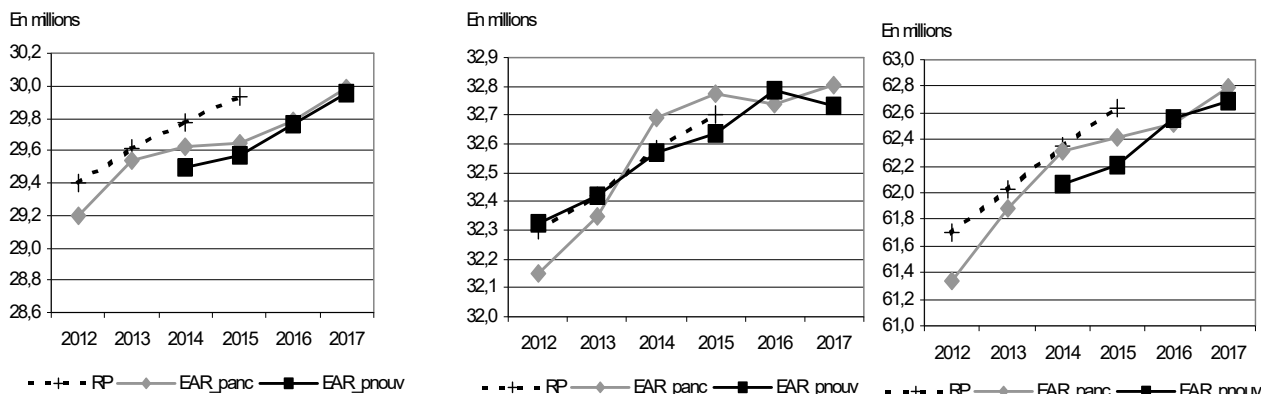
En grandes communes, l'EAR sous-estime la population par rapport au RP, en raison de la sous-estimation du nombre de logements (figure 14a). Cette sous-estimation se répercute au niveau national tous types de communes confondus (figure 14c).

Figure14 : estimations de la population en France métropolitaine selon l'EAR et le RP

Figure 14a : grandes communes

Figure 14b : petites communes

Figure 14c : ensemble



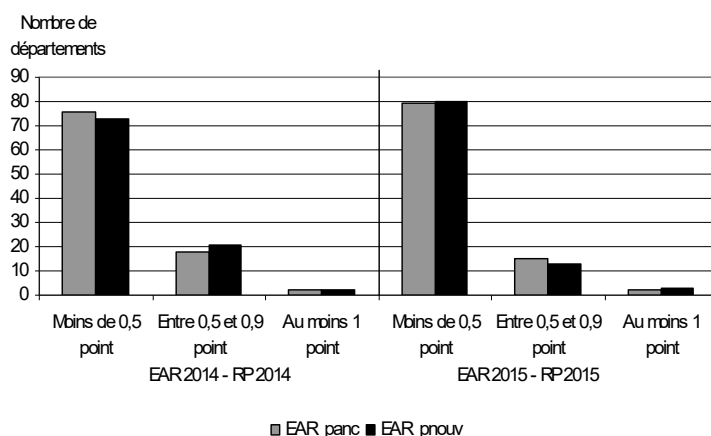
Champ : individus résidant en logements ordinaires de France métropolitaine.  
 Source : EAR 2012 à 2017 et RP 2012 à 2015.

### V.1.2 Comparaison des données EAR et RP en taux

Contrairement aux niveaux, les estimations de proportions issues de l'EAR sont proches des estimations du RP. C'est d'ailleurs pour cette raison que les études menées sur l'EAR 2015 et l'EAR 2016 qui se basaient encore sur les anciennes pondérations ne présentent que des proportions ou des taux à des niveaux infranationaux (cf. [3] et [4]).

La nouvelle pondération n'apporte donc pas d'amélioration sur l'estimation de taux. Au niveau national, quelle que soit la pondération utilisée, l'EAR 2015 estime par exemple 17,4 % d'actifs occupés à temps partiel, proportion identique à celle estimée par le RP 2015. Pour la quasi-totalité des départements, quelle que soit la pondération utilisée, 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 1 point au maximum en 2014 et en 2015 (figure 15).

Figure 15 : répartition des départements selon l'écart (en valeur absolue) entre le pourcentage des actifs occupés travaillant à temps partiel estimé par l'EAR et celui estimé par le RP

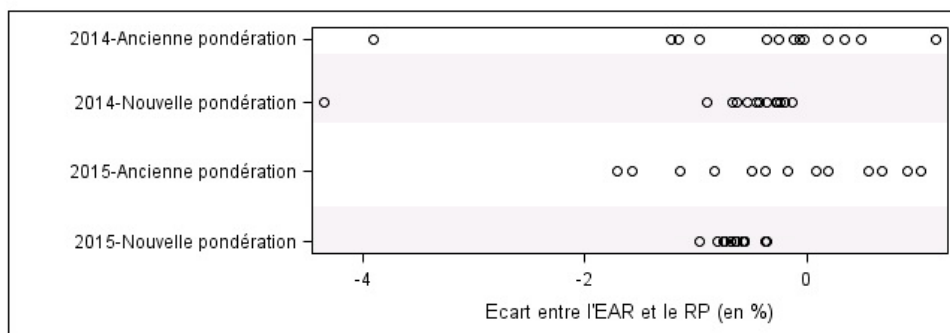


Note : écart en points de pourcentage (en valeur absolue) entre le pourcentage d'actifs occupés à temps partiel estimé par l'EAR et celui estimé par le RP du même millésime, selon la pondération de l'EAR utilisée.  
 Champ : 96 départements de France métropolitaine (individus de logements ordinaires).  
 Source : EAR 2014 et 2015, RP 2014 et 2015.

## V.2 Comparaison des données EAR et RP aux niveaux régional et départemental

En 2014, les écarts sur la population estimée entre EAR et RP au sein de chaque région sont beaucoup plus proches d'une région à l'autre avec la nouvelle pondération (figure 16). On retrouve au niveau régional le fait que l'EAR sous-estime la population par rapport au RP (les écarts avec la nouvelle pondération ne sont pas centrés autour de 0).

Figure 16 : écart entre la population estimée par l'EAR et celle estimée par le RP selon les régions



Note : chaque cercle représente une région.

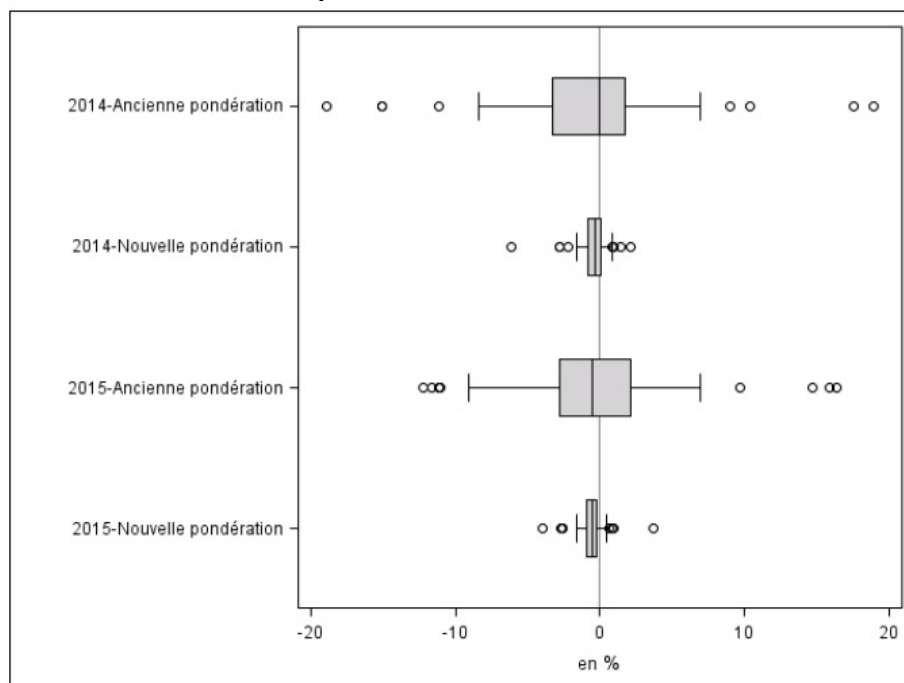
Lecture : en 2015, les écarts entre l'EAR ancienne pondération et le RP au niveau régional s'étendent de -1,7 % et +1,0 %. Avec la nouvelle pondération, ils s'étendent de -1,0 % à -0,4 %

Champ : 13 régions de France métropolitaine (individus de logements ordinaires).

Source : EAR 2014 et 2015, RP 2014 et 2015.

La distribution complète des écarts de population entre EAR et RP par département est plus resserrée avec la nouvelle pondération (figure 17).

Figure 17 : distribution des écarts entre EAR et RP sur la population départementale selon la pondération de l'EAR et l'année considérée



Note : les « moustaches » s'étendent du 5<sup>e</sup> percentile au 95<sup>e</sup> percentile. La ligne divisant les « boîtes » représentent la valeur médiane de l'écart. Les bords des boîtes sont positionnés aux premiers et troisièmes quartiles des distributions.

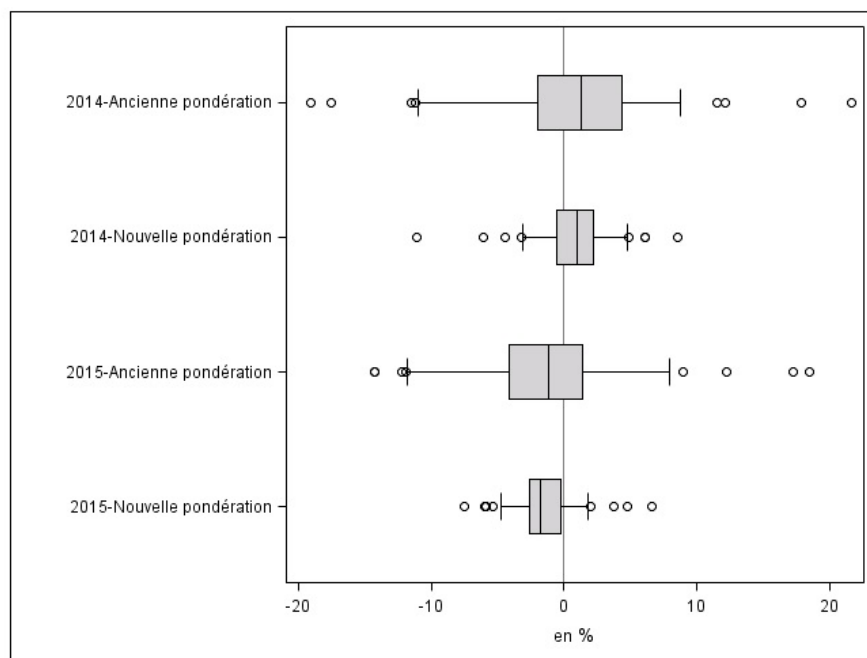
Lecture : en 2014, les écarts entre l'EAR ancienne pondération et le RP au niveau départemental s'étendent de -19 % et +19 %. Avec la nouvelle pondération, ils varient de -7 % à +4 %.

Champ : 96 départements de France métropolitaine (individus de logements ordinaires).

Source : EAR 2014 et 2015, RP 2014 et 2015.

Ce « resserrement » peut également être observé pour l'estimation du total de populations plus spécifiques comme le nombre d'actifs occupés à temps partiel par département. Désormais, pour la majorité des estimations départementales, l'écart avec le RP n'excède pas 2 % en valeur absolue (figure 18).

Figure 18 : **distribution des écarts entre EAR et RP sur le nombre d'actifs occupés à temps partiel par département selon la pondération de l'EAR et l'année considérée**



Note : les « moustaches » s'étendent du 5<sup>e</sup> percentile au 95<sup>e</sup> percentile. La ligne divisant les « boîtes » représentent la valeur médiane de l'écart. Les bords des boîtes sont positionnés aux premiers et troisièmes quartiles des distributions.

Lecture : en 2014, les écarts entre l'EAR ancienne pondération et le RP au niveau départemental s'étendent de -19 % à +22 %. Avec la nouvelle pondération, ils varient de -11 % à +9 %.

Champ : 96 départements de France métropolitaine (individus de logements ordinaires).

Source : EAR 2014 et 2015, RP 2014 et 2015.

## V.4 Comparaison des données EAR et RP en évolution

Si l'on ne s'attend pas à un rapprochement des niveaux entre EAR et RP, on espère en revanche un rapprochement de l'estimation EAR des évolutions entre deux dates à l'estimation RP.

Les données disponibles permettent de comparer l'évolution de la population entre 2014 et 2015 estimée par les EAR d'une part à celle estimée par le RP d'autre part.

Au niveau national, l'évolution de la population entre 2014 et 2015 estimée par les EAR avec la nouvelle pondération se rapproche légèrement de celle estimée par le RP (figure 19a). Sur le champ des petites communes de France métropolitaine, on dispose de trois évolutions disponibles pour l'analyse 2012-2013, 2013-2014 et 2014-2015 : avec les EAR, elles se rapprochent davantage de celles du RP lorsque la nouvelle pondération est utilisée (figure 19b).

**Figures 19 : estimations de l'évolution de la population en France métropolitaine entre 2014 et 2015 selon les EAR et le RP**

Figure 19a : ensemble

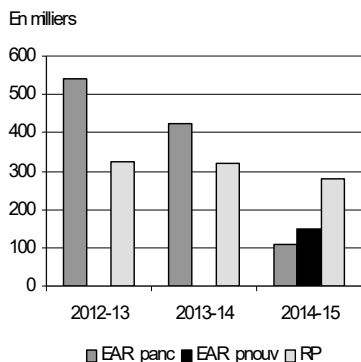


Figure 19b : petites communes

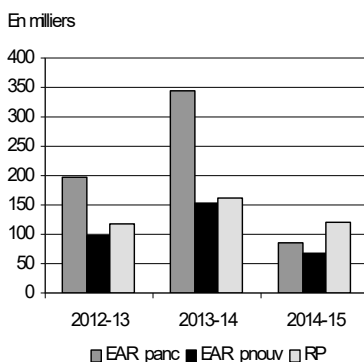
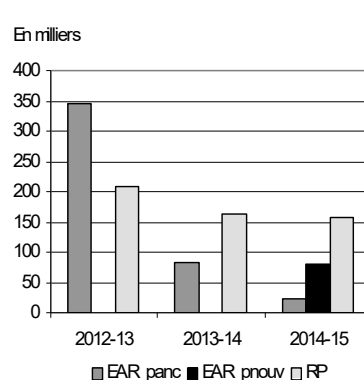


Figure 19c : grandes communes



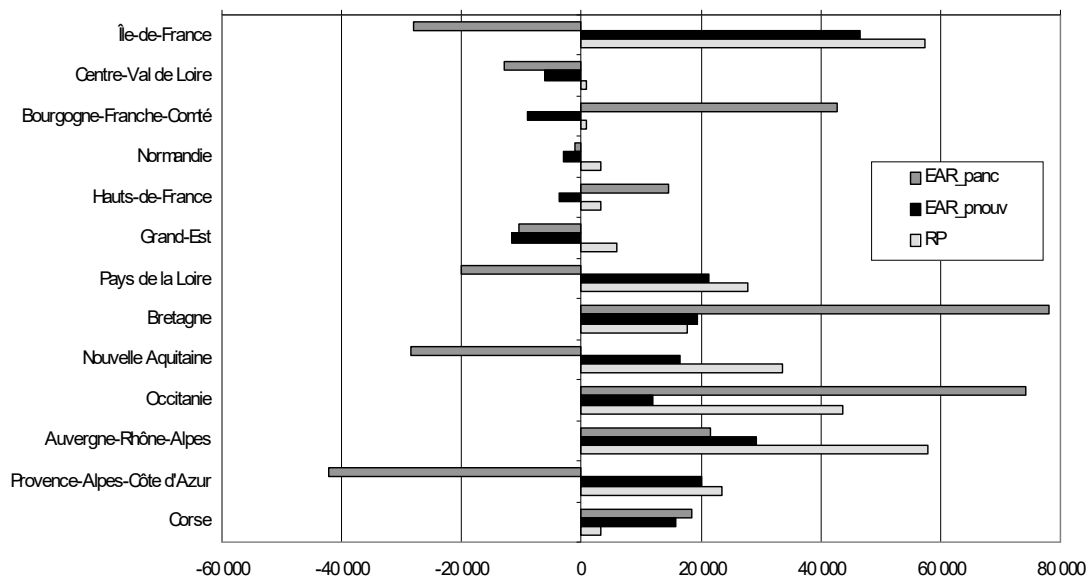
Lecture : entre 2014 et 2015, la population croît de 279 000 habitants d'après le RP, 147 000 d'après l'EAR avec la nouvelle pondération, 106 000 d'après l'EAR avec l'ancienne pondération.

Champ : individus de ménages ordinaires de France métropolitaine.

Source : EAR 2012 à 2015 et RP 2012 à 2015.

Les résultats appréhendés via l'évolution 2014-2015 selon les régions sont également rassurants : pour la majorité des régions, les écarts avec l'évolution estimée par le RP sont réduits avec la nouvelle pondération (figure 20). Quand ils s'éloignent, l'écart avec le RP ne se creuse que très faiblement.

Figure 20 : évolution de la population entre 2014 et 2015 selon les régions de France métropolitaine



Champ : individus de logements ordinaires de France métropolitaine.

Source : EAR 2014 et 2015, RP 2014 et 2015.

## Conclusion

L'exploitation d'une seule EAR a permis d'obtenir plus rapidement les résultats issus d'une question modifiée à l'EAR 2015 et 2016 [3] [4]. Seuls des taux ont toutefois été publiés au niveau régional et départemental, les indicateurs en volumes étaient trop instables. Il est donc crucial d'améliorer la précision des résultats estimés à partir d'une seule EAR.

La méthode proposée dans ce document repose sur une utilisation accrue des données administratives. Elle s'adapte à chaque champ de l'enquête : petites communes/grandes communes, métropole/DOM, population en logement ordinaire/en communauté/HMSA, etc.

Deux méthodes de validation ont été mises en œuvre pour montrer que la nouvelle méthode conduit à des gains de précision. D'une part, la volatilité empirique des variables d'intérêt décroît pour la quasi-totalité des départements et des régions. D'autre part, la variance des estimateurs diminue de manière significative. Ces améliorations de précision concernent en premier lieu les estimations sur le champ des petites communes, et sur les variables les plus corrélées au variable de calage : population par sexe et âge.

Cette nouvelle méthode de pondération a été mise en œuvre sur les EAR à partir de l'année 2019. La mobilisation des EAR pour les études statistiques sera ainsi facilitée, auprès de différents publics : services statistiques ministériels et chercheurs notamment. Les données de l'EAR avec ces nouvelles pondérations seront mises à disposition par le biais du Centre accès sécurisé distant (CASD) en 2020.

Les résultats issus d'une seule EAR resteront moins robustes que ceux établis à partir du recensement de la population, et ne pourront pas être produits sur des mailles territoriales trop petites. L'Insee va réaliser en 2020 des estimations de variance des estimateurs à partir d'une seule EAR, qui n'ont pas pu être produites dans le cadre de ce document de travail. Il faut par ailleurs bien garder à l'esprit qu'il existe une différence de concept entre les données des EAR et celles du RP (cf. partie 5 de ce document) : le nombre de logements de l'EAR étant calé sur le RIL au 1<sup>er</sup> janvier précédant l'année de collecte en grande commune, ce qui conduit à une sous-estimation structurelle du nombre de logements.

La nouvelle méthode de pondération pour la population des logements ordinaires en grande commune repose sur un appariement entre la base de sondage d'adresses et le fichier Fidéli. Cet appariement permet d'enrichir la base de sondage et de disposer d'information auxiliaire. Ces données pourraient être mobilisées à d'autres étapes du processus pour gagner en qualité (tirage de l'échantillon, amélioration des pondérations du recensement de la population, etc.) ou pour mener des travaux de simulation d'un changement de méthode.

## 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.
- [5] Sautory, O. (Insee), « La macro Calmar. Redressement d'un échantillon par calage sur marges », *Document de travail F9310 de la DSDS*, 1993.
- [6] Deville J.-C., Särndal C.-E., "Calibration Estimators in Survey Sampling", *Journal of the American Statistical Association*, Vol. 87, No. 418, pp.376-382
- [7] Deville J.-C., Tillé Y. (2004), "Efficient Balanced Sampling: The cube method", *Biometrika*, **91**, 4, pp. 893-912
- [8] Le Guennec J., Sautory O. (2002), « Calmar 2 : Une nouvelle version de la macro Calmar de redressement d'échantillon par calage », *Journées de méthodologie statistique*

## Annexe : méthode de calage simultané envisagée pour les grandes communes de La Réunion

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 (Tableau 4 dans le corps du document). 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. Une approche par calage simultané est susceptible de concilier ces deux approches.

L'évaluation des performances de cette méthode de repondération n'a pas permis de mettre en évidence de gains significatifs par rapport à la méthode de calage à l'îlot. C'est donc la méthode de calage à l'îlot qui a été finalement mise en œuvre à La Réunion, comme dans les autres DOM (cf. partie III.2.4.1 dans le corps du document).

### A.1 Description de la méthode

- 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
- 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 :

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

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.

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

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 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 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 de la BSA appariée à Fidéli.
- **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.
  - Le vecteur  $\bar{v}_m$  comprend les estimations pondérées du nombre total de logements issu de la BSA 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 non appariée à Fidéli.

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

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

### A.3 Calcul des variables auxiliaires du 1<sup>er</sup> degré de tirage

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 de Fidéli enrichis du code îlot, on calcule pour tous les îlots des grandes communes de la Réunion les variables auxiliaires (vecteurs  $X_m$ ).
- À partir de l'échantillon d'adresses de l'EAR :
  - On calcule les estimations des variables auxiliaires 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 2017 (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 de la BSA appariée à Fidéli.
  - 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.

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

## Série des documents de travail de la DSDS

- F2020-02** : Difficultés de logement ; les situations de logement appréciées selon les dimensions proposées par le Conseil national de l'information statistique - Catherine Rougerie
- F2020-01** : Effets des réformes 2018 de la fiscalité du capital des ménages sur les inégalités de niveau de vie en France : une évaluation par micro-simulation - Félix Paquier et Michaël Sicsic
- F1908** : Simulation des effets redistributifs de la transformation de l'ISF en IFI à l'aide du modèle Ines - Félix Paquier, Kevin Schmitt et Michaël Sicsic
- F1907** : Le projet Muse : 5 ans d'expérimentations pour préparer l'introduction d'Internet dans l'enquête Emploi - Marguerite Garnero
- F1906** : Enseignements sur les estimations d'emploi de l'appariement entre l'enquête Emploi et les bases administratives en 2012 – Claude Picart
- F1905** : Le revenu des ménages selon la comptabilité nationale et selon l'Enquête Revenus Fiscaux et Sociaux (ERFS) - Jérôme Accardo
- F1904** : Appariement entre l'enquête Emploi et le fichier Historique de Pôle emploi sur la période 2012-2017 - Méthode et premiers résultats - Alexis Hameau, Sylvain Larrieu, Anis Marrakchi, Alexis Montaut, Christophe Dixte, Yohan Coder et Sophie Hamman
- F1903-G2019/04** : L'économie numérique fausse-t-elle le partage volume-prix du PIB ?  
Lorraine Aeberhardt, Florian Hatier, Marie Leclair, Benoît Pentinat et Jean-Denis Zafar
- F1902** : Les modèles de contrôle des effets de structure : comparaisons et application à une analyse de la disparité départementale des revenus non salariés moyens - Christophe Bertran et Laurianne Salembier
- F1901 - G2019/01** : Effets de moyen terme d'une hausse de TVA sur le niveau de vie et les inégalités : une approche par microsimulation - Mathias André et Anne-Lise Biotteau
- F1808** : La construction d'un indice des prix des logements neufs - Thomas Balcone
- F1807** : Estimer les effectifs de couples de personnes de même sexe au recensement : expérimentation d'une solution de validation du sexe par le prénom - Élisabeth Algava et Sébastien Hallépée
- F1806** : L'effet d'une variation du montant de certains transferts du système socio-fiscal sur le niveau de vie : résultats sur 2016 à partir du modèle de microsimulation Ines (Cahier de variantes) - Maëlle Fontaine et Michaël Sicsic
- F1805** : Family, Firms and the Gender Wage Gap in France - Elise Coudin, Sophie Maillard et Maxime Tô
- F1804** : Mode de collecte et questionnaire, quels impacts sur les indicateurs européens de l'enquête Emploi ? - Klara Vinceneux  
How do the collection mode and questionnaire used affect the European indicators in the French Labour Force Survey ? - Klara Vinceneux
- F1803** : Trois versions du taux d'effort en matière de logement - Pascal Godefroy
- F1802** : Heterogeneous exposure to labor earnings risk - Pierre Pora et Lionel Wilner

- F1801** : L'espérance de vie par niveau de vie Méthode et principaux résultats - Nathalie Blanpain
- F1708** : Les données fiscales de l'EDP : une nouvelle source d'informations sur les couples et les familles ? - Vianney Costemalle
- F1707** : La situation matrimoniale dans le recensement : impact de la refonte du questionnaire de 2015. - Guillemette Buisson
- F1706** : Indices de prix à la consommation - Patrick Sillard
- F1705** : Effet d'un choc d'inflation sur le revenu disponible et ses composantes deux ans après : une approche par microsimulation - Anne-Lise Biotteau et Maëlle Fontaine
- F1704** : Scanner data and quality adjustment - Isabelle Léonard, Patrick Sillard, Gaëtan Varlet et Jean-Paul Zoyem
- F1703** : Les structures familiales en France : comparaison entre le recensement, l'enquête famille et logements et l'enquête emploi - Guillemette Buisson et Aude Lapinte
- F1702** : Projections de la population active à l'horizon 2070 - Malik Koubi et Anis Marrakchi
- F1701** : Les taux marginaux effectifs de prélèvement pour les personnes en emploi en France en 2014 - Juliette Fourcot et Michaël Sicsic
- F1606** : Projections de population 2013-2070 pour la France : méthode et principaux résultats - Nathalie Blanpain et Guillemette Buisson
- F1605** : Les durées passées en famille monoparentale - Méthode d'estimation des durées et résultats - Vianney Costemalle
- F1604** : ESeG = European Socio economic Groups - Nomenclature socio-économique européenne - Monique Meron, Michel Amar, Charline Babet, Milan Bouchet-Valat, Fanny Bugeja-Bloch, François Gleizes, Frédéric Lebaron, Cédric Huguée, Étienne Penissat et Alexis Spire
- F1603** : Catégorie sociale d'après les déclarations annuelles de données sociales et catégorie sociale d'après le recensement : quels effets sur les espérances de vie par catégorie sociale ? Comparaison entre les déclarations annuelles de données sociales et les recensements de la population. Comparaison de méthodes d'estimation des espérances de vie - Vianney Costemalle
- F1602** : L'espérance de vie par catégorie sociale et par diplôme - Méthode et principaux résultats - Nathalie Blanpain
- F1601** : Échantillonnage des agglomérations de l'IPC pour la base 2015 - Laurence Jaluzot et Patrick Sillard
- F1508** : Worker-firm matching and the family pay gap: Evidence from linked employer-employee data - Lionel Wilner
- F1507** : Effet des nouvelles mesures sociales et fiscales sur le niveau de vie des ménages : méthodologie de chiffrage avec le modèle de microsimulation Ines - Mathias André, Marie-Cécile Cazenave, Maëlle Fontaine, Juliette Fourcot et Antoine Sireyjol
- F1506** : Nowcasting du taux de pauvreté par la micro-simulation - Maëlle Fontaine et Juliette Fourcot

- F1505/376-501** : Bilan du projet EDP++ - division Camap et division Enquêtes et études démographiques
- F1504** : Contrôles des rémunérations dans les déclarations annuelles de données sociales (DADS) - Une analyse exploratoire pour améliorer la détection des points atypiques - Claire Jacod
- F1503** : Précision de l'enquête Patrimoine 2010 - Pierre Lamarche et Laurianne Salembier
- F1502** : Pourquoi l'indicateur de pauvreté en conditions de vie baisse malgré la crise économique ouverte en 2008 ? Jean-Louis Pan ké shon
- F1501** : Évolution de la population de la France entre 1981 et 2011 : contributions de la fécondité, de la mortalité, du solde migratoire et de la structure de la pyramide des âges - Catherine Beaumel et Pascale Breuil-Genier
- F1410** : "Personal network" and retirement: Is retirement bad for friendship and good for family relationships ? Anne Laferrère
- F1409** : Retraités mais pas en retrait : La retraite pousse-t-elle à de nouvelles activités ? Anne Laferrère
- F1407** : Production "aval" de l'enquête emploi en continu EEC2 2013 - 20XX - Fabien Guggemos
- F1406 bis** : La constitution de l'échantillon démographique permanent de 1968 à 2012 - Stéphane Jugnot
- F1405 (tome 1)** : Hommes et femmes vivant en couple en 2009, 1999 et aux recensements précédents - Fabienne Daguet
- F1405 (tome 2)** : Hommes et femmes vivant en couple en 2009, 1999 et aux recensements précédents - Fabienne Daguet
- F1404** : L'addition est-elle moins salée ? La réponse des prix à la baisse de TVA dans la restauration en France - Quentin Lafféter et Patrick Sillard
- F1403** : Estimer les flux d'entrées sur le territoire à partir des enquêtes annuelles de recensement - Chantal Brutel
- F1402** : Une rotation de la main d'œuvre presque quintuplée en 30 ans : plus qu'un essor des formes particulières d'emploi, un profond changement de leur usage - Claude Picart
- F1401** : Calculs statistiques de stock et de flux sur la révision électorale 2012-2013 - Christelle Rieg