

La repondération des enquêtes annuelles de recensement pour une diffusion complémentaire du RP

10e colloque francophone sur les
sondages





01

· Présentation de l'EAR



02

· Équilibrage imparfait des groupes de rotation



03

· Nouvelle méthode de repondération



04

· Impact sur la qualité des estimations



01

Présentation de l'EAR

1.1. L'EAR, une contribution au cumul de recensement

1.2. Une diffusion encore restreinte

1.3. Un produit perfectible

- ◆ **Moyen de répartir la collecte sur un cycle de 5 ans**
 - Petites communes (PC, < 10 000 habitants) réparties en 5 groupes, collectées exhaustivement
 - Unité d'échantillonnage = la commune
 - Grandes communes (GC, \geq 10 000 habitants) interrogées par échantillon d'adresses tous les ans
 - Unité d'échantillonnage = l'adresse
- ◆ Un millésime de résultat du RP formé par 5 EAR consécutives

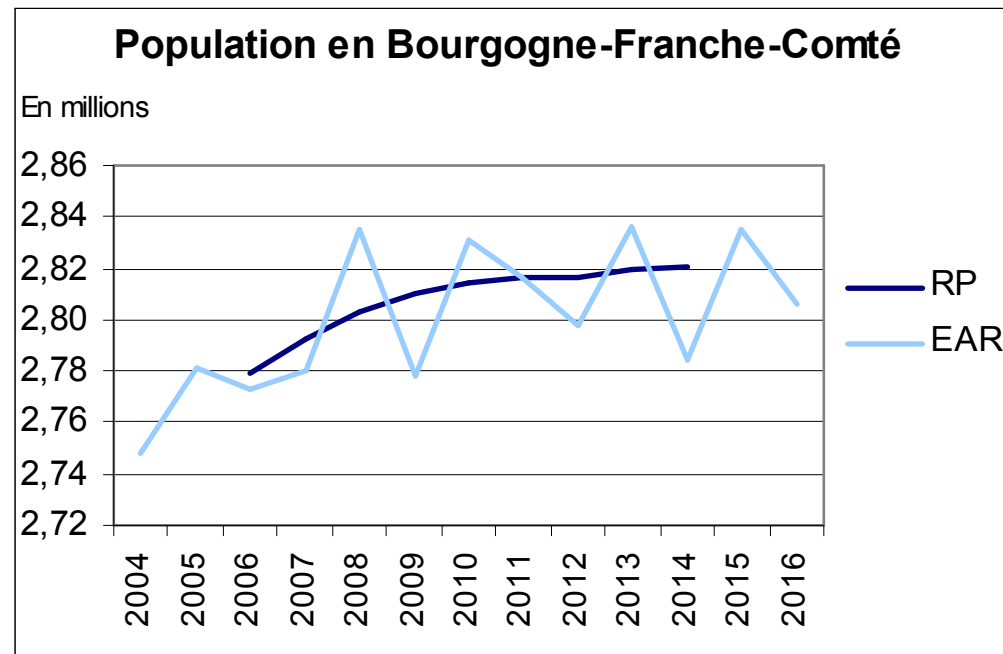
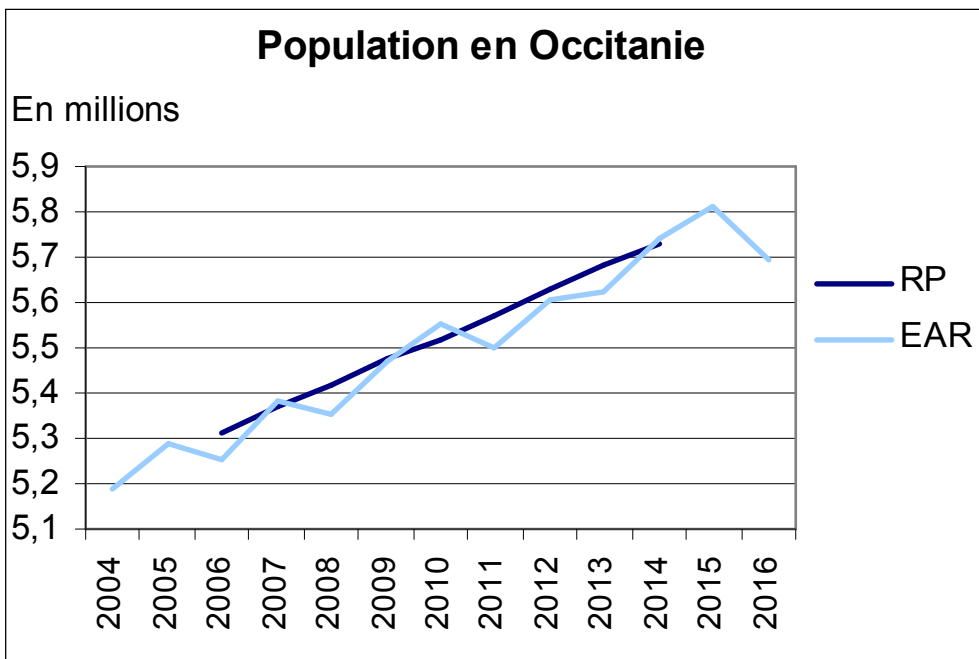
- ◆ Un produit facilement accessible
 - librement en interne à l'Insee
 - aux SSM et assimilés sur convention
- ◆ Recommandations d'utilisation restrictives
- ◆ malgré certains intérêts spécifiques
 - changements de questionnaires
 - utilisations plus conjoncturelles
(indicateurs annuels, évolutions récentes, ...)



L'EAR, une enquête de très grande taille ?

	Ménages Ordinaires		CMT
	PC-méto	GC-méto	
Nb. individus de l'échantillon	6 480 000	2 325 000	330 000
Nb. unités échantillonnées	7 000	350 000	6 000
Taux de sondage	20 %	8 %	20 %
Part de la population	49,3 %	45,2 %	2,4 %

Des résultats actuels volatils, même au niveau régional

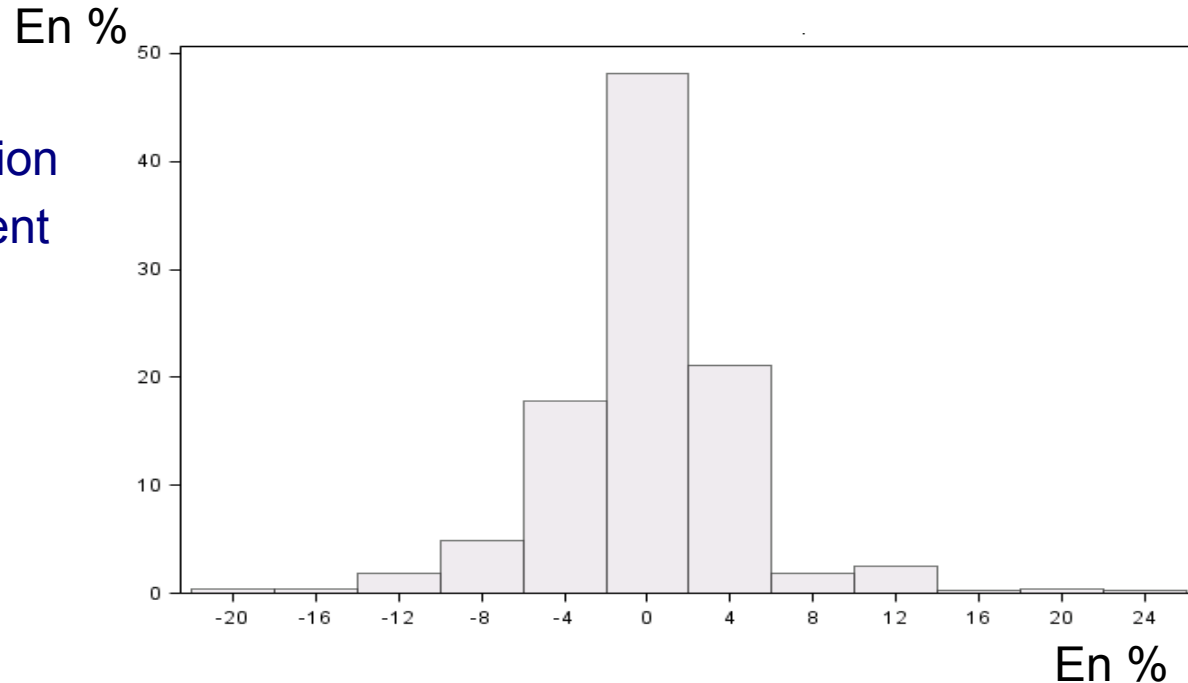


02

Équilibrage imparfait des groupes de rotation

- ◆ Par construction
 - difficulté d'équilibrer la répartition des PC dans les GR initialement

Ex. déséquilibres (en %) des GR départementaux vis-à-vis de la population (RP99)



- Déséquilibres pouvant s'amplifier au fil des ans
- ◆ Au niveau départemental, les petites communes contribuent toujours majoritairement aux déséquilibres des groupes de rotation

03

Nouvelle méthode de repondération

3.1. Détermination de la source auxiliaire

3.2. Petites communes de métropole et des DOM

3.3. Grandes communes de métropole

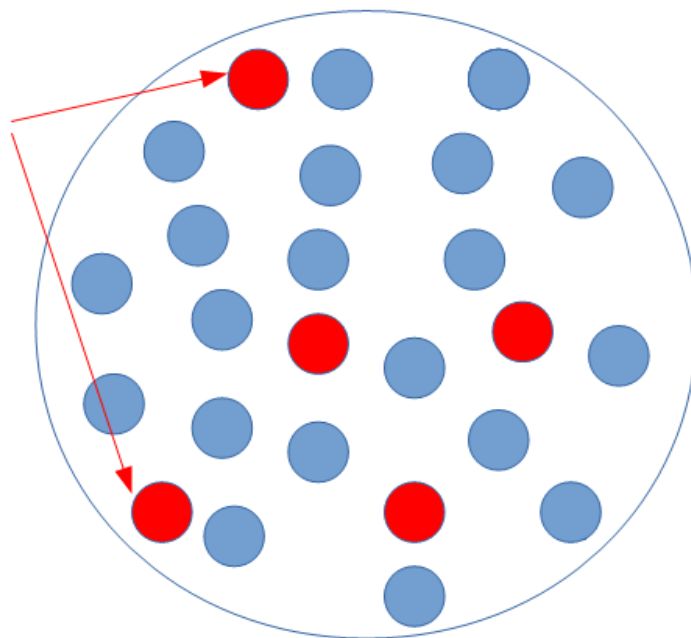
3.4. Grandes communes des DOM

3.5. Communautés


- ◆ **Pour le calage de l'EAR 2018 : Fidéli 2017**
 - Données exhaustives sur les logements et les individus mises à jour annuellement
 - Variables auxiliaires proches de celles du RP
 - Variables d'adressage
 - code commune, identifiant de la voie, numéro...
 - Deux ans de fraîcheur supplémentaires par rapport au dernier millésime de RP disponible (2015)

- ♦ Fidéli N-1 : fichiers logements et individus
⇒ variables auxiliaires pour chaque PC
- ♦ Dans chaque région :

Échantillon : les
PC recensées



Marges de calage :
Totaux régionaux
des var. aux.

- ◆ **Dans chaque commune :**
 - la Base de Sondage d'Adresses (BSA)
 - Échantillonnage en deux phases des adresses pour l'EAR
- ◆ Calcul des variables auxiliaires
 -  Pas d'identifiant commun entre la BSA et Fidéli

1ère phase



2nde phase



- ♦ **Comment apparier les adresses de la BSA et les adresses de Fidéli ?**
 - Plusieurs catégories d'adresses

Dégroupées

9 rue de la Tour, bâtiment A
9 rue de la Tour, bâtiment B

Non normalisées

0 rue Gambetta, villa Matisse
0 rue Gambetta, villa Monet

Avec suffixe

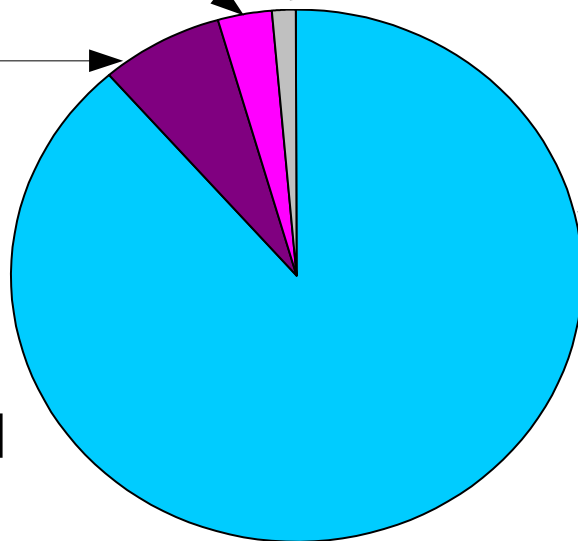
7 av de la Marne
7 bis av de la Marne

Simple

86 avenue Verdier, Montrouge

Avec un hôtel

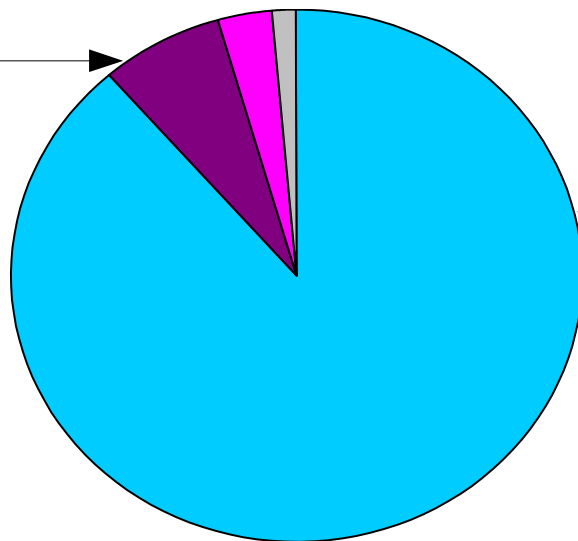
11 rue de Lille [habitation]
11 rue de Lille [hôtel]



- ♦ **Comment apparier les adresses de la BSA et les adresses de Fidéli ?**
 - Plusieurs difficultés possibles d'appariement

Avec suffixe
7 av de la Marne
7 bis av de la Marne

Difficulté possible dans Fidéli :
2 logements au 7 av de la Marne
mais **rien** au 7 bis



Difficulté possible dans Fidéli :
1 logement ayant pour adresse fiscale de traitement **86** avenue Verdier mais pour adresse foncière le n°**84**

Simples
86 avenue Verdier, Montrouge

- ♦ **Comment apparier les adresses de la BSA et les adresses de Fidéli ?**
 - Cas de recours à la référence cadastrale

Dégroupées

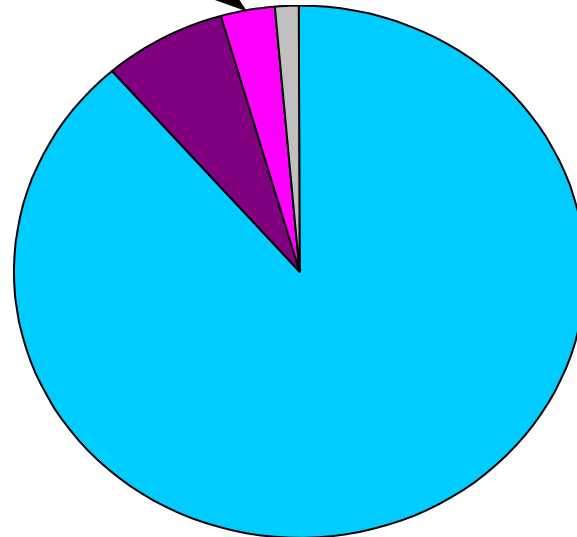
9 rue de la Tour, bâtiment A
9 rue de la Tour, bâtiment B

Non normalisées

0 rue Gambetta, villa Matisse
0 rue Gambetta, villa Monet

Difficultés possibles :

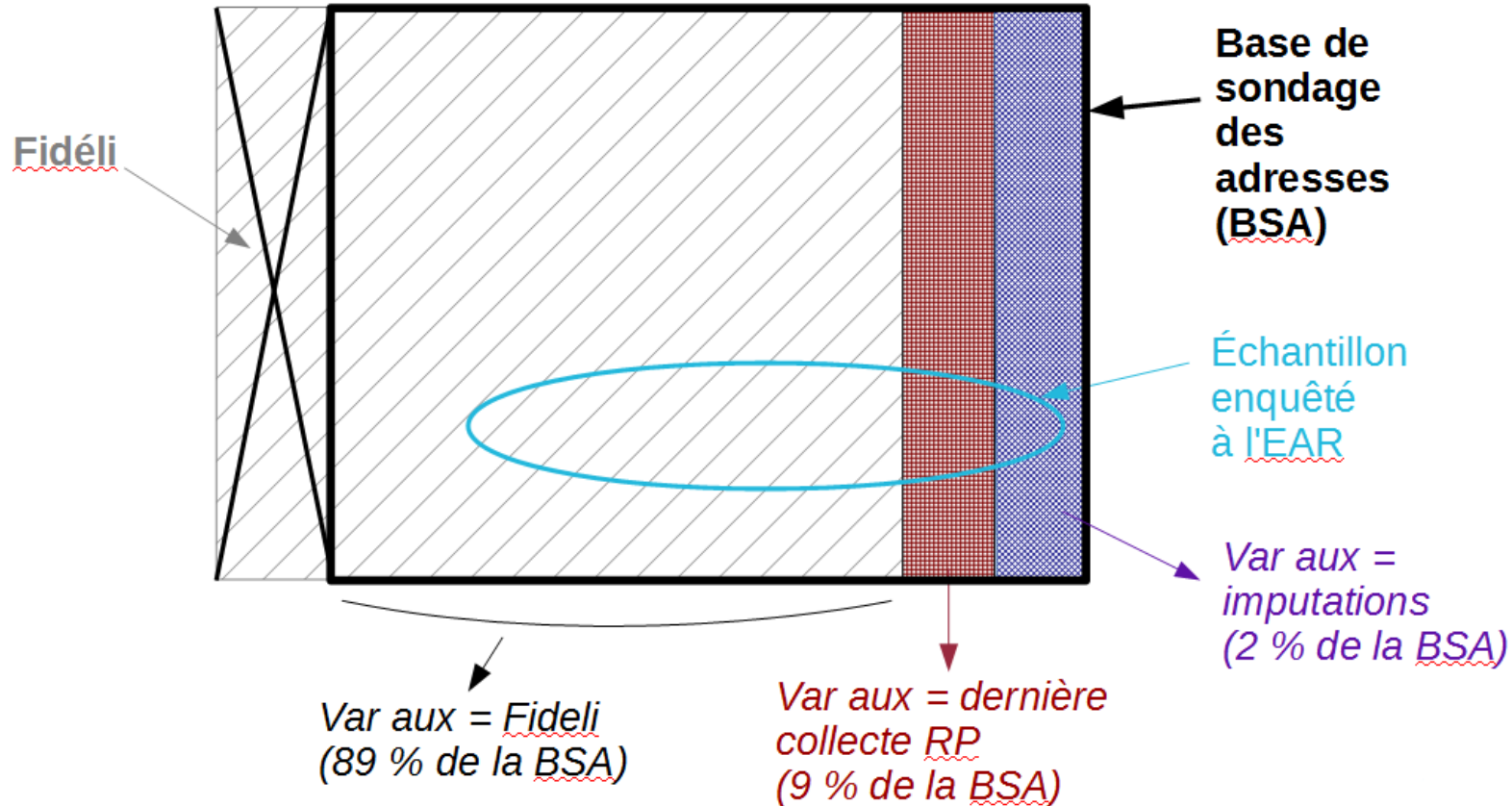
- pas de distinction suivant le bâtiment dans Fidéli
- tous les bâtiments sur la même référence cadastrale



Difficulté possible :

tous les bâtiments sur la même référence cadastrale

- ◆ **Calcul de variables auxiliaires composites**

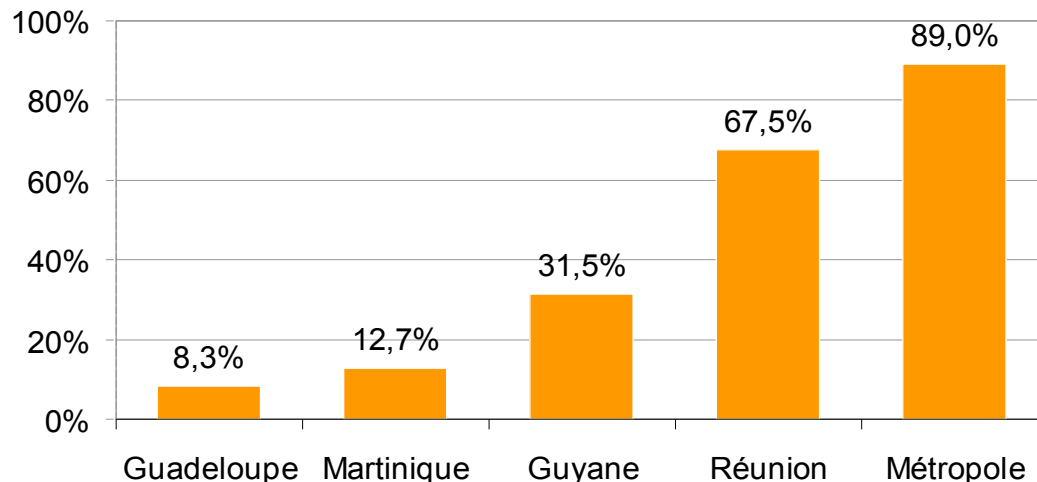


- ◆ Un calage commune par commune ($\approx 1\ 000$ calages)
 - Base de calage : BSA enrichie des variables auxiliaires
 - ⇒ Marges de calage : totaux des variables auxiliaires sur la BSA
 - Échantillon à caler : adresses d'habitation enquêtées à l'EAR (N)

- ◆ Variables auxiliaires composites
 - Nombre de logements de la BSA
 - Variables caractérisant la population (nombre d'individus, par tranche d'âge et sexe) résidant à l'adresse multi-sources :
 - Fideli (N-1)
 - Dernière collecte de recensement
 - Imputations

- ♦ Qualité hétérogène de l'appariement BSA – Fidéli

Taux d'appariement des adresses par zone

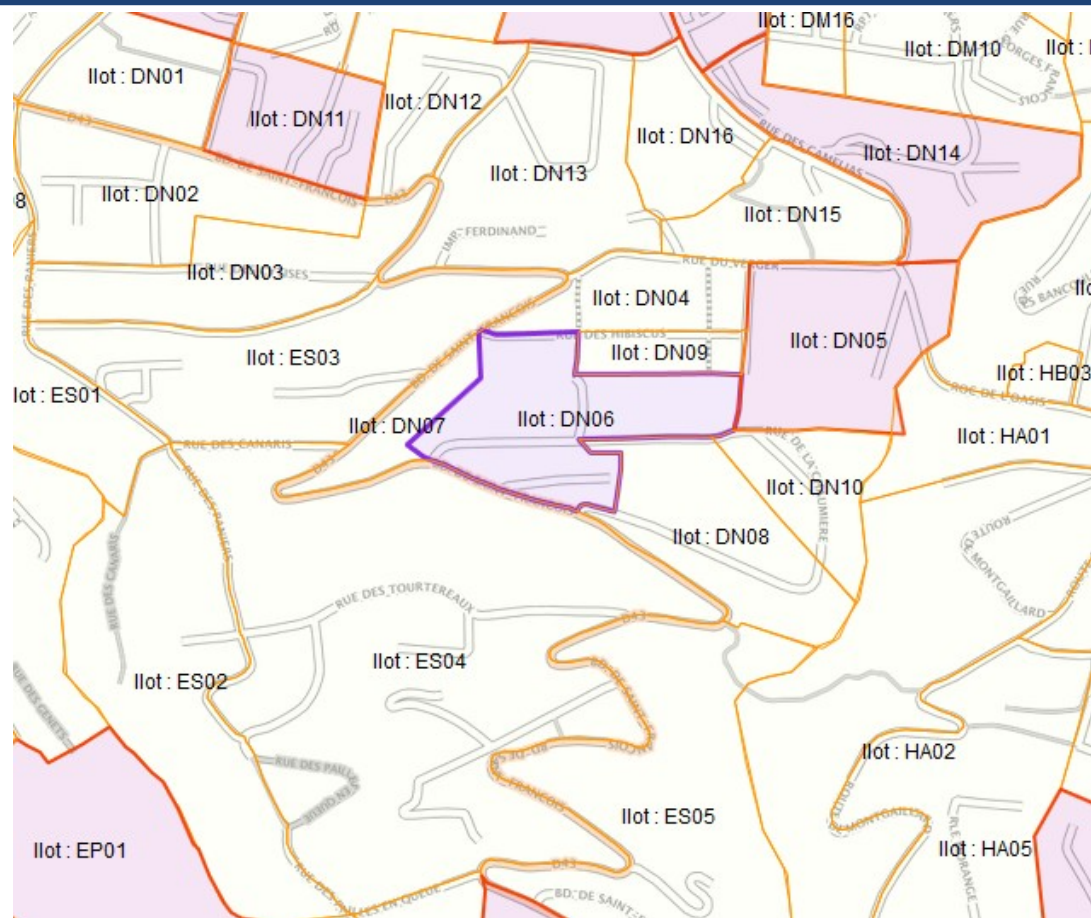


- ♦ Adresses non normalisées (sans numéro sur la voie)

- Majoritaires aux Antilles
- > 40 % en Guyane
- < 6 % à la Réunion et en métropole

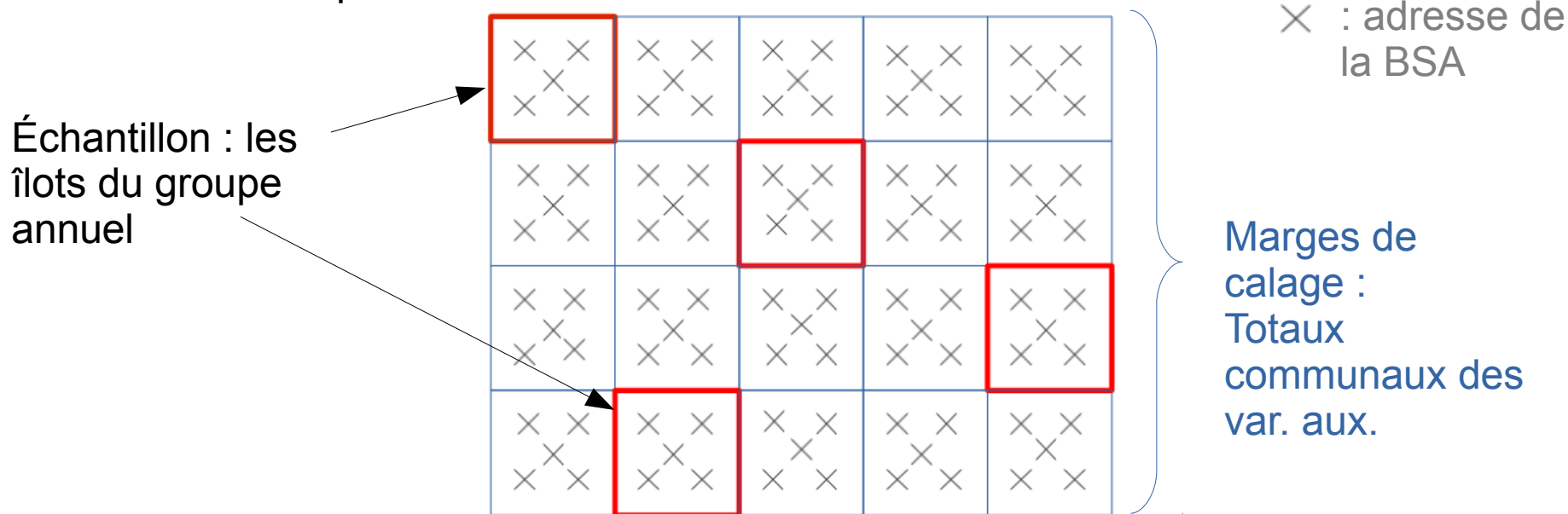
◆ Spécificités de l'échantillonnage

- 1^{er} degré de tirage : chaque commune est partitionnée en îlots répartis entre les GR
- 2nd degré de tirage : adresses tirées dans les îlots du GR annuel
- Jusqu'à l'EAR 2018, pas d'identifiant pérenne des entités adressées



- ♦ **Calage des îlots pour les Antilles et la Guyane**

- Dans chaque commune :





- Fidéli N-1 : variables auxiliaires pour chaque îlot

♦ Calage simultané à deux degrés pour la Réunion

- Dans chaque commune :

Échantillon

- du 1^{er} degré = îlots du groupe annuel (S_1) : 

- du 2nd degré = adresses tirées (S_2) : 



× : adresse non appariée

× : adresse appariée

\bar{v}_j = estim du nb de logements dans l'îlot j sur le champ de la BSA non appariée à Fidéli

x_i = nb total d'hommes ... dans l'îlot i (Fidéli)

v_i = estim du nb d'hommes ... dans l'îlot i sur le champ de la BSA appariée à Fidéli

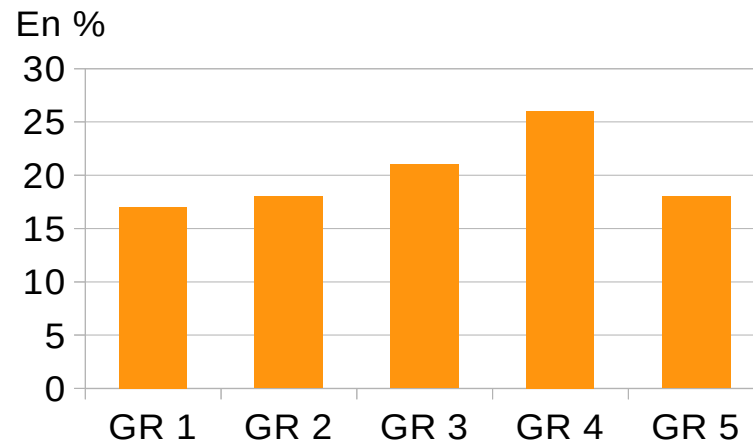
♦ Calage simultané à deux degrés pour la Réunion

- Dans chaque commune les informations auxiliaires propres aux adresses sont « remontées » au niveau de l'îlot (variables v_i, \bar{v}_i)
- Finalement : calage des îlots du groupe annuel
- Équations de calage :

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

- Marges :
 - X = totaux Fidéli des variables auxiliaires
 - V = totaux Fidéli des var. aux. sur la fraction appariée à la BSA
 - \bar{V} = nb log BSA sur la fraction non appariée à Fidéli

- ♦ GR parfois très déséquilibrés. Ex : population en Centre-Val de Loire



- ♦ Calage :
 - Échantillon à caler : communautés enquêtées à l'EAR (N)
 - Variables auxiliaires : populations par région, par catégorie de communauté, et selon certains croisements région x catégorie
 - Source : RP (N-3) car pas d'info identifiant les communautés dans Fidéli

04

Impact sur la qualité des estimations

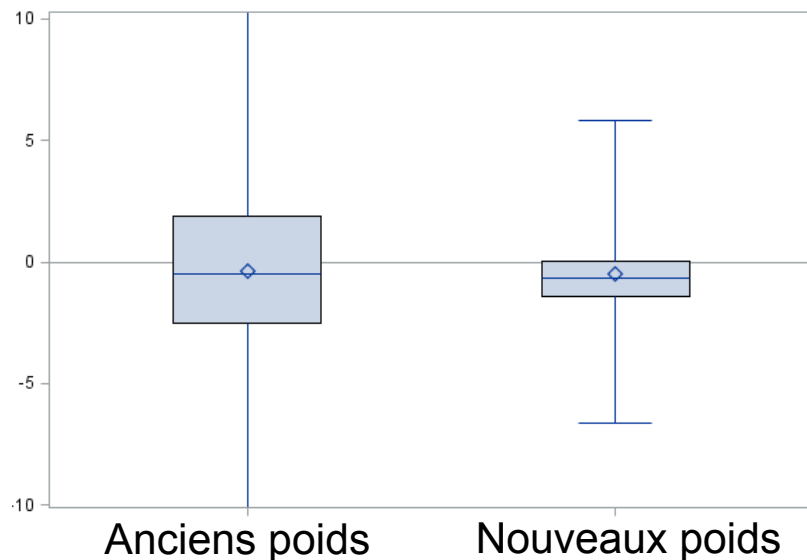
- 4.1. Comparaison EAR 2015-RP 2015
- 4.2. Séries chronologiques des EAR
- 4.3. Approche par les calculs de précision

- **La nouvelle pondération permet-elle d'améliorer les estimations de totaux à partir des EAR ?**

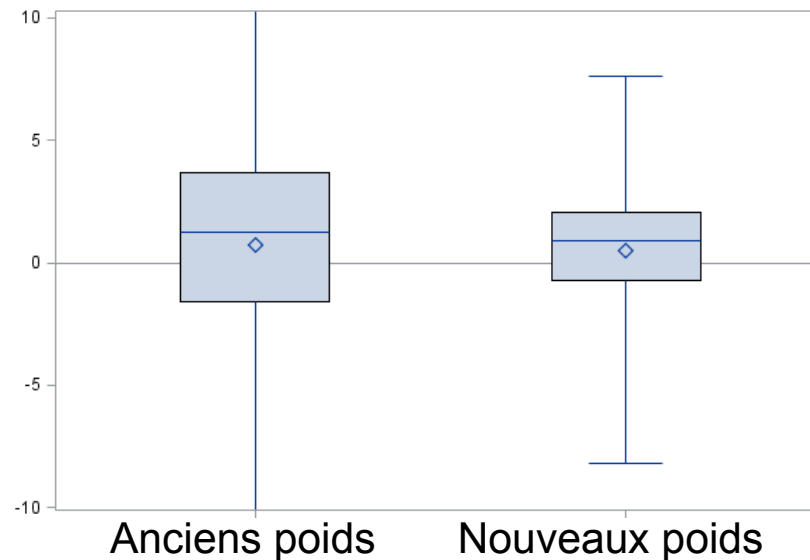
→ oui pour la population totale comme pour des populations plus spécifiques

Distribution des écarts EAR-RP (en %) par département en 2015

Pour la population totale



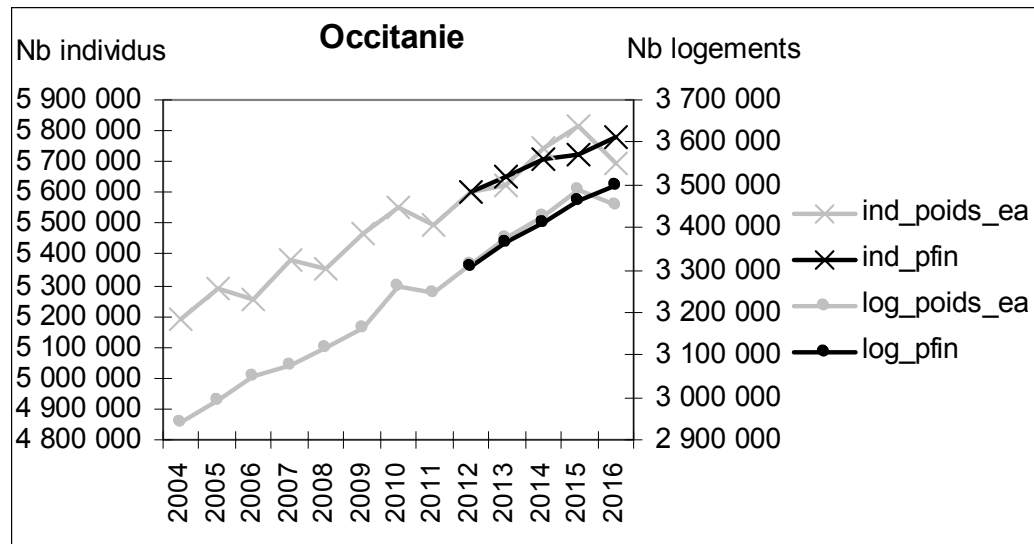
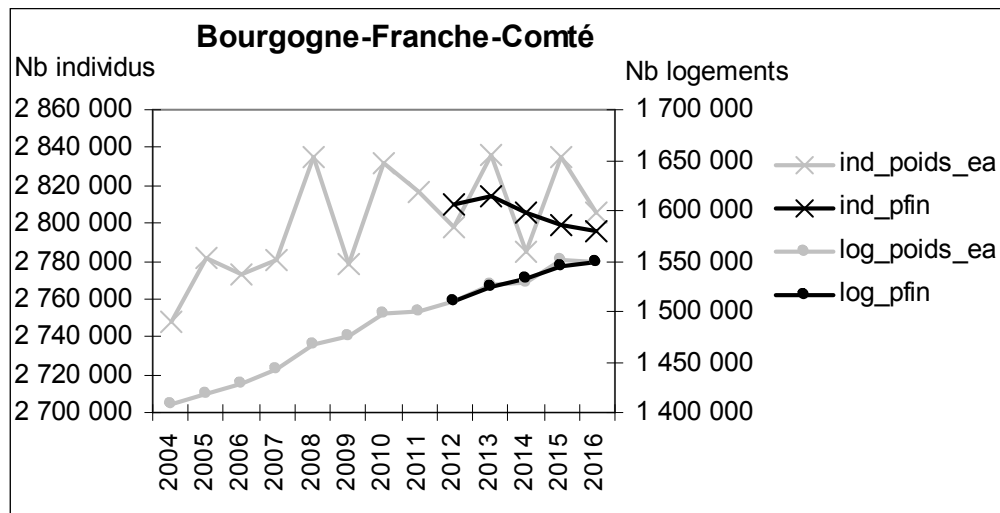
Pour le nombre d'actifs occupés à temps partiel



4.2. Séries chronologiques des EAR

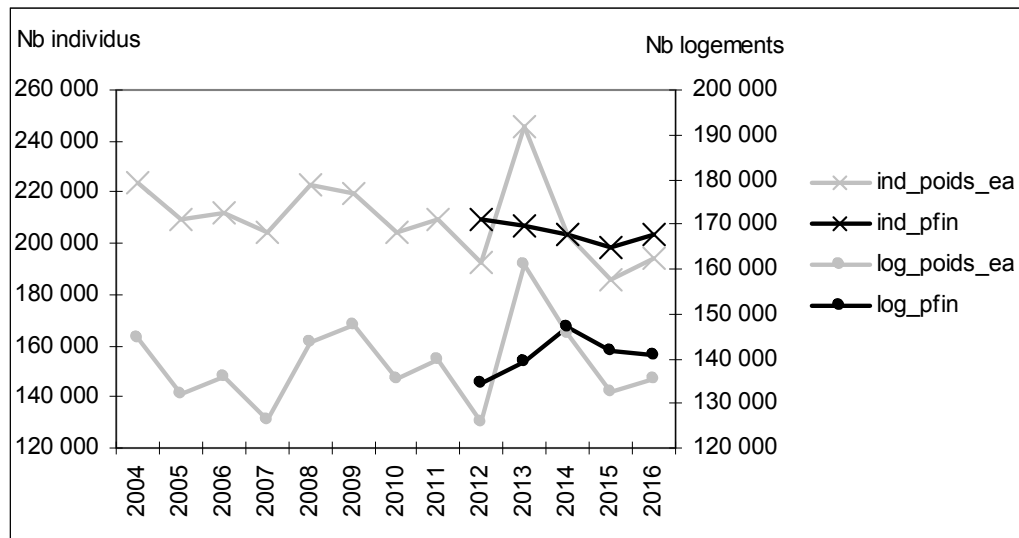
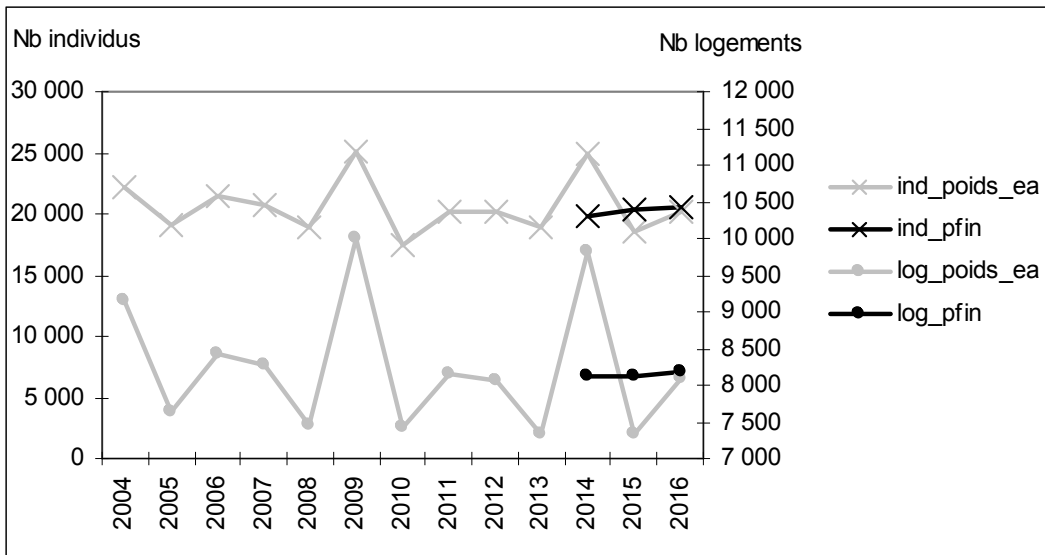
La nouvelle pondération lisse-t-elle les séries des logements et de population calculées à partir des EAR ?

→ oui pour 11 des 13 régions de France métropolitaine



Département de la Nièvre (58)

Logements ordinaires et leurs habitants



Grande commune Le Mée-sur-Seine (77285)

Logements ordinaires et leurs habitants

La nouvelle pondération permet de réduire la volatilité des séries de logements et de population pour ...

	Logements	Population	Total
Nb régions France métro.	11	11	13
Nb dépts France métro.	86	95	96
Nb DOM	4	3	4
% GC France métro.	79 %	67 %	100 %

Champ : logements ordinaires et leurs habitants.

- ♦ **Des résultats théoriques conformes à l'approche empirique**
- ♦ **Fort gain de précision grâce aux nouvelles pondérations**
 - Plus faible sur la série de logements et pour des populations « rares »
 - Plus élevé à des niveaux géographiques fins et sur les petites communes

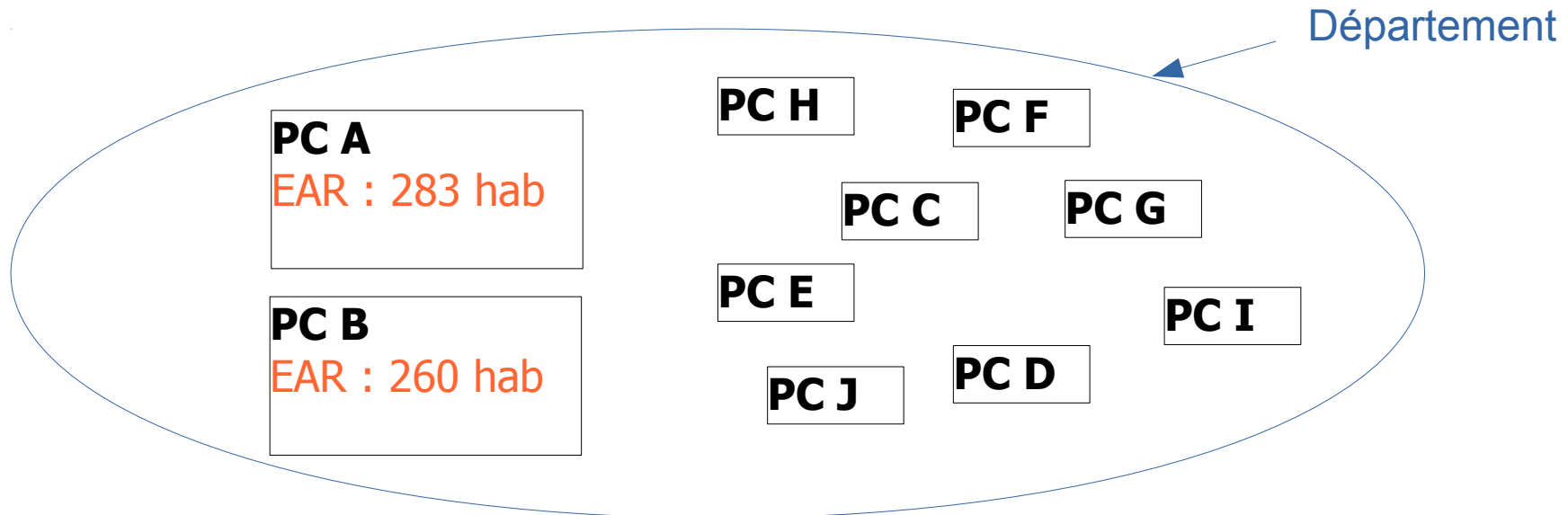
Retrouvez-nous sur :

[insee.fr](https://www.insee.fr)



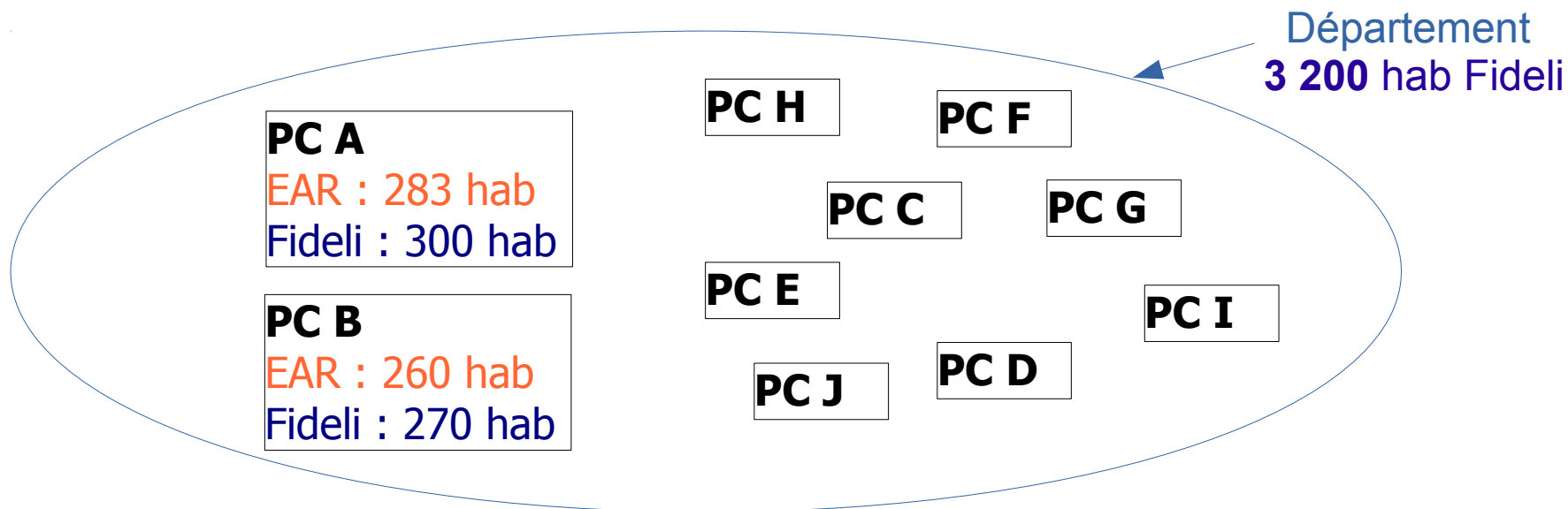
Sébastien Hallépée
Pierre-Arnaud Pendoli
Olivia Sautory

Division méthodes et traitements des recensements
Insee - Direction générale



Les deux PC A et B appartiennent au même groupe de rotation et sont donc enquêtées à la même EAR.

Nb habitants EAR du département estimé sans calage : $(283 + 260) * 5 = 2\ 715$

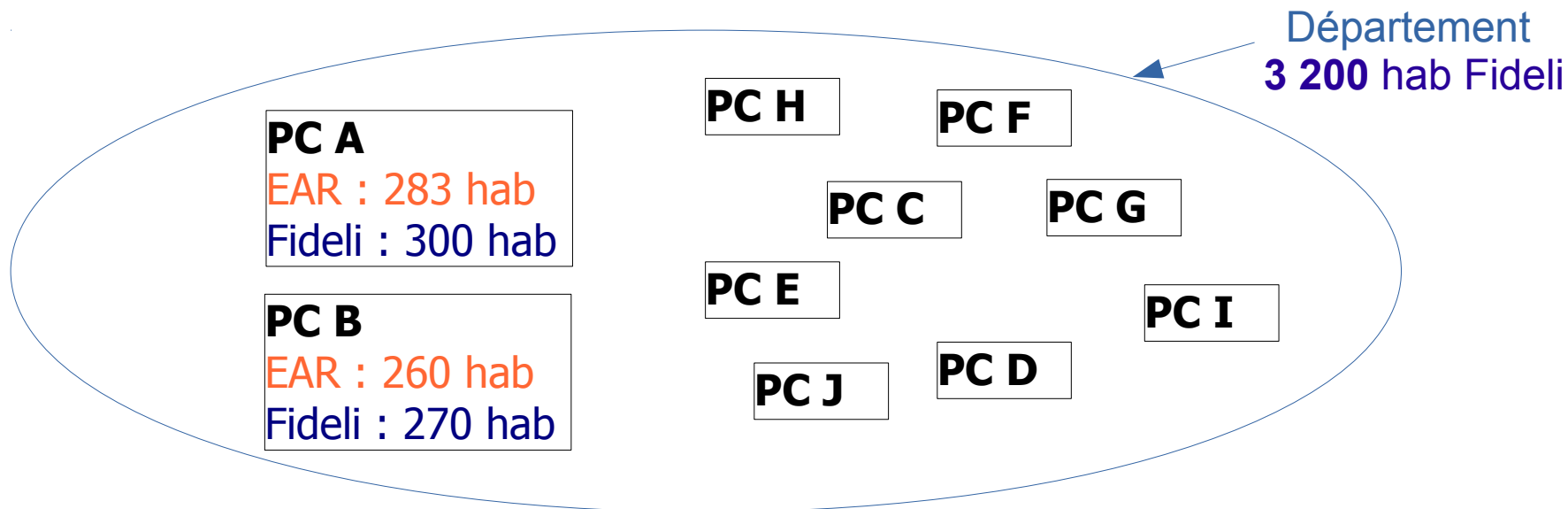


Les deux PC A et B appartiennent au même groupe de rotation et sont donc enquêtées à la même EAR.

Nb habitants EAR du département estimé sans calage : $(283 + 260) * 5 = 2\,715$

Nb habitants Fideli du département « estimé » : $(300 + 270) * 5 = 2\,850$

Facteur de calage Fideli : $3\,200 / 2\,850 = 1,12$



Les deux PC A et B appartiennent au même groupe de rotation et sont donc enquêtées à la même EAR.

Nb habitants EAR du département estimé sans calage : $(283 + 260) * 5 = 2\ 715$

Nb habitants Fideli du département « estimé » : $(300 + 270) * 5 = 2\ 850$

Facteur de calage Fideli : $3\ 200 / 2\ 850 = 1,12$

⇒ Nb habitants du département estimé avec calage : $2\ 715 * 1,12 = 3\ 041$ ($\neq 3\ 200$)

♦ **EAR 2016 :**

- Variables auxiliaires issues de Fideli 2015
- Rapports de poids : de 0,25 à 2 (contre 0,95 à 1,06 avec l'ancienne pondération)
- Par rapport à l'ancienne pondération : +0,44 % logements et +0,14 % individus.
- Distribution des écarts relatifs (en %) par département

	Min	P5	Q1	Méd.	Q3	P95	Max
Nb log	-17,9	-7,7	-0,9	+0,1	+1,0	+9,7	+33,2
Nb ind	-16,8	-10,1	-3,2	+0,2	+5,0	+12,3	+32,7