

Vers une utilisation accrue des sources externes au RP

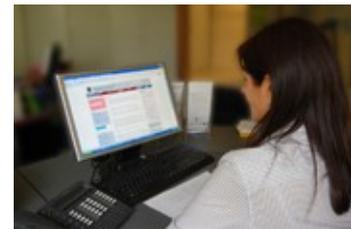
Réunion de division

Sébastien HALLÉPÉE - DMTR

Division Méthodes et traitements des recensements



Mesurer pour comprendre



30 sept.
2016

Contexte du recensement français

- Recensement en continu bâti sur un cycle de cinq enquêtes annuelles
- Petites communes réparties en cinq groupes et recensées exhaustivement sur une des années du cycle
- Grandes communes enquêtées chaque année sur un échantillon d'adresses représentant au total 40 % des logements de la commune sur le cycle
- Estimations des populations et des résultats statistiques tous les ans sur les 36 000 communes

Sources auxiliaires

Une utilisation déjà effective

- Fichiers des permis de construire pour mettre à jour la base de sondage d'adresses des grandes communes
- Fichiers fiscaux
 - pour contrôler la collecte
 - pour estimer les petites communes en extrapolation
- Améliorations possibles
 - Utilisation affinée grâce à une meilleure connaissance de la source

La méthode par sondage en grande commune

- Méthode de tirage équilibré d'adresses
- Limiter l'influence des grandes adresses
 - Strate exhaustive à partir de 60 voire plus de 200 logements
- Tirer parti de la base de sondage
 - Calage sur le nombre de logements par quartier (Iris)
- Quelques fragilités
 - Plus petites des GC, Non normalisation, profil touristique
 - Sensibilité sur certaines adresses atypiques (flux et effet d'échantillonnage)

Méthode de calcul de population pour les grandes communes jusqu'au RP 2013

- Calcul mené au sein de chaque Iris de la commune pour toutes les adresses de logements « ordinaires »

$$\begin{array}{r} \text{Ril « médian »} \longrightarrow \text{pop_men RP}_{2013} \\ = \\ \text{nb log au 01/01/2013} \\ \times \\ \text{[nb pers par log GC]} \\ \uparrow \\ \text{Cumul de 5 EAR} \\ \text{2011 à 2015} \end{array}$$

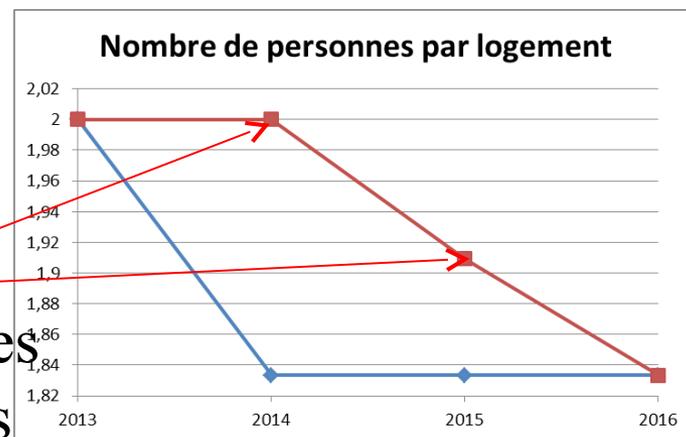
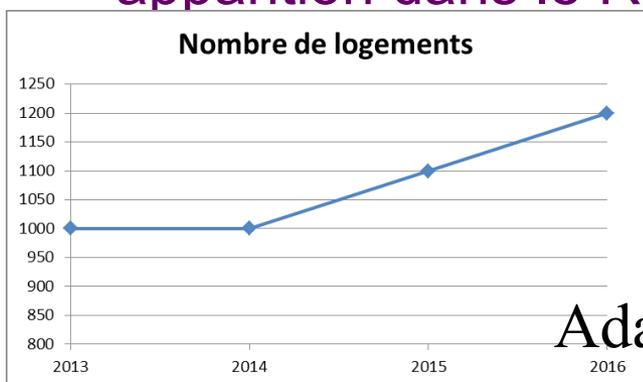
Remédier aux fragilités

- Adresses atypiques
 - Différence marquée du nombre de personnes par logement au sein de la strate d'estimation
 - Détection sur un critère d'impact → winsorisation
 - Mise en cohérence des flux
- L'apport de la strate sur les campings et les résidences hôtelières
- Apport potentiel du recours à une autre source

Adresses atypique

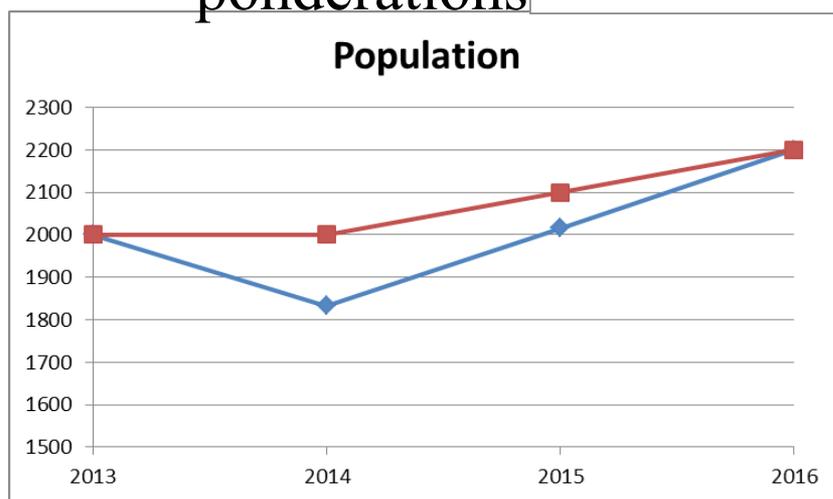
Un traitement en aval

- Coordonner arrivée d'une grande adresse neuve entre
 - fenêtre d'estimation du RP
 - apparition dans le Ril



Adaptation des pondérations

Exemple :
Résidence étudiante
apparue en juillet
2015 interrogée dès
2016

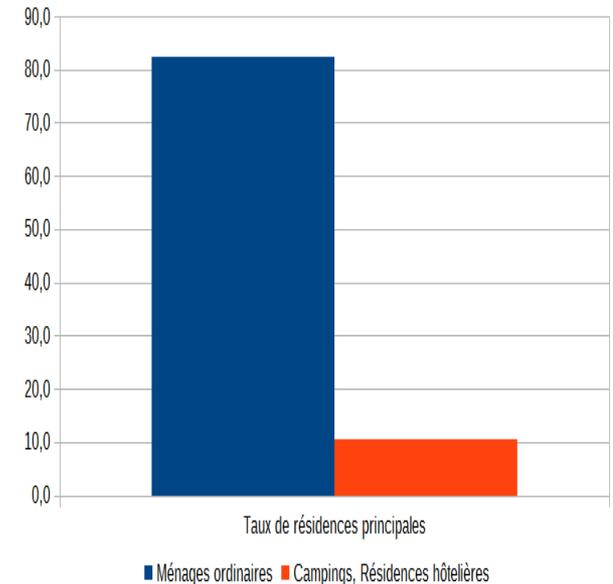
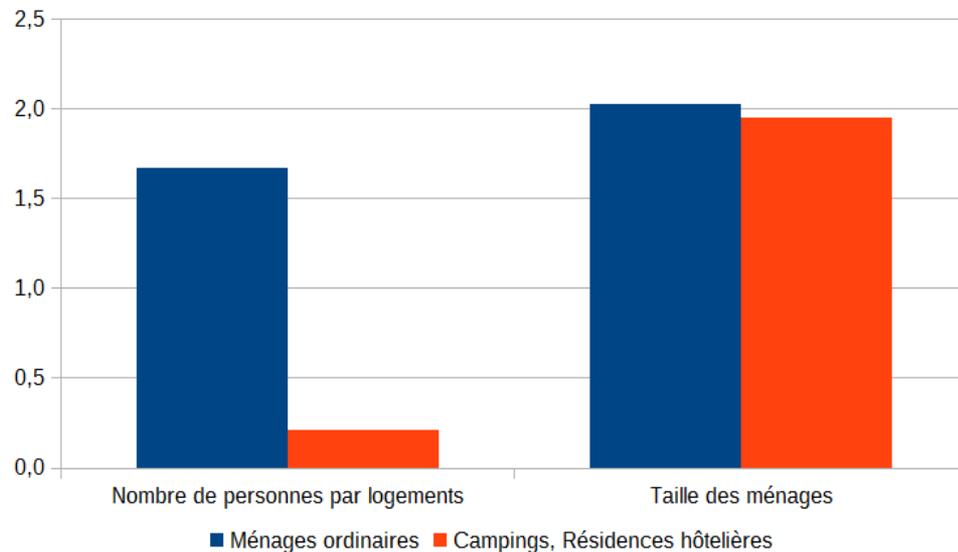


Correction des fragilités en amont

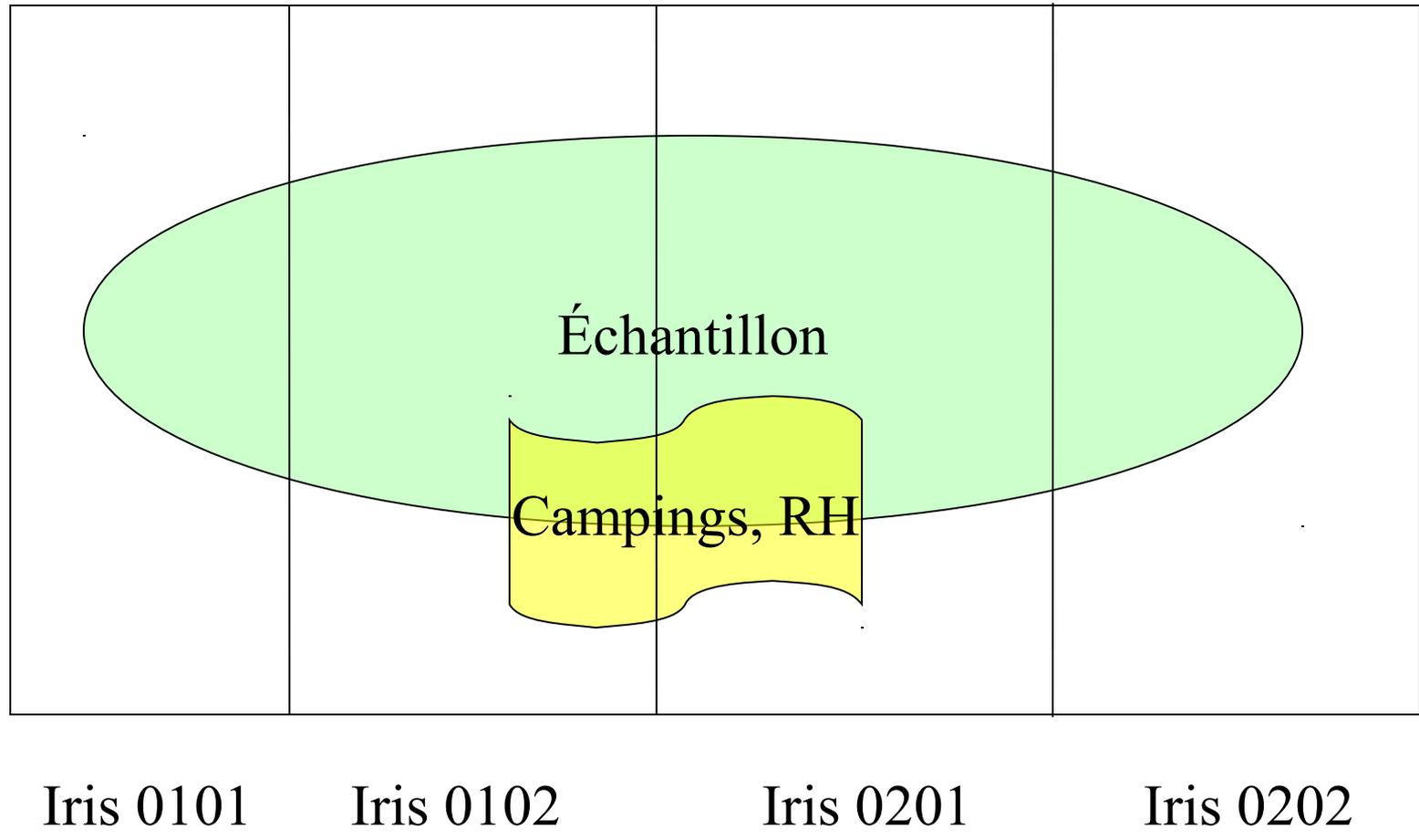
- Repérage de ces adresses dans la base de sondage
- Prise en compte dans l'échantillonnage
 - Strate à part
 - Nouvelles variables d'équilibrage
 - L'exhaustivité, ça n'a pas de prix ?
- Prise en compte dans la pondération
 - Strate à part
 - Nouvelles variables de calage

Pourquoi les structures touristiques ?

- Source récurrente de problèmes identifiés
- Repérage possible à partir d'une liste externe
- Caractéristiques spécifiques



Schématisation de la méthode



Nouvelle post-stratification

- Élimination de l'aléa de sondage sur cette population

Résultats des
estimations

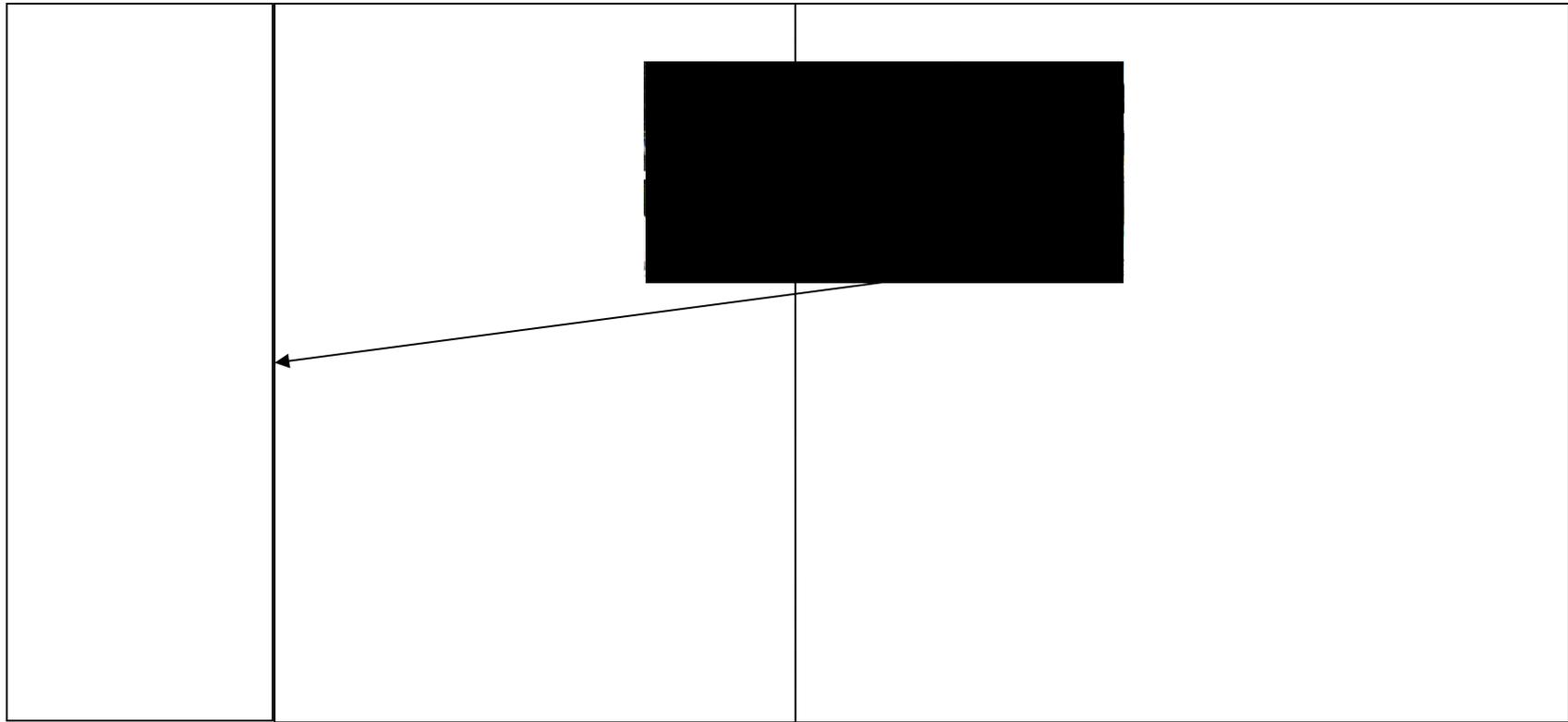


Iris 0102

Iris 0201

Nouvelle post-stratification

- Élimination de l'aléa de sondage sur cette population



Structures de
tourisme

Iris 0102

Iris 0201

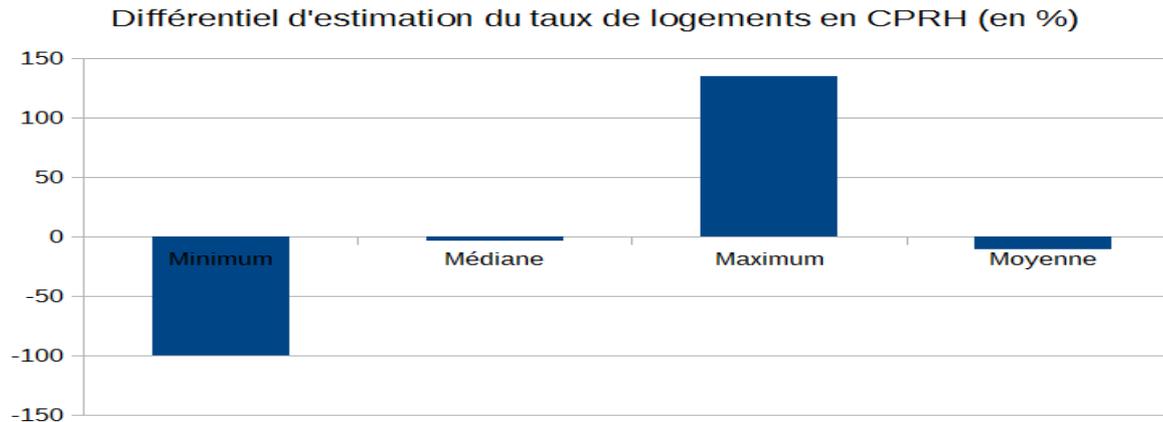
Principe nouvelle méthode

- Hôtels → Élargissement du principe pour les hôtels isolés
 - Hôtel compte tel quel dès son premier recensement
- Camping et Résidences hôtelières
 - Construction d'une strate, indépendante des habitations de leurs Iris
 - Nécessaire rétopolation de ces structures en phase transitoire
 - Meilleure prise en compte des flux par un Ril médian spécifique

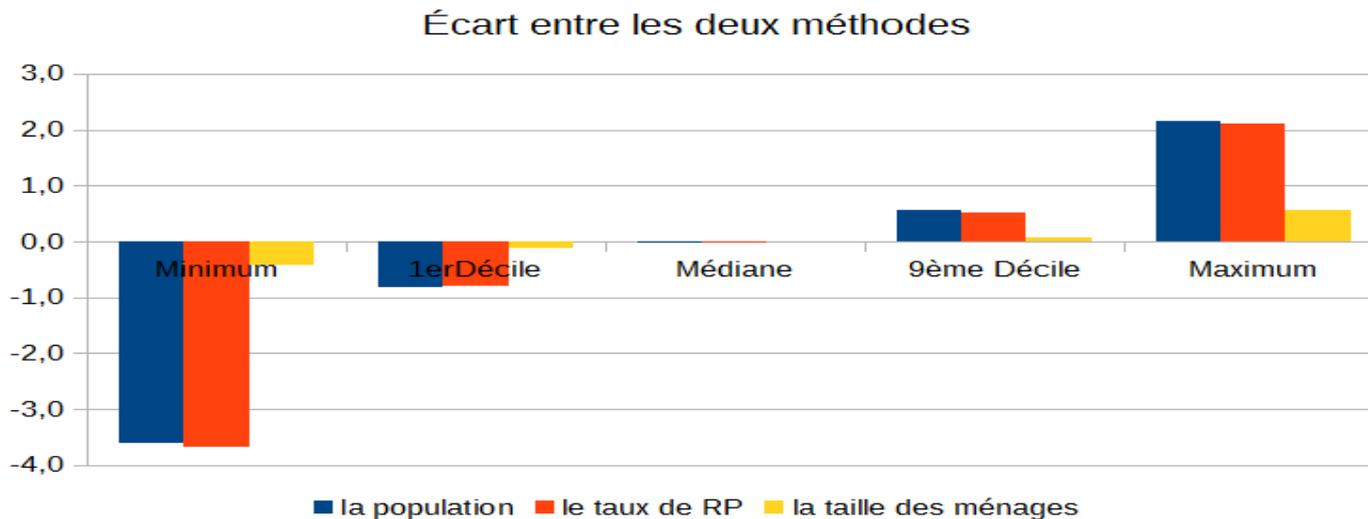
Effet sur les estimations

Comparaison sur une année

- Meilleure représentativité de la strate atypique CPRH

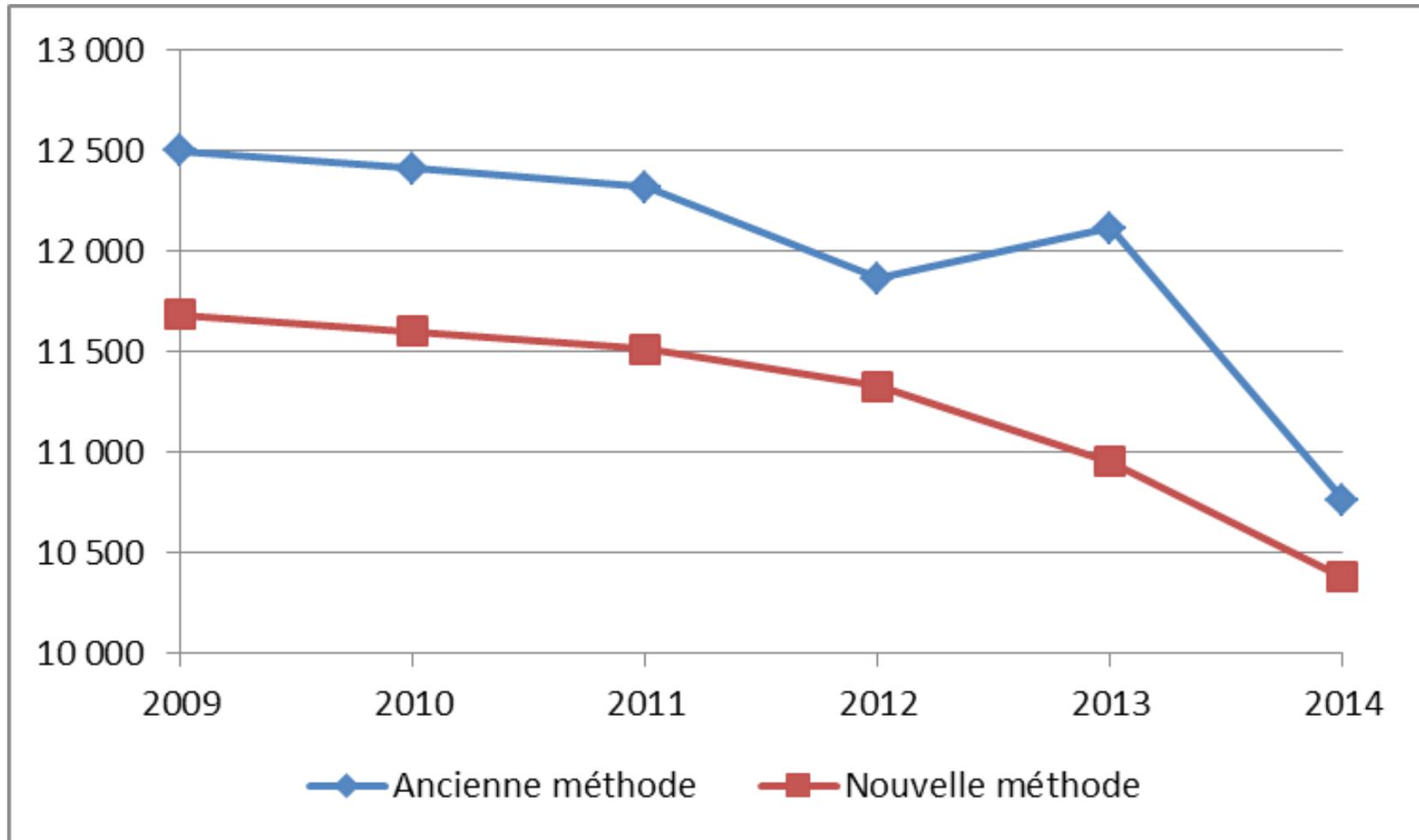


- Correction concentrée sur une meilleure appréciation du taux de résidences principales



Effet sur plusieurs années

Stabilisation de la population

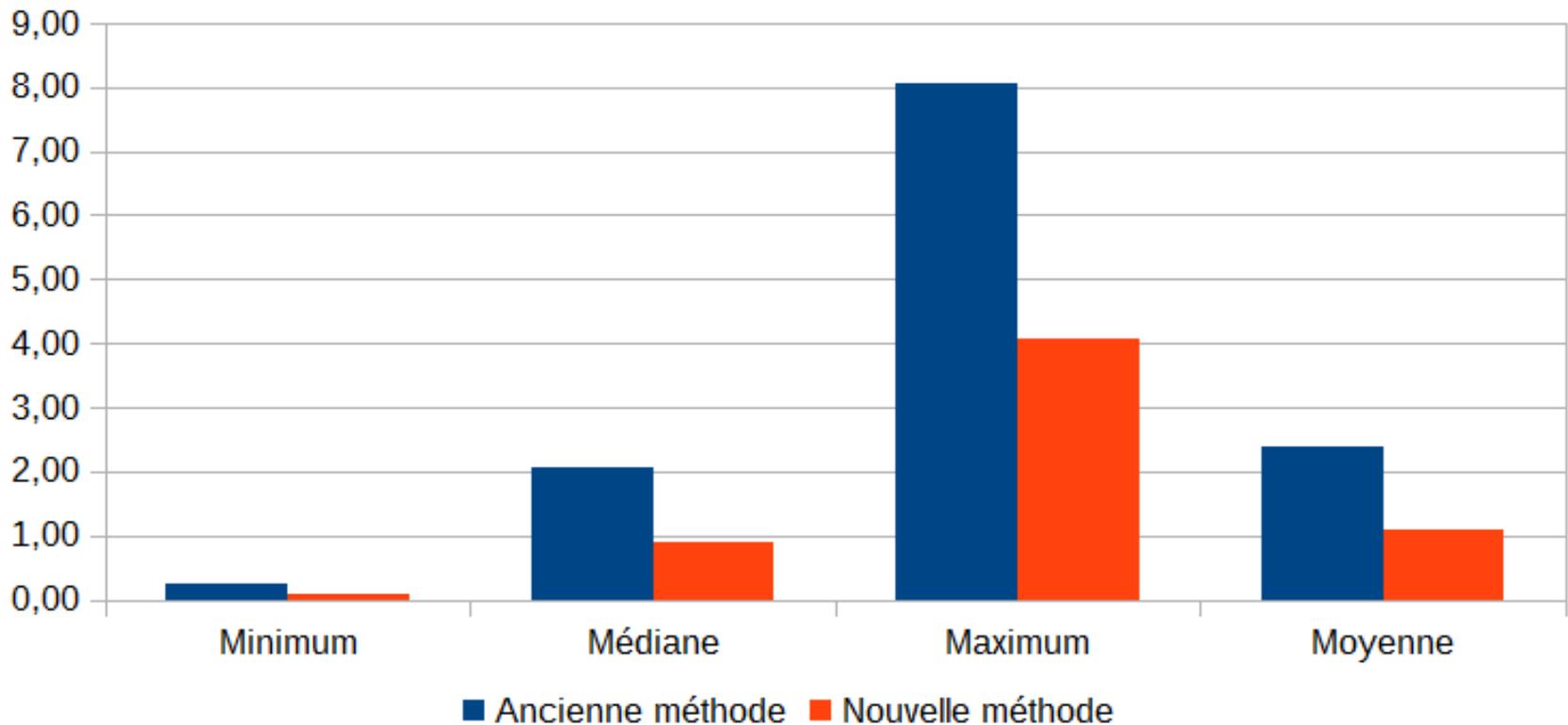


Effet sur les estimations

Stabilisation de la population

- Méthode en moyenne deux fois plus stable.

Écart par rapport à une évolution lissée sur 5 ans



Élargir le principe grâce aux sources fiscales

- Possibilité de constituer une strate d'adresses atypiques
 - Plus grandes des adresses s'écartant du nombre moyen de personnes par logement de la commune
 - Au tirage et pour les estimations
- Possibilité d'utiliser la source fiscale pour introduire des variables auxiliaires complémentaires
 - Principe : on essaie que l'échantillon soit à l'image de la base de sondage (nmpl, structure démographique...)
 - Variables d'équilibrage
 - Variables de calage

Un apport stratégique de l'appariement

- Affranchissement d'un travail manuel
 - Plus d'asymétrie de traitement
 - Réduction des coûts de la procédure
- Difficultés de l'appariement au niveau adresse
 - Concept d'adresse différent selon la source
 - Absence de normalisation
 - Un effort pour rapprocher les sources
- Base existante solide en cours d'amélioration
 - Taux d'appariement « direct » de 92 %
 - Amélioration en vue grâce au cadastre et rapprochement des données individuelles

Des bénéfices collatéraux

- Méthode petits domaines pour données carroyées
- Diffusion d'une enquête de recensement plus complète et robuste
- Amélioration des méthodes de correction de non réponse

Merci de votre attention !

Sébastien HALLÉPÉE

Division Méthodes et traitements des recensements

Insee

18 bd Adolphe-Pinard
75675 Paris Cedex 14

www.insee.fr  

Informations statistiques :

www.insee.fr / Contacter l'Insee

09 72 72 4000

(coût d'un appel local)

du lundi au vendredi de 9h00 à 17h00