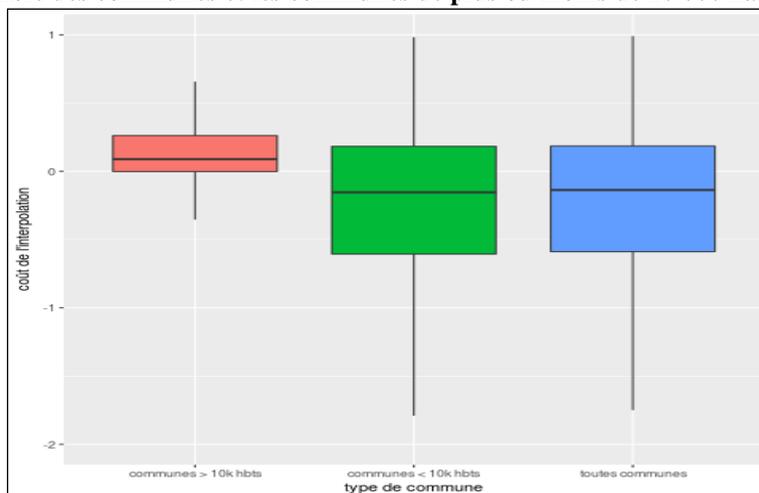


Complément en ligne C1 – Écart produit par l'interpolation spatiale

Figure C1-1

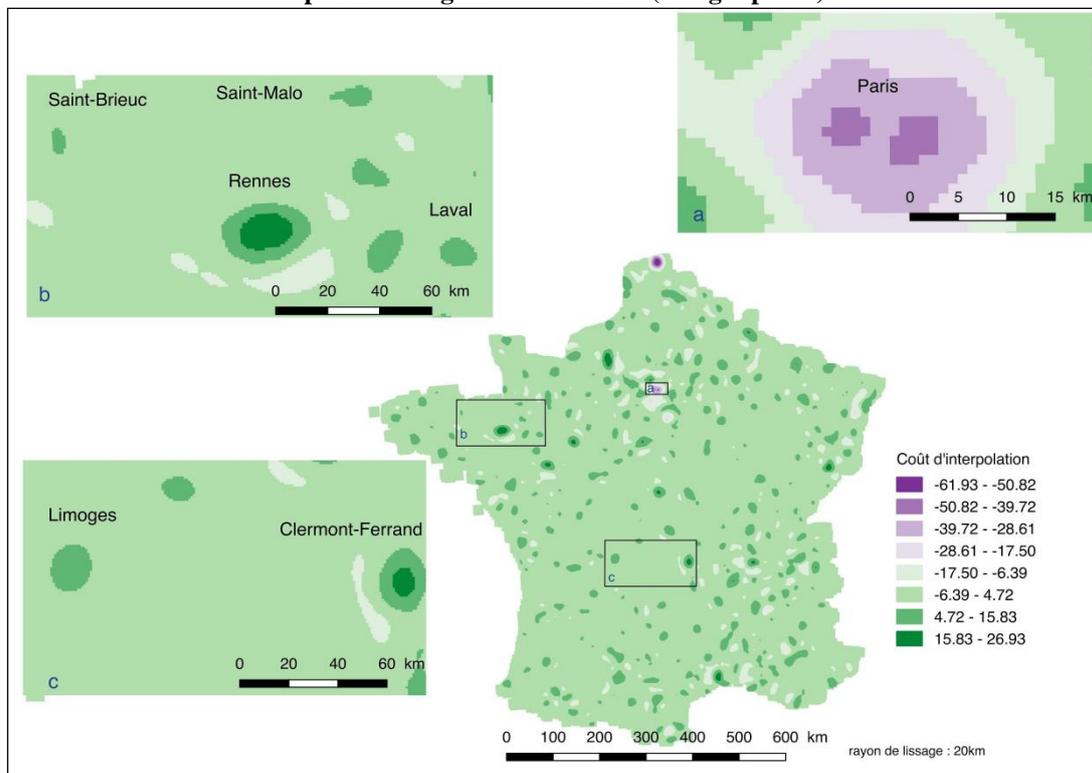
Distribution de l'écart relatif de la population communale et de l'estimation obtenue par interpolation spatiale, ensemble des communes et les communes de plus ou moins de 10 000 habitants



Lecture : un écart négatif correspond à une surestimation, un écart positif à une sous-estimation.
 Source : Filosofi 2011.

Figure C1-2

Écart brut de la population communale et de l'estimation obtenue par interpolation spatiale à partir de la grille des Voronoï (lissage spatial)

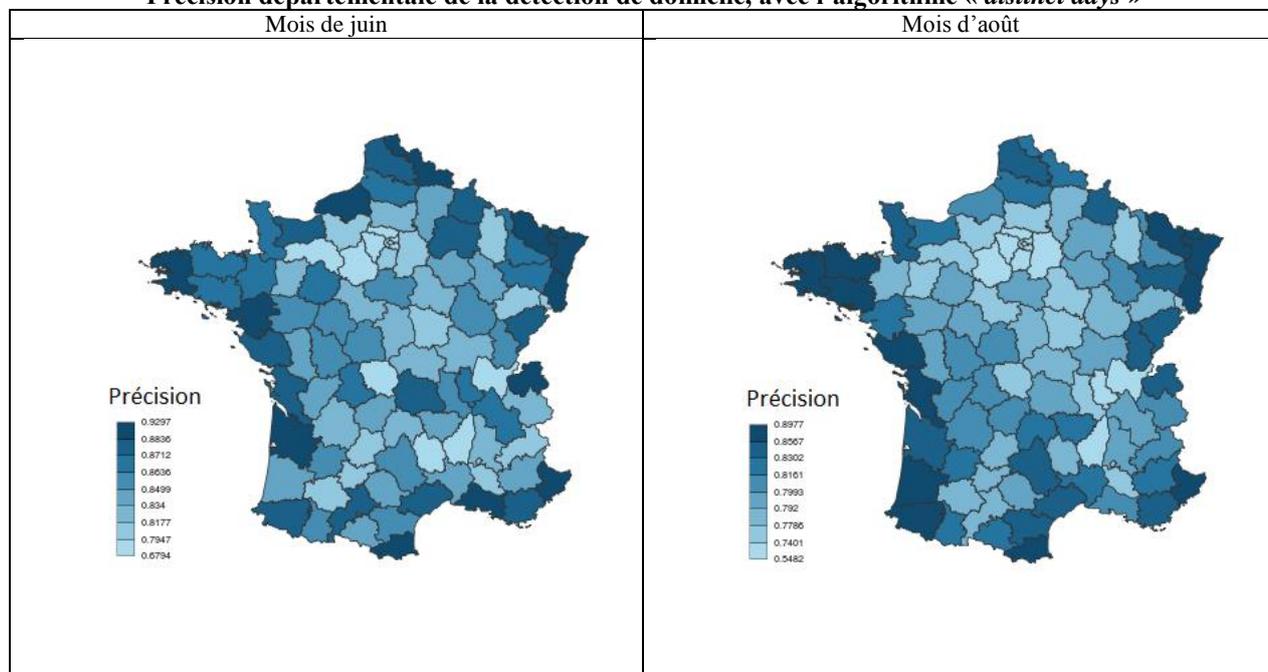


Lecture : un écart négatif correspond à une surestimation, un écart positif à une sous-estimation.
 Source : Filosofi 2011.

Complément en ligne C2 – Détection de domicile

Figure C2-1

Précision départementale de la détection de domicile, avec l'algorithme « *distinct days* »



Source : CDR, fichier clients.

Estimer la population résidente à partir de données de téléphonie mobile, une première exploration*

Benjamin Sakarovitch, Marie-Pierre de Bellefon, Pauline Givord et Maarten Vanhoof

Compléments en ligne / Online complements

Complément en ligne C3 – Maille d’analyse

Pour estimer les coefficients du modèle de l’équation (4) inspiré de la méthode de Deville *et al.* (2014) pour les densités de population nous utilisons un modèle linéaire généralisé avec une famille de Poisson adaptée aux comptages que nous cherchons à produire. Les CDR auxquels nous avons eu accès sont les mêmes que ceux qui ont été agrégés dans cette précédente étude.

Le tableau A1 représente les différentes estimations obtenues pour le coefficient β . Ce coefficient varie assez largement selon la maille considérée. Il est très largement inférieur pour les polygones de Voronoï et les Iris par rapport aux trois mailles à une échelle plus élevée pour lesquelles β est très proche de 1. À titre de comparaison, Deville *et al.* (2014) obtenaient une valeur plus faible pour ce coefficient, de 0.84 au niveau communal. Les deux études diffèrent néanmoins à la fois sur l’objet d’étude (des densités de population plutôt que des comptes absolus) et sur les données (nombre d’utilisateurs quotidiens par antenne plutôt que nombre de populations résidentes).

Tableau C3
Coefficient de superlinéarité dans l’équation (4) selon la maille

Maille	Voronoï	Iris	Commune	Zone d’emploi	Département
β	0.39	0.50	0.96	1.02	1.04

Source : CDR, Filosofi 2011.

Estimer la population résidente à partir de données de téléphonie mobile, une première exploration*

Benjamin Sakarovitch, Marie-Pierre de Bellefon, Pauline Givord et Maarten Vanhoof

Compléments en ligne / Online complements

Complément en ligne C4 – Analyse des écarts d'estimation au niveau communal à partir des caractéristiques communales

Pour une analyse plus systématique des écarts constatés entre les populations communales fournies par la source fiscale et les estimations obtenues au niveau communal par les données de téléphonie mobile, on régresse ces erreurs d'estimation (relativement au niveau de la population communale de référence) sur plusieurs caractéristiques communales: le nombre de polygones de Voronoï intersectant le territoire de la commune, le coût de l'interpolation – comme décrit dans la partie 2.2 – pris relativement à la population de la commune, les taux de pénétration régionaux comme donnés par l'Arcep, la taille de la commune et enfin les parts de marchés départementales estimées à l'aide du fichier client. Dans une deuxième spécification, on utilise outre le coût de l'interpolation et le nombre de Voronoï à la place de ces variables « internes », une variable issue de la base communale de *Corinne Land Cover*, qui permet de construire le pourcentage de la commune appartenant à des zones naturelles (forêt, glaciers et neiges éternelles, marais, plans d'eau, lagunes littorales, estuaires, mers et océans), pour lesquelles on peut supposer que sur ces zones la couverture est mauvaise. La variable dépendante est la différence entre le niveau de population communale (source fiscale) et l'estimation fournie par les données de téléphonie mobile, rapportée à la population fiscale. Un signe négatif d'une variable signifie donc que les estimations sur-estiment la population.

L'estimation confirme que le facteur déterminant pour expliquer l'ampleur des divergences est lié à l'approximation liée à l'interpolation spatiale (Tableau C4). Le coefficient est en effet proche de un, suggérant une relation directe entre ces deux erreurs. En fait, un modèle ne comprenant que cette variable explique plus des deux tiers de la variance des écarts relatifs entre population estimée et de référence. Les autres caractéristiques communales ne permettent pas d'améliorer significativement la capacité prédictive du modèle. Le coefficient correspondant à la taille de la commune est positif (alors même qu'on s'intéresse à une variable où elle intervient au dénominateur), ce qui conforte le fait que les estimations ont plutôt tendance à sous-estimer la taille des grandes communes. En revanche le nombre de Voronoï intersectant la commune a un coefficient négatif – suggérant à l'inverse une tendance à la sur-estimation quand ce nombre est élevé (plus cette intersection est élevée, plus le risque existe d'attribuer à tort à une commune un habitant d'une commune voisine appartenant au même Voronoï). En revanche, le pourcentage de zones naturelles n'a pas d'effet significatif une fois contrôlée de ces caractéristiques (même si on obtient une corrélation négative entre cette variable et les écarts dans un modèle avec cette seule variable).

Tableau C4

Corrélation entre erreurs d'estimation de la population communale et les caractéristiques des communes

Constante	0.57*** (0.07)	0.19*** (0.01)
Coût de l'interpolation relatif	0.95*** (3.40e-3)	0.95*** (0.003)
Population communale	1.16.e-5*** (1.75.e-6)	
Nombre de Voronoï	- 3.64e-2*** (4.12e-3)	- 2.4e-2*** (0.002)
Taux de pénétration régional (Arcep)	- 0.65*** (5.88e-2)	
Parts de marché départementales	0.52*** (0.14)	
Proportion du territoire en zones naturelles		-5.5e-4 (2.8e-3)
R2	0.69	0.69

Note : modèle linéaire simple, la variable dépendante est la différence relative entre la population communale enregistrée dans la source fiscale et l'estimation fournie par les données de téléphonie mobile redressée à l'aide du fichier client et *, **, *** indiquent la significativité aux seuils de 10, 5 et 1 %.

Estimer la population résidente à partir de données de téléphonie mobile, une première exploration*

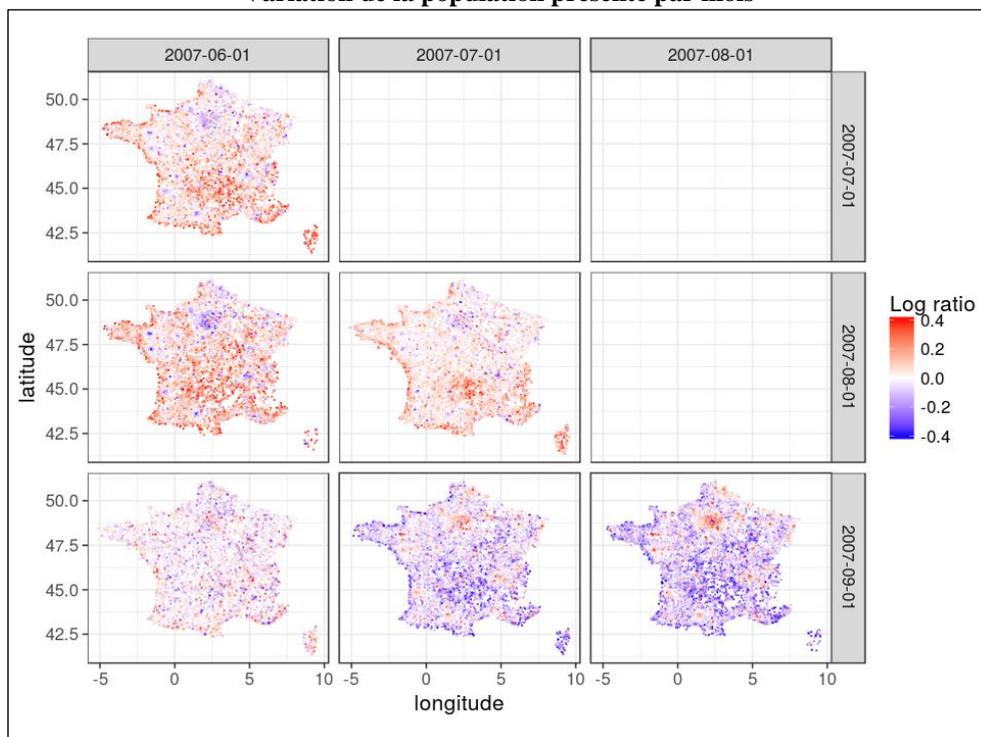
Benjamin Sakarovitch, Marie-Pierre de Bellefon, Pauline Givord et Maarten Vanhoof

Compléments en ligne / Online complements

Complément en ligne C5 – Variations saisonnières de population résidente

Figure C5

Variation de la population présente par mois



Lecture : entre juin et août les antennes montrent que la population détectée comme habitant autour des antennes bleues claires a diminué de moins de la moitié.

Source : CDR.