

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

Question clé

Les données issues de la téléphonie mobile fournissent des enregistrements avec une résolution spatiale élevée et à une haute fréquence temporelle. Elles sont considérées comme une source prometteuse pour mesurer la mobilité des personnes et la fréquentation des territoires à des niveaux spatiaux ou temporels très fins. Leur utilisation à des fins statistiques soulève un certain nombre de questions (qualité des informations collectées, représentativité des données disponibles, retraitements nécessaires), que cet article se propose d'examiner en s'intéressant à l'estimation d'indicateurs de populations résidentes.

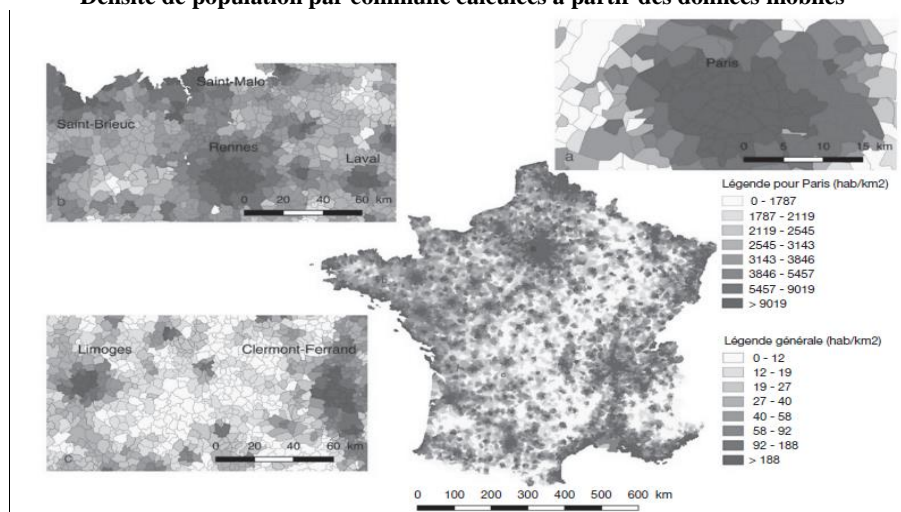
Méthodologie

Des données anonymisées fournies par un opérateur de téléphonie mobile sur cinq mois de l'année 2007 permettent d'assigner le lieu de résidence de chaque utilisateur à la zone couverte par l'antenne la plus récurrente dans la transmission d'appels. La population résidente ainsi construite est comparée à celle fournie par une source fiscale géolocalisée.

Principaux résultats

- Les estimations de population résidente à partir de téléphonie mobile s'écartent significativement de la population de référence fournie par la source fiscale
- La modélisation de la couverture des antennes par tessellation de Voronoï puis la transposition à la grille communale des estimations de population résidente expliquent l'essentiel des écarts constatés.
- Les estimations au niveau communal sont plus précises qu'au niveau de la zone de couverture de l'antenne - et presque autant qu'à une maille plus large (zone d'emploi, département).
- Le redressement de l'hétérogénéité locale des parts de marché de l'opérateur améliore la qualité de l'estimation, mais nécessite une information auxiliaire telle qu'un fichier client, dont la disponibilité et la fiabilité ne sont pas garanties.
- Les écarts entre la population résidente estimée et de référence présentent une structure spatiale marquée.
- Les données mobiles semblent toutefois prometteuses pour l'étude des variations saisonnières de population.

Densité de population par commune calculées à partir des données mobiles



Note : les estimations sont redressées au niveau départemental à l'aide du fichier client.

Source : CDR, fichier client et Filosofi ; calculs des auteurs.

Message

L'exploitation des données mobiles pour des comptages de population soulève plusieurs questions méthodologiques. La modélisation de la zone de couverture des antennes, le choix de la maille d'analyse et la localisation des événements sur cette maille jouent un rôle déterminant, tout comme les méthodes de redressements pour passer des estimations sur les clients d'un opérateur à la population générale. L'utilisation de données avec des enregistrements plus fréquents de la position des abonnés que les données de facturation pourrait permettre de gagner en précision pour la détection des points d'ancrage (comme le lieu de résidence). L'étude de la population présente et des variations spatio-temporelles de la densité de population font partie des applications les plus prometteuses pour la statistique publique.