

Utilisation des données d'opérateurs de téléphonie mobile par la Statistique Publique

Exemple avec l'utilisation pendant le Covid-19

Marie-Pierre de Bellefon, Lino Galiana

Insee

JEMU - 17 octobre 2022

Plan

L'usage par l'Insee des données de téléphonie mobile

L'expérience de mesure de population lors du Covid

Perspectives

L'usage par l'Insee des données de téléphonie mobile

Les données mobiles : pour répondre à quel besoin ?

Mesure de la population présente

- ▶ Avis de moyen terme du CNIS 2019 - 2023 : "le Conseil souhaite que les travaux d'expertise des nouvelles sources de données se poursuivent dans l'objectif de mesurer la population présente sur un territoire à différents moments de la journée, différents jours de la semaine ou mois de l'année."
- ▶ Besoin exprimé par les acteurs locaux pour calibrer les offres de soin, traitement des déchets, services de secours,...

Les données mobiles : pour répondre à quel besoin ?

Mesure de la mobilité

- ▶ Mieux comprendre les interactions entre territoires et les comportements quotidiens des individus, vers leur lieu de travail et vers les centres de loisir.
- ▶ Confrontation aux zonages d'étude.
- ▶ Eclairage complémentaire aux données traditionnelles, plus récent et plus fréquemment mis à jour.

Groupe de travail des instituts de statistique européens

- ▶ Partage d'expérience (partenariats existant en Belgique, Hollande, Italie,...)
- ▶ Harmonisation des méthodologies : packages R pour la localisation des évènements et pour le redressement, rapports en accès libre,...
- ▶ Poursuite des réflexions au niveau européen

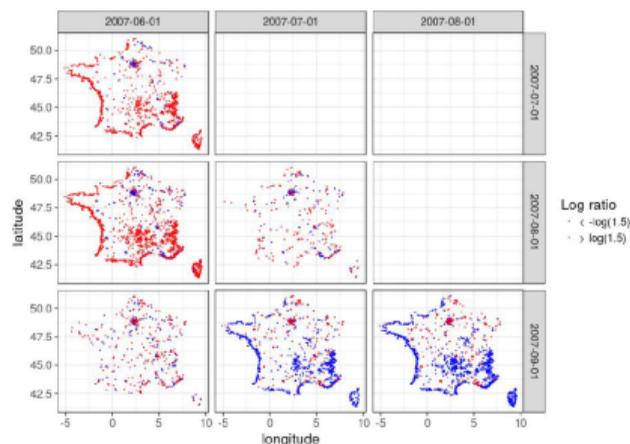
2016 - 2022 : partenariat de recherche avec Orange et Eurostat

Données

- ▶ CDR (Call Detail Record) de 2007 conservées à des fins de recherche. Accessibles dans les locaux d'Orange.
- ▶ Limites
 - ▶ trace uniquement lors d'un appel \Rightarrow pas info continue
 - ▶ pas d'info sur la couverture réelle de l'antenne \Rightarrow approximation géométrique (Voronoi)
- ▶ Résultats
 - ▶ plusieurs articles de recherche qui éclairent des sujets d'intérêt pour la statistique publique
 - ▶ une meilleure compréhension des données \Rightarrow bonne base pour de futures collaborations avec Orange

Illustration : travaux sur la population présente

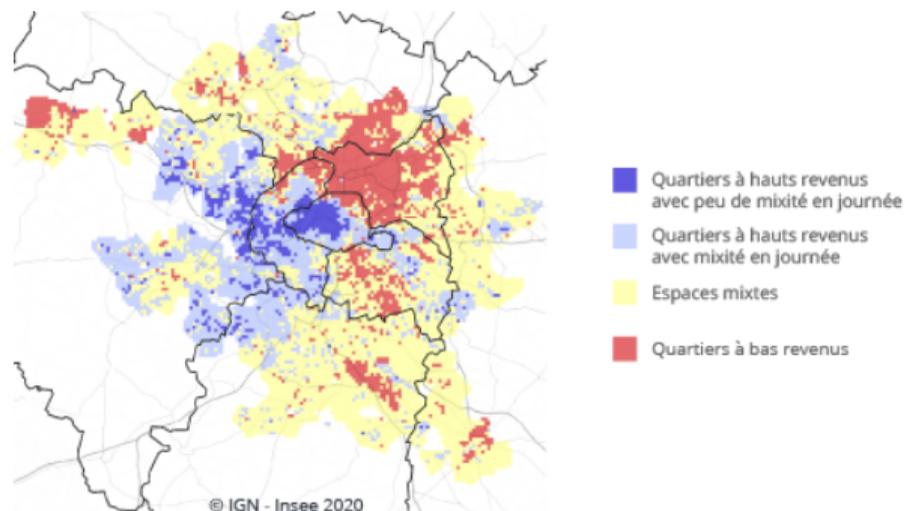
Figure – Variation de la population présente par mois



Lecture : Entre juin et août la population détectée comme habitant autour des antennes rouges foncées a plus que doublé (essentiellement sur le littoral et en montagne). *Source* : Orange, CDR 2007 ; calculs Insee - Bellefon, Givord, Sakarovitch, Vanhoof 2019

Illustration : travaux sur la ségrégation résidentielle

Figure – Typologie des espaces en fonction des dynamiques de population



Les espaces franciliens sont regroupés en quatre catégories en fonction de leur dynamique de population. *Source* : Orange, CDR 2007 ; calculs Insee - *Galiana, Sakarovitch, Sémécurbe, Smoreda 2020*

L'expérience de mesure de population lors du Covid

Contexte

- ▶ Mise en place rapide du confinement
- ▶ Rumeurs puis annonce (16 mars 2020 au soir) avant la mise en oeuvre (17 mars à 12h)
- ▶ Prolongation à plusieurs reprises.
- ▶ Fin le 11 mai, entrée dans une phase de déconfinement graduel (juqu'au 2 juin)

Enjeux

- ▶ Très tôt, médias se focalisent sur les *"Parisiens [qui] s'exilent à la campagne"*;
- ▶ Rapidement, chiffre de *"plus d'un million de Franciliens"* ayant quitté la région parisienne circule dans la presse
- ▶ Besoin de statistiques objectives et fiables :
 - ▶ Combien de personnes se sont déplacées ?
 - ▶ Quel profil des personnes s'étant déplacées ?

Nature des données

- ▶ Données commerciales mises à disposition gratuitement;
 - ▶ Usage exclusif pour des sujets Covid-19
 - ▶ Accès aux données 3 opérateurs : d'abord Orange et Bouygues puis SFR
 - ▶ Données du 16 janvier au 31 mai 2020 (mais pas mêmes plages selon opérateurs);
- ▶ Données journalières de présence au niveau départemental :
 - ▶ Agrégées et repondérées par les opérateurs;
 - ▶ Besoin de recalage supplémentaire pour cohérence avec données Insee.
- ▶ Conditionnées en fonction du département de résidence :
 - ▶ Déterminé par les opérateurs de téléphonie mobile;
- ▶ Données plus fines (EPCI) non exploitées faute de temps;

Challenge

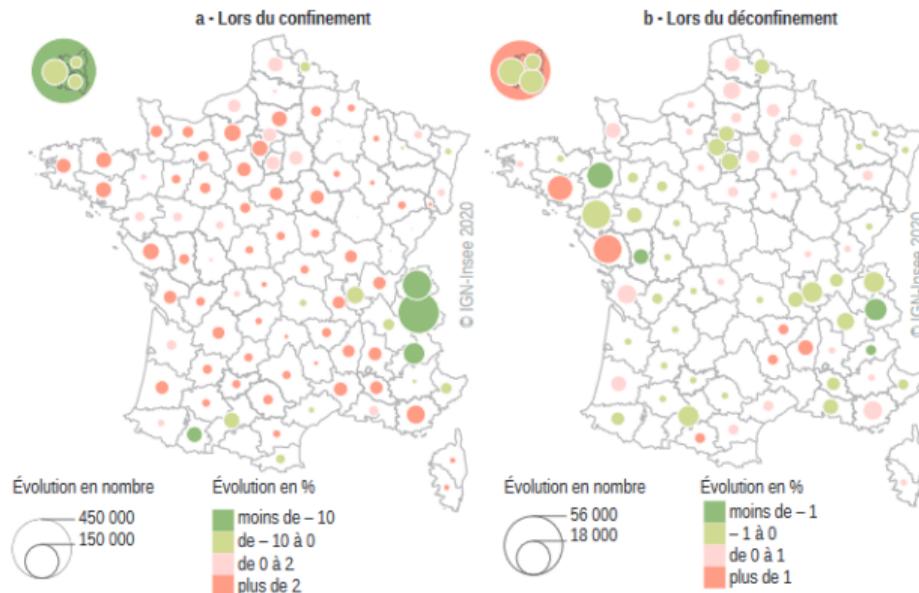
- ▶ Utilisation de données :
 - ▶ dont on ne contrôle pas la représentativité;
 - ▶ dont on ne contrôle pas le processus de production;
 - ▶ collaboration avec de nouveaux acteurs.
- ▶ Contrôle qualité de la donnée nécessaire :
 - ▶ Recalage avec données de recensement pour être cohérent
- ▶ Temps très contraint
 - ▶ Demande sociale forte;
 - ▶ Risque de mésinformation

Difficulté et limites

- ▶ Fixe une période de "référence", où la carte de présence est "habituelle"
 - ▶ Difficulté : vacances juste avant le confinement
- ▶ Etudie la distribution des présences pendant confinement et déconfinement par rapport à la référence
- ▶ Pas d'information sur les caractéristiques des porteurs de carte :
 - ▶ On fournit des pistes d'interprétations
 - ▶ On délègue certaines caractérisations aux opérateurs (exemple : lieu de résidence)
- ▶ Expérience sur données de téléphonie permettait de comprendre certaines limites des données

Valorisation principale

1 Évolution du nombre de métropolitains présents dans le département



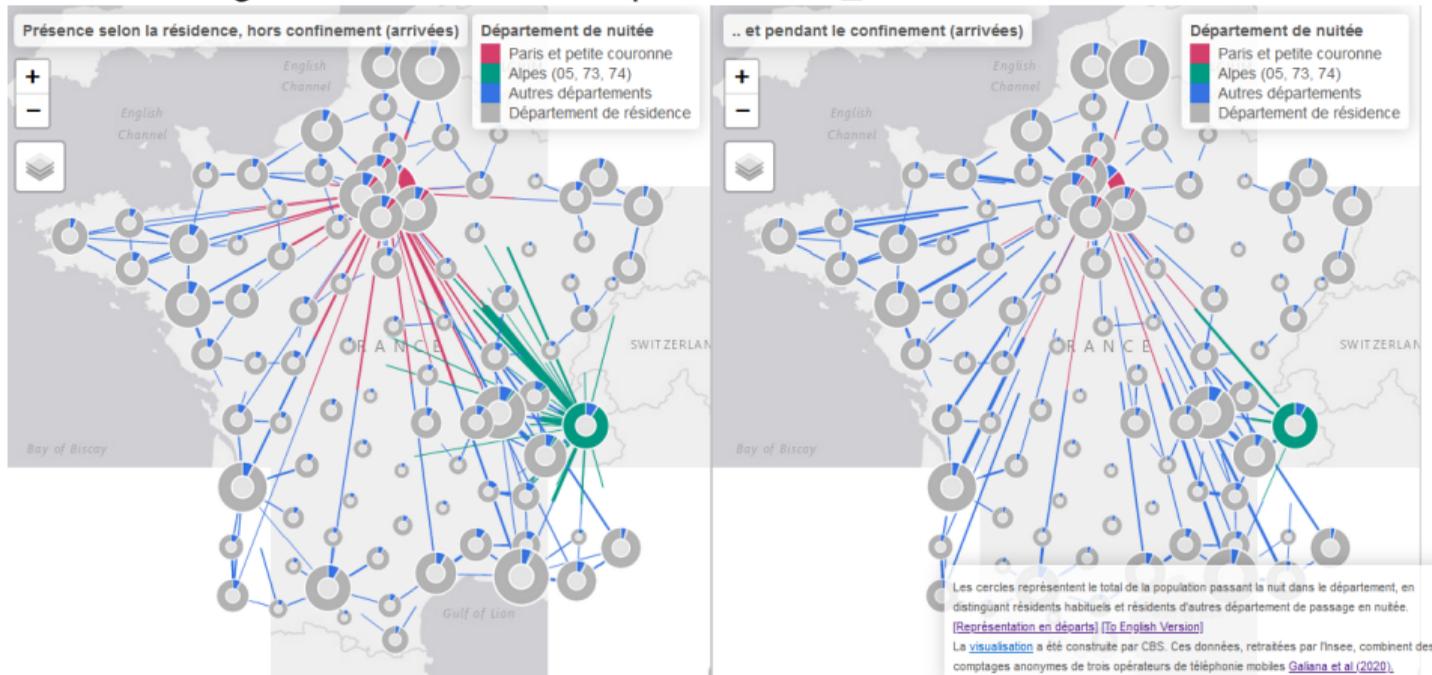
Note : les évolutions inférieures à 1 000 personnes ne sont pas représentées.

Lecture : le nombre de personnes présentes la nuit en Ardèche a augmenté de 19 000 avec le confinement, soit + 6 %. Au début du déconfinement, l'Ardèche compte 4 000 personnes de plus en moyenne que pendant le confinement, soit + 1 % par rapport à la population présente pendant le confinement.

Champ : période de référence hors confinement du 16 janvier au 17 mars 2020 au matin, période de confinement du 17 mars 2020 au soir au 11 mai 2020 au matin, période déconfinée du 11 mai au soir au 31 mai 2020 au matin.

Valorisations : visualisation interactive

https://inseefrlab.github.io/lockdown-maps-R/inflows_FR.html



Perspectives

Perspectives

Partenariat de recherche

- ▶ Collaboration en cours avec Orange et l'université Gustave Eiffel dans le but de travailler sur des données agrégées.

Travaux européens

- ▶ Subvention d'Eurostat pour la construction d'un protocole standard de traitement des données de téléphonie mobile.

Merci de votre attention

Marie-Pierre de Bellefon *marie-pierre.de-bellefon@insee.fr*

Lino Galiana *lino.galiana@insee.fr*

SSP Lab - Direction de la Méthodologie et de la Coordination Statistique Internationale