

# Géographie sociale de la téléphonie mobile en France

Éric Guichard, Anthony Busson, Thomas Begin

2022–2024

## Résumé

Pour quantifier d'éventuelles discriminations sociales ou spatiales en matière d'accès à la téléphonie mobile, nous utilisons deux sources : celle relative à la localisation des antennes téléphoniques et celle issue du découpage par l'Insee de la France habitée en carrés de 200 mètres de côté. Nous détaillons la méthodologie employée pour traiter ces données massives et leurs limites. Nous élaborons ensuite une modélisation spatiale centrée sur les humains pour mesurer la distance des ménages et des jeunes aux antennes : les ménages pauvres en sont plus proches que les autres ménages, alors que la classe d'âge n'a pas d'effet. Nous étudions enfin les évolutions diachroniques de la couverture des villes : les plus pauvres sont équipées légèrement avant les autres mais les écarts à moyen terme entre villes ne sont pas significatifs. Tout ceci prouve que les modèles simples privilégiant la géographie des équipements et les lois de Poisson peuvent être aisément améliorés avec d'autres statistiques et en intégrant pleinement des paramètres socio-démographiques, malgré leur caractère lacunaire sinon estompé.

## Abstract

In an effort to quantify possible social or spatial discrimination in access to mobile telephony, we use two sources : the one relating to the location of telephone antennas and the one derived from the INSEE's division of inhabited France into 200-meter squares. We detail the methodology used to process these massive data, and their limitations. We then develop a human-centered spatial model to measure the distance of households and young people from antennas : poor households are closer than other households, while age class has no effect. Finally, we study the diachronic evolution of the coverage of cities : the poorest cities are equipped slightly before the others, but the medium-term differences between cities are not significant. Our study, therefore, shows that simple models based on the geography of equipment and Poisson laws can be easily improved with other statistics and by fully integrating socio-demographic parameters, despite the fact that the latter are incomplete or even blurred.

**Mots clés** Antennes téléphoniques, aménagement du territoire, analyse spatiale, atlas électronique, habitat, télécommunications, ségrégation socio spatiale.

**Keywords** Mobile phone antennas, spatial analysis, electronic atlas, habitat, telecommunications, socio-spatial segregation.

## 1 Projet et hypothèses

Les supports de la téléphonie mobile sont rarement mis en regard avec des indicateurs socio-démographiques de la France<sup>1</sup>.

Par exemple, il est difficile de savoir si, une fois prises en compte les contraintes de l'Arcep (Autorité de régulation des communications électroniques, des postes et de la distribution de la presse, <https://www.arcep.fr>), la distribution spatiale des « antennes » téléphoniques privilégie ou non des quartiers populaires ou des classes d'âge.

Par ailleurs, nous voulions préciser les formes que pouvaient prendre les territoires numériques en France, au regard de la téléphonie mobile, en prolongeant donc les travaux de l'Arcep : <https://www.monreseauemobile.fr> et <https://maconnexioninternet.arcep.fr> [11]. Pour cela, nous avons réalisé l'atlas <http://barthes.enssib.fr/sociotel/antennes.html>, qui illustre et synthétise cet article tout en explorant les possibilités d'une écriture numérique poussée à ses limites. Il permet aux utilisateurs de choisir les données à cartographier (population, ménages pauvres, jeunes) pour chaque département de l'hexagone, offre des agrandissements particulièrement élevés et rend accessibles les informations relatives aux supports d'antennes et aux carrés de population.

Pour tester notre hypothèse de départ, nous avons utilisé deux types de sources.

## 2 Sources

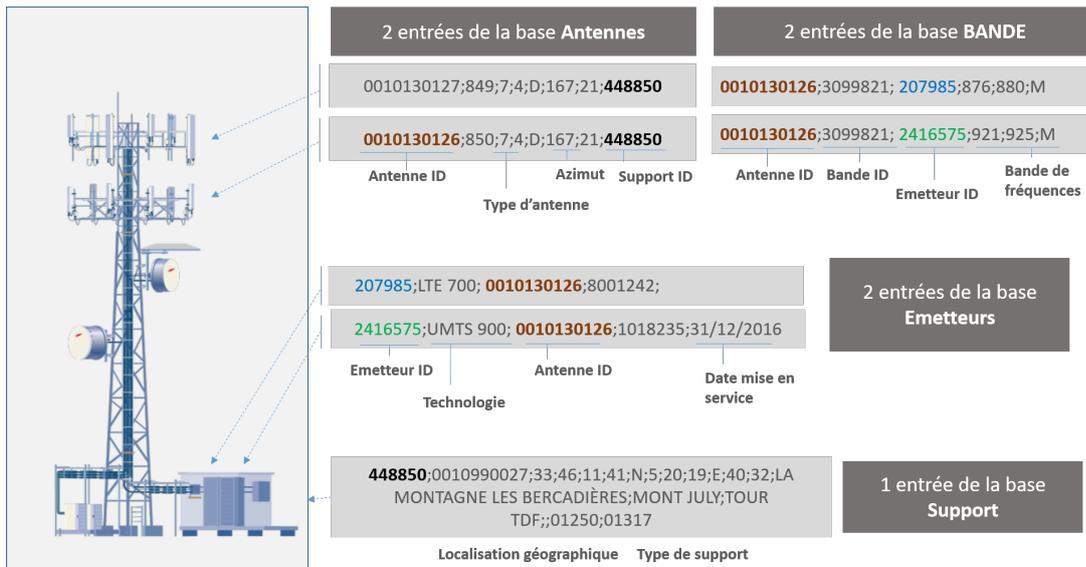
### 2.1 Téléphonie mobile

#### 2.1.1 Installations radioélectriques

La base de données ANFR (agence nationale des fréquences) répertorie toutes les installations radioélectriques de plus de 5 Watts en France. Mise à jour tous les mois, elle est constituée de 5 tables décrivant les antennes, les bandes de fréquences, les émetteurs, et les supports. Il s'agit de tables relationnelles utilisant des clés/identifiants qui permettent d'établir les relations entre les différents éléments, par exemple pour retrouver toutes les antennes d'un même support. Pour rappel, un support désigne l'infrastructure physique sur laquelle sont installés les équipements radios (antennes, émetteurs, etc.).

---

1. Nous remercions ici l'IPL *Better-Net*, fédéré par l'Inria et piloté par Isabelle Chrisment, qui a partiellement financé ce travail.



**Figure 1** – Extraits de 4 des 5 bases de données ANFR, qui précisent les différents éléments d'un même support ou d'une même station. Les liens entre les tables se font via les identifiants (antennes, émetteurs, bande, support) en couleur dans cet exemple.

La table « support » indique la localisation du support (adresse et coordonnées) ainsi que son type (château d'eau, pylône, mât béton, éolienne, etc.).

Les antennes sont les équipements qui émettent et reçoivent les ondes électromagnétiques (onde radio donc) et les transforment en ondes électriques transmises à l'émetteur (et réciproquement). Les communications entre les émetteurs d'un support et leurs antennes se font donc en filaire.

L'émetteur est l'équipement qui reçoit et envoie les données. Il les formate et les transmet à l'antenne (et inversement). L'émetteur est lié à une technologie (PMR, 3G/4G, FM, etc.). À noter qu'il faut interroger une autre base de données pour établir les correspondances entre les valeurs numériques utilisées dans la première base et leurs significations : numéro d'exploitant d'un opérateur, numéro de type d'antenne, etc.

Nous illustrons cette base de données de l'ANFR avec la figure 1. On repère un support d'identifiant (448450) dont l'entrée décrit ses propriétés physiques en particulier son type (la valeur 33 correspond à « Tour Hertzienne ») et sa localisation (coordonnées géographiques et adresses : N ;5 ;20 ;19 ;E ;40 ;32 ; « Montages les Bercadières », etc.). Nous identifions, grâce à la table « antennes », deux antennes installées sur ce support. Les entrées correspondantes décrivent entre autres leur type (7 correspond à « Antenne directive ») et leur azimut (167° ici pour les deux antennes). Dans la table « émetteurs », nous retrouvons deux émetteurs qui utilisent la même antenne. Ces deux émetteurs utilisent chacun une technologie spécifique : 3G (UMTS) et 4G (LTE). La table « bande » permet de retrouver les fréquences que chaque paire émetteur/antenne utilise : ici 876-880 MHz et 921-925 MHz.

La 5<sup>e</sup> table n'est pas représentée sur la figure 1. Elle décrit pour chaque antenne l'exploitant, les dates d'installation et de mise en service.

Ainsi, à partir de cette base de données de l'ANFR, il est possible de faire le lien entre la position géographique des supports, les générations de réseaux cellulaires, les dates de mises en service, et les exploitants.

En revanche, il n'est pas possible d'en déduire des paramètres de qualité de service précis, notamment en termes de débit. Cela dépend en pratique d'un grand nombre de paramètres inconnus de l'étude : les puissances d'émissions, les gains d'antennes, l'allocation de ressources radio faites à chaque utilisateur, etc. Au-delà de ces informations dont certaines peuvent être connues de l'opérateur, la qualité du lien radio et donc du service réseaux dépendra aussi de la localisation précise de l'utilisateur, de l'environnement géographique (montagne, forêt, etc.), des obstacles (murs, maisons, etc.).

Il existe quand même un invariant sur lesquels tous les ingénieurs s'accordent, qui est que la qualité des liens radios est en moyenne décroissante avec la distance entre l'utilisateur et son antenne.

### 2.1.2 Mise en pratique

Par la suite et sauf précision contraire, les mots « antenne », « émetteur » et « support » seront utilisés comme synonymes, puisqu'ils renvoient tous à une même localisation.

Au final, nous avons sélectionné, pour les émetteurs publics<sup>2</sup> leurs latitude, longitude, code postal, technologie, opérateur et date de mise en fonction : en avril 2020, sur environ 1 400 000 émetteurs<sup>3</sup>, on dénombrait 54 000 supports dédiés aux usages de la téléphonie grand public.

## 2.2 Socio-démographie de la France

L'autre source de données provient de l'Insee. Cet institut a découpé la France en carrés habités de 200 mètres de côté. Nous nous sommes limités à l'Hexagone et à la Corse, décrits par 2 278 216 tels carrés. Chaque carré est décrit par ses coordonnées et le nombre exact de ses habitants. D'autres variables, essentiellement démographiques sont récupérables *via* une autre base spatiale, composée de rectangles, dont la fonction est de garantir l'anonymat. Ces 698 659 rectangles (unions des carrés précités) sont composés d'au moins 11 ménages fiscaux<sup>4</sup> et leurs variables s'appuient sur les revenus fiscaux localisés et sur la taxe d'habitation des ménages. Un carré se confond avec le rectangle qui le contient dans les

---

2. Ces fichiers décrivent aussi des stations « virtuelles », au sens où elles peuvent être des regroupements de supports (ex. : parc militaire) ou non destinées au public (ex. : EdF).

3. Dont 307 868 pour Bouygues, 256 906 pour Free, 377 269 pour Orange, 311 766 pour SFR.

4. Le document « Diffusion de données carroyées – novembre 2013. Documentation complète sur les données carroyées à 200 mètres » précise cette méthodologie et les sources transmises. Cf. <https://www.insee.fr/fr/statistiques/2520034>.

zones relativement denses<sup>5</sup> ; sinon, les informations que l'on peut tirer d'un carré résultent d'une interpolation à partir de son rectangle<sup>6</sup>.

### 2.2.1 Variables de type démographique

Ces variables étaient tout d'abord : le nombre total d'individus, de ménages, de ménages de 5 personnes et plus et nombre d'enfants de 0 à 3 ans, de 4 à 5 ans, de 6 à 10 ans, de 11 à 14 ans, de 15 à 17 ans, de 25 ans et plus. S'y ajoutaient, parfois de façon interpolée pour éviter des repérages de personnes dans les zones peu denses, le total des individus de 65 ans et plus et de 75 ans et plus.

De ces variables, nous avons retenu les nombres d'individus et de ménages, leur taille moyenne, les nombres de personnes de moins de 15 ans, de 15 à 25 ans, d'adultes et de plus de 65 ans.

D'autres variables, comme le nombre de ménages en logement collectif, le nombre de ménages propriétaires ou la surface cumulée des résidences principales, ne nous ont pas semblé utiles.

### 2.2.2 Variables de type économique

Nous avons estimé les profils économiques des carrés à partir des autres variables disponibles : somme des revenus fiscaux par unité de consommation (UC) « winsorisés »<sup>7</sup> des individus et nombre de ménages dont le revenu fiscal par UC se situe en dessous du seuil de bas revenu.

Nous avons retenu de ces informations le nombre de ménages « pauvres », c'est-à-dire déclarant un revenu fiscal de moins de 7500 Euros.

## 2.3 Mise en correspondance des deux sources

Pour articuler ces variables socio-démographiques avec les données des antennes, nous avons converti<sup>8</sup> les coordonnées des carrés (décrites au format Lambert 2 étendu) au format des antennes : le WGS84.

Au final, nous avons conservé les 895 767 carrés de plus de 10 habitants, afin d'éviter de travailler sur des carrés aux données trop erratiques du fait de l'ano-

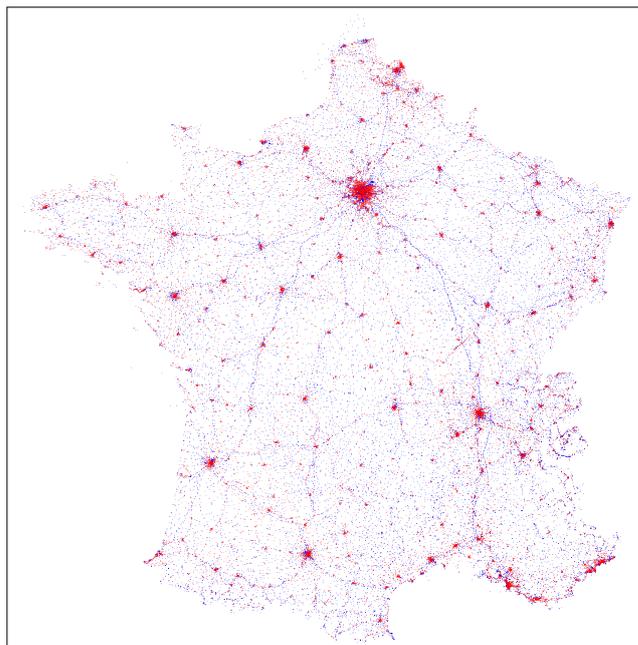
---

5. Si on considère qu'un ménage fiscal se compose de 2,3 personnes en moyenne (source : <https://www.insee.fr/fr/statistiques/2012717>), un carré peut se confondre avec le rectangle qui le contient quand il est composé d'environ 25 personnes, ce qui renvoie à une densité d'environ 600 habitants par km<sup>2</sup> — et d'au moins 275 par km<sup>2</sup> si tous les ménages du carré sont composés d'une seule personne.

6. Si un rectangle de 50 habitants contient 4 carrés dont un de 10 habitants, les informations estimées de ce dernier seront le cinquième de celles du rectangle (et non pas le quart).

7. La winsorisation est une méthode statistique de nivellement des données. Cf. <https://www.insee.fr/fr/statistiques/2520034>, déjà cité. Pour chaque ménage, on attribue 1 UC au premier adulte, 0,5 UC aux autres personnes de plus de 14 ans, 0,3 UC aux enfants. Le niveau de vie d'un ménage résulte alors de son revenu total divisé par son total d'UC. Cf. <https://www.insee.fr/fr/metadonnees/definition/c1802>.

8. Ce qui s'est avéré long : 130 heures de calcul.



**Figure 2** – Antennes de France métropolitaine. En rouge, les supports en des zones habitées (carrés de plus de 10 habitants). En bleu, les autres. Villes et axes de circulation apparaissent clairement.

mysation précitée. Il y a 577 660 carrés de plus de 20 habitants (seuil à partir duquel un carré commence à se confondre avec le rectangle qui le décrit), 277 465 carrés de plus de 50 habitants et 130 007 carrés de plus de 100 habitants. Nous avons estimé qu’une zone pavillonnaire avec des jardins de 1000 m<sup>2</sup> inscrite dans un tel carré contient environ 20 ménages (en considérant que la voirie et l’espace public attenants constituent la moitié de la surface) et correspond donc à une population de 80 habitants. Le rectangle-carré le plus habité a 4384 habitants, ce qui correspond à une densité (rare) de 100 000 habitants au km<sup>2</sup>.

Rappelons que nous disposons aussi de 54 000 supports de téléphonie mobile.

Nous avons ensuite vérifié si l’un des 54 000 points-supports était ou non dans l’un des 900 000 carrés, puis repéré les supports en zone habitée et ceux qui ne l’étaient pas : un sur deux est au milieu des champs, principalement le long des principales voies de circulation. Cf. figure 2.

## 2.4 Remarques

Avant de décrire nos résultats, précisons quelques points et perspectives.

1. Les données socio-démographiques de l’Insee et des carrés associés datent de 2010–2011 et ont été publiées fin 2013. Depuis, l’Insee a publié une base plus récente<sup>9</sup>. Il était tentant d’utiliser cette dernière. Pour autant,

9. <https://www.insee.fr/fr/statistiques/4176290?sommaire=4176305>.

les équipements de téléphonie mobile ayant majoritairement été déployés autour des années 2010, c'est bien la première qui est la plus pertinente.

2. Nous mettons ce déploiement d'antennes en correspondance avec les habitants (individus, ménages). Nous n'avons pas tenu compte de la géographie du tertiaire (bureaux, etc.), dont l'influence sur les stratégies des opérateurs n'est pas négligeable, par exemple dans les quartiers dédiés comme les tours de la Défense près de Paris.
3. Notre récolte des données peut se poursuivre et il sera aisé d'étendre notre enquête à la 5G.

## 3 Premiers résultats

### 3.1 Limites du *big data*

Les 895 767 carrés de plus de 10 habitants concernent 56 628 478 habitants. Or il y en a 67 millions en France (en négligeant la différence entre 2010 et 2018). Manquent donc environ 11 millions. Après avoir enlevé des 67 millions d'habitants du dernier recensement ceux de Martinique, de Guadeloupe, de Guyane, de la Réunion et de Mayotte, il aurait dû en rester 64,9 millions. Or les 698 660 rectangles de plus de 11 ménages fiscaux de l'Insee (France métropolitaine + Corse) concernent 61 914 632 de personnes. Trois millions de Français disparaissent. Après vérification, ces trois millions de personnes résident en des zones quasiment vides<sup>10</sup>. Aussi, notre enquête sur la socio-démographie fine de la France, telle qu'elle nous est permise par l'Insee sur la base des carrés, masque finalement 8,3 millions de personnes : 3 millions du fait du choix direct de l'Insee (rectangles) et 5,3 millions de notre fait, mais en corollaire de ce choix : la non prise en compte des carrés de moins de 10 habitants, pour lesquels nous ne voulions pas énoncer d'extrapolations abusives<sup>11</sup> ; soit 12,7% de la population.

Preuve que le *big data*, articulé à des questions éthiques (préservation légitime de l'anonymat) induit beaucoup d'imprécisions, malgré les espoirs de certains de ses promoteurs, qui voudraient qu'il nous permette d'accéder à des informations très fines. Comme le fait remarquer D. Pumain, la question des données massives ne se réduit pas à celle de la puissance des ordinateurs [24]. Ces données sont construites [16], approximatives et leur grande précision peut s'accompagner de grosses lacunes, comme nous le voyons ici.

### 3.2 Des classes sociales estompées

Il n'est pas aisé de repérer des équivalents de classes sociales *via* les données de l'Insee : l'institut rabat dans sa base les 9<sup>e</sup> et 10<sup>e</sup> déciles des revenus fiscaux

---

10. Façon de redécouvrir la réalité rurale de la France : plus de 9 millions de personnes vivent dans des rectangles (donc plus étendus que nos carrés de 200 m de côté) de moins de 40 habitants.

11. Pour information, 5 286 520 personnes vivent dans des carrés de 9 habitants ou moins.

des ménages sur le 8<sup>e</sup>; ce qu'explique l'institut (logique de « winsorisation » — cf. note 7) et ce que l'on découvre en calculant le revenu moyen des individus par carré, qui ne dépasse jamais 29 244 Euros<sup>12</sup>. Aussi les classes supérieures sont-elles estompées<sup>13</sup>.

De façon analogue, ce revenu des ménages n'est jamais en dessous de 7 500 Euros. Pourtant, l'Insee dénombre, pour chaque carré, le « nombre total de ménages dont le revenu fiscal par unité de consommation se situe en dessous du seuil de bas revenu » (moins de 11 250 Euros par an) mais précise que cette variable ne correspond pas vraiment à un indicateur de pauvreté (cf. <https://www.insee.fr/fr/statistiques/fichier/2520034/documentation-complete-donnees-a-200m-1.pdf>, p. 11). L'institut précise que ce revenu fiscal n'inclut pas les « revenus sociaux non déclarés (minima sociaux tels que RMI ou RSA et minimum vieillesse, prestations familiales, aides au logement) ».

En bref, les riches et les très pauvres nous échappent. Ce que nous voyons avec la carte 3, qui décrit, pour les 4 départements de la couronne parisienne, les carrés des plus hauts revenus, ceux des plus bas et ceux contenant plus de 25% de ménages aux plus bas revenus fiscaux; et il n'y a que 73 carrés parisiens dont le revenu fiscal moyen des ménages excède 26 000 Euros. Ces statistiques ne reflètent donc pas la réalité économique des ménages des quartiers riches de Paris (carte 4).

### 3.3 Point méthodologique

Notre espoir de mesurer d'éventuelles discriminations sociales ou spatiales en matière d'accès à la téléphonie mobile fut donc freiné pour les raisons précitées, que nous synthétisons :

- pas de données sur la puissance ou la couverture des antennes;
- pas de prise en compte de la géographie du tertiaire;
- disparition de 12,7% de la population du fait direct ou indirect de l'Insee (zones rurales);
- estompement des classes sociales par l'Insee.

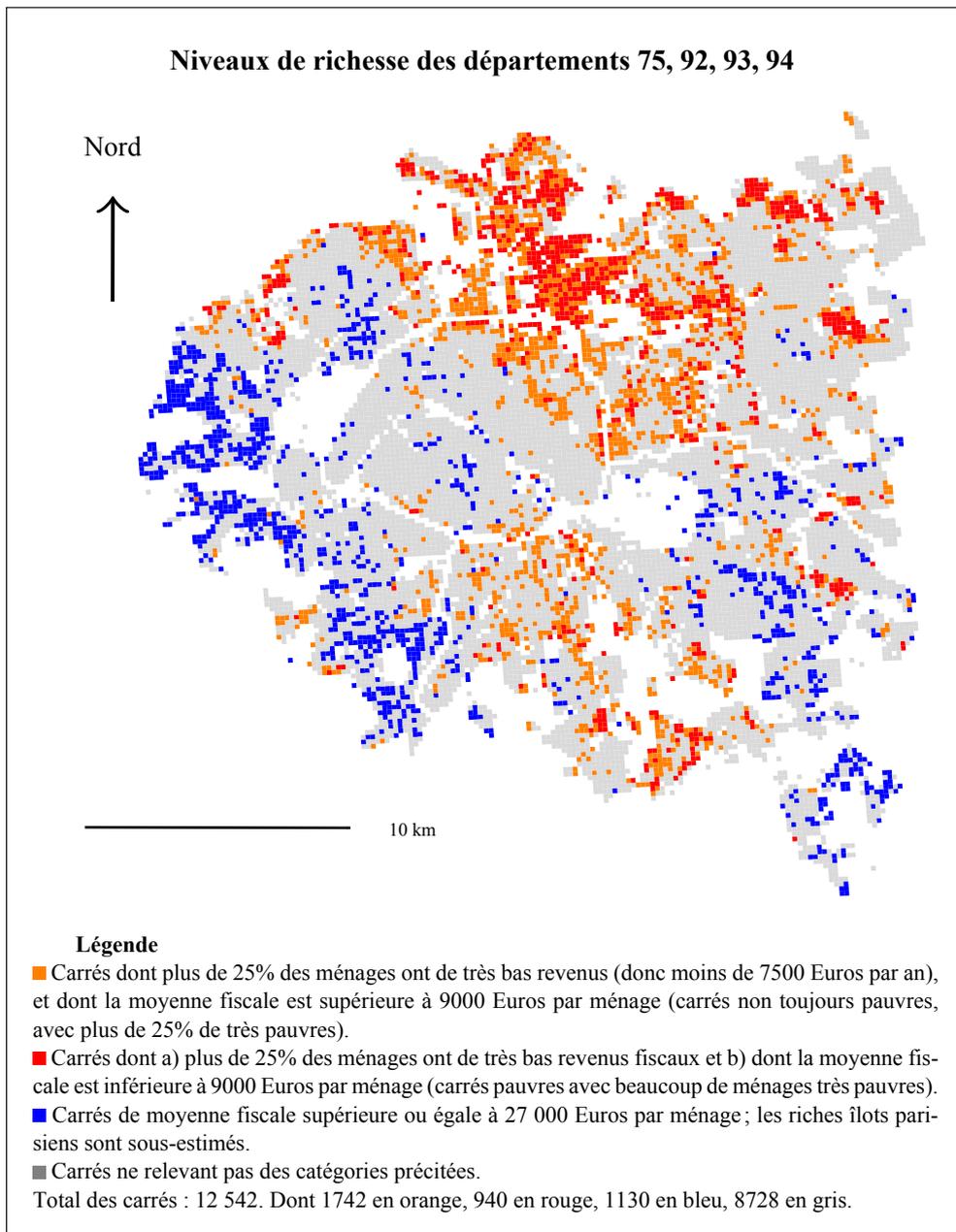
Nous avons néanmoins décidé de poursuivre notre étude, car

- elle reste pertinente, à condition de garder en mémoire les contraintes précédentes : de spécifier précisément les conditions de validité de nos conclusions;
- elle garde son intérêt premier : conjuguer les méthodes usuelles de la modélisation réseaux avec des données sociales, ne serait-ce que pour montrer aux industriels et au monde de la recherche que de tels croisements sont féconds.

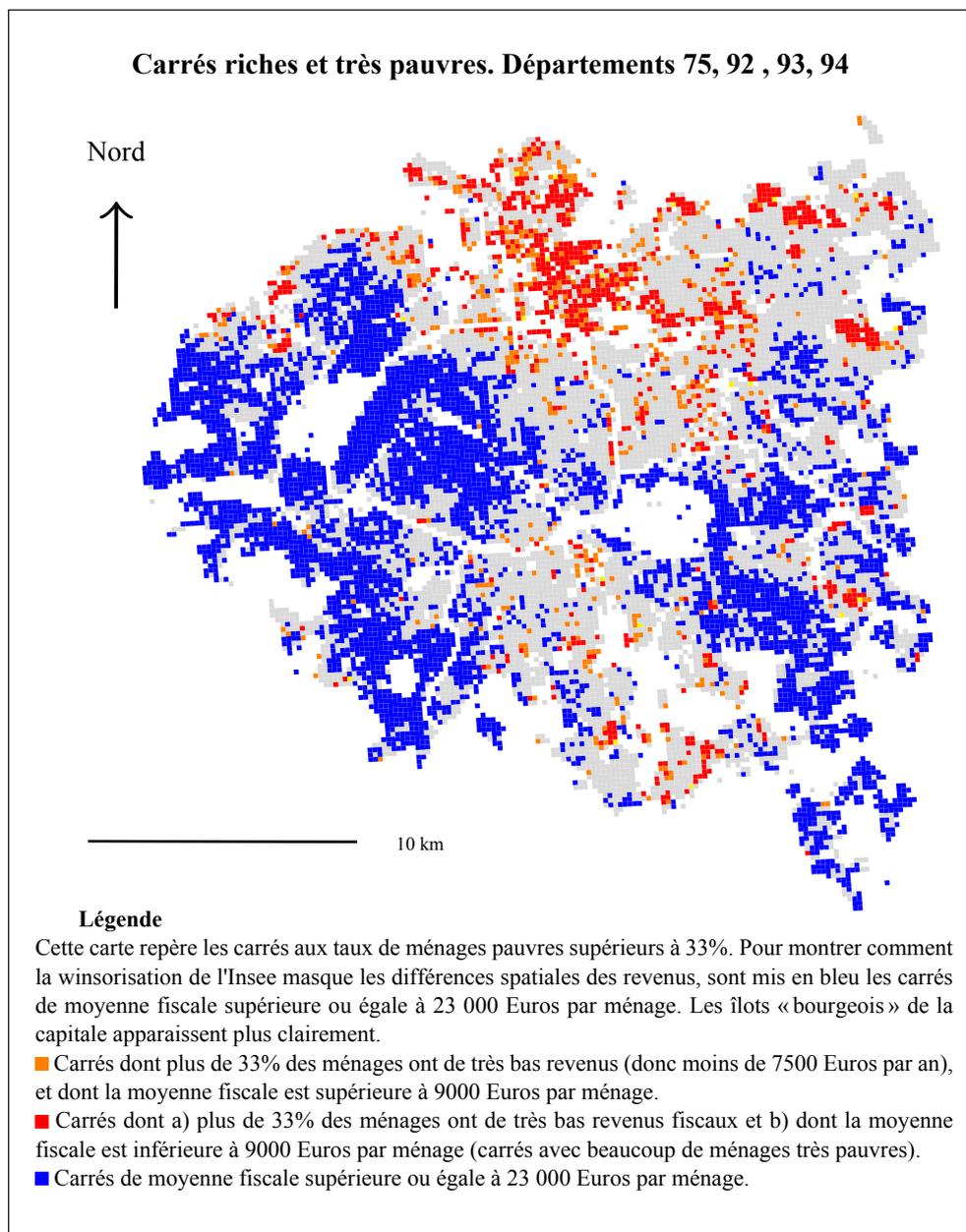
---

12. Alors que le 9<sup>e</sup> décile des revenus individuels commence à 33 000 Euros, et le 10<sup>e</sup> à 57 000 Euros. Source : <https://www.insee.fr/fr/statistiques/2417897> pour 2018.

13. Un cadre célibataire majeur, qui se confond donc avec un ménage, voit donc son revenu limité à 2500 Euros par mois.



**Figure 3** – Niveaux de richesse des carrés en région parisienne.



**Figure 4** – Carrés riches et très pauvres.

### 3.4 Repérer les ménages pauvres

Pour autant, les carrés où le taux de ménages à très bas revenu fiscal (en orange) excède 25% permettent de repérer des territoires explicitement connus comme pauvres : quartiers Nord de Paris (malgré la gentrification des anciens quartiers populaires), abords du périphérique, proche banlieue. Pour préciser ce fait, nous avons mis en rouge les carrés qui ont cette première propriété et qui ont une très basse moyenne de revenus fiscaux (moyenne des ménages du carré inférieure à 9 000 Euros par an). Nous repérons certaines zones d'Aubervilliers, La Courneuve, Montfermeil, Villepinte, etc.

Ainsi avons-nous choisi de considérer ces ménages « à très bas revenus fiscaux hors prestations sociales » comme des « ménages pauvres », voire très pauvres, malgré les recommandations de l'Insee. Et, dans la suite de l'article, nous appellerons ménages pauvres ces ménages à très bas revenus.

## 4 Apports de la modélisation

### 4.1 La distance des ménages aux antennes

#### 4.1.1 État de l'art

La distance des ménages à leur antenne la plus proche est un des paramètres clés permettant d'évaluer la couverture radio et sa qualité.

La modélisation spatiale des réseaux a débuté à la fin des années 1990, avec des modèles stochastiques qui intègrent les positions géographiques des points de présence des opérateurs [3, 4]. L'idée première de ces études était d'évaluer le coût des déploiements filaires à grande échelle.

La modélisation spatiale a ensuite été appliquée aux réseaux sans fil où les points d'un processus ponctuel représentent les positions géographiques des stations de bases [1, 10, 13]. Les métriques déduites de ces modèles étaient alors relativement simples ce qui permettait de les exprimer sous forme de formules closes. Les modèles décrivant les ressources radio se sont ensuite enrichis tendant à être de plus en plus réalistes [18, 12, 15, 19, 7].

Dans les articles ici précités, la modélisation réseau (canal radio, allocation des ressources) s'est donc complexifiée mais les processus ponctuels décrivant les positions géographiques sont restés simples. Ils utilisent des processus de Poisson. Le choix du processus de Poisson est un choix pragmatique car il présente des propriétés qui simplifient les calculs. Il existe des exceptions avec des modèles spatiaux différents mais pour lesquels les calculs se complexifient nettement [26, 27] rendant l'outil théorique moins attractif.

En pratique, il existe peu ou pas d'études à grande échelle sur la distribution spatiale réelle des équipements réseaux et en particulier des stations de base.

Notons également, que la qualité des liens radios ne dépend pas de la position des stations de base seule mais des positions relatives stations de base / utilisateurs. Pour ces derniers, les statistiques sont encore plus rares. Il existe des travaux sur

la mobilité des utilisateurs [5, 22], mais à notre connaissance, rien sur la position réelle des utilisateurs des réseaux cellulaires. Dans la grande majorité des études citées précédemment, un processus de Poisson indépendant de celui des stations de base est utilisé.

Notons enfin que les distributions des stations de base et des utilisateurs sont très probablement corrélées, d'où la nécessité d'une étude conjointe. Ce que nous proposons dans cet article.

#### 4.1.2 Notre méthode

Nous ne distinguons pas ici les générations d'antennes (3G ou 4G par exemple) ni les opérateurs.

Les carrés Insee sont divisés en quatre sous-carrés de tailles égales (appelés sous-carrés dans la suite) afin d'affiner ce calcul des distances, en particulier pour les zones denses où plusieurs antennes peuvent se trouver dans un même carré Insee.

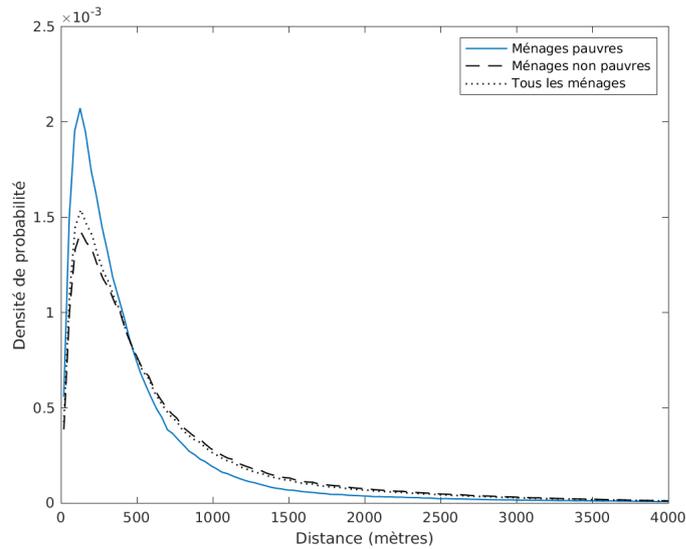
En pratique, nous repérons les cellules de Voronoï formées par les antennes dans lesquelles se trouvent les centres des sous-carrés Insee. Comme nous ne connaissons pas les localisations exactes des ménages au sein de leur carré ou sous-carré, nous les confondons avec le centre de ce sous-carré. Les ménages d'un carré sont alors répartis de manière homogène dans ces quatre sous-carrés. Cela signifie qu'un utilisateur est rattaché/associé au support qui lui est le plus proche.

Quand nous calculons le pavage de Voronoï associé à la position des supports, cela revient à rattacher tous les utilisateurs au support de la cellule de Voronoï dans laquelle ils se trouvent. Dans nos différentes statistiques, nous générons un échantillon pour chaque utilisateur. Le nombre d'utilisateurs dans une cellule de Voronoï donnée pondère donc la statistique : un support sera alors plus représenté si plus d'utilisateurs lui sont rattachés. Les statistiques obtenues seraient différentes si un seul échantillon par support avait été pris en compte. Dans ce cas, chaque support serait représenté avec un poids égal, même s'il a très peu d'utilisateurs (ce serait donc moins pertinent). Énoncée de façon simple, la question à laquelle nos statistiques répondent est « quelle est ma distance à l'antenne si je suis un ménage tiré au hasard ? » et non pas un point géographique tiré au hasard sur le territoire ou dans un des carrés Insee.

Les positions des antennes-supports sont, nous l'avons vu, connues de façon très précise, en deçà du mètre. Nous avons dénombré 23 801 292 ménages distribués en nos 900 000 carrés devenus environ 3 600 000 petits carrés de 100 mètres de côté.

#### 4.1.3 Distance aux antennes : ménages pauvres et autres

La figure 5 affiche les densités de probabilité empirique des distances des ménages pauvres (avec notre définition), non pauvres (les autres), et de tous les ménages confondus. Les trois distributions montrent un pic à quelques centaines de mètres avant d'entamer une décroissance assez prononcée. Une différence de distribution entre les ménages pauvres et non pauvres apparaît clairement, avec



**Figure 5** – Distributions empiriques de la distance des ménages à leur antenne la plus proche.

	Moyenne	Médiane
Ménages pauvres	532m	317m
Ménages non pauvres	735m	452m
Tous les ménages	698m	422m

**Table 1** – Moyennes et médianes des distances ménage-antenne.

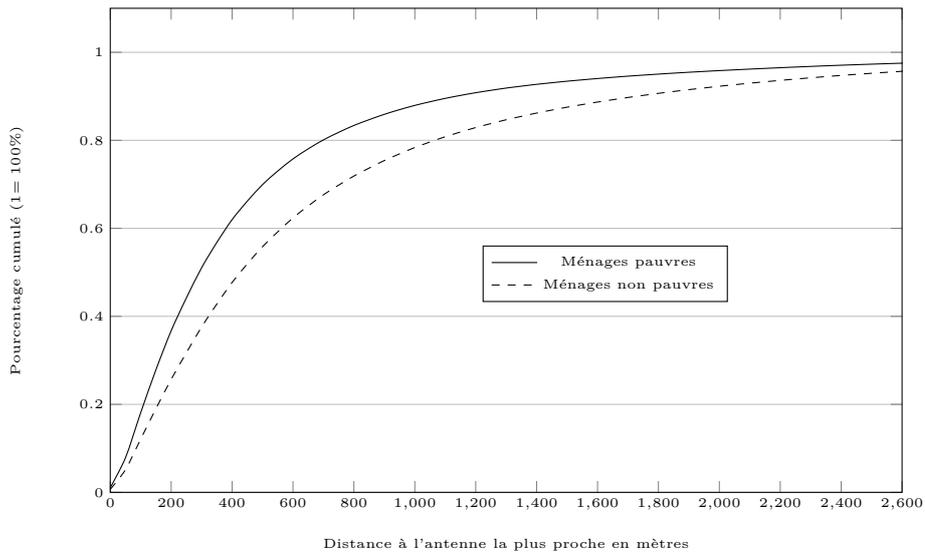
un pic supérieur pour les ménages pauvres suivi d’une décroissance plus rapide.

Les ménages pauvres sont donc distribués plus près de leurs antennes que les autres ménages. Cela apparaît clairement dans la table 1. La moyenne pour les ménages pauvres est de 203 mètres inférieure à celle des ménages non pauvres. On observe aussi un écart important entre médiane et moyenne qui s’explique par la décroissance lente des distributions empiriques. Il est difficile d’expliquer cette différence de distance entre les classes de ménages sur la base de ces statistiques et des données traitées. (cf. point 3.3).

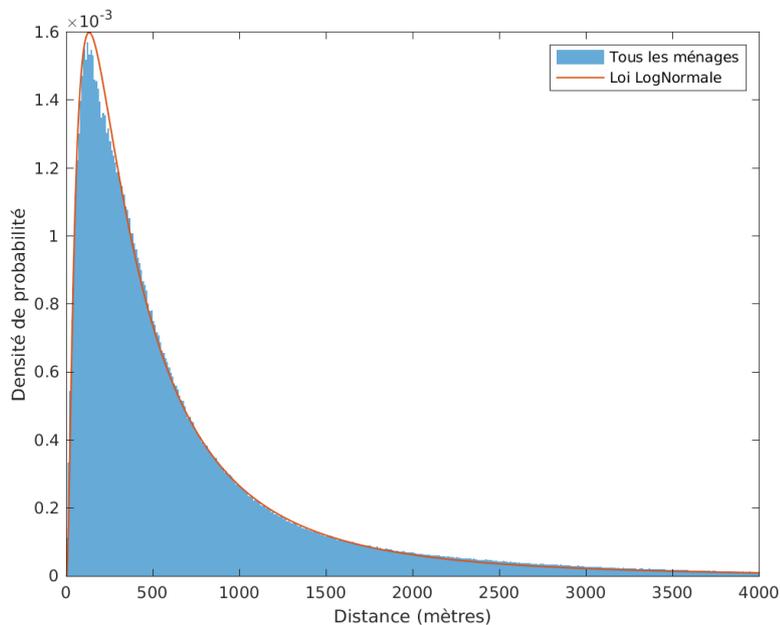
Cet écart d’environ 200 mètres entre ménages pauvres et autres ménages semble se maintenir, quelle que soit la distance du ménage au support, une fois le pic dépassé ; ce que montre le graphique 6, qui traduit de façon plus intuitive le résultat de la figure 5. Avant de tenter une conclusion, nous allons fouiller la nature de cette statistique.

#### 4.1.4 Extrapolation de la distribution empirique

La puissance du signal en réception étant décroissante en moyenne avec la distance, la distance entre l’utilisateur et l’antenne (ou son support) est un facteur



**Figure 6** – Histogramme cumulé normalisé des ménages pauvres et non pauvres. Quatre points par carré. Abscisse : distance à  $x$  mètres d’(au moins) une antenne. Ordonnée : pourcentage cumulé de la catégorie sociale (1=100%). Le graphique se lit ainsi : quand 60% des ménages pauvres sont à moins de 400 m de la plus proche antenne, 60% des non pauvres sont à moins de 600 de la plus proche.



**Figure 7** – Distribution de la distance des ménages à l’antenne la plus proche : loi empirique et log-normale.

clé de la qualité des communications<sup>14</sup>. Un certain nombre d'études [2, 20, 17] s'appuie sur une modélisation aléatoire de la distance utilisateur-antenne pour évaluer les systèmes de communications. La distribution choisie a un fort impact sur les performances du système et doit refléter la distribution réelle du terrain.

Aussi devons-nous comparer cette distribution empirique de la distance de nos 23 801 292 ménages aux antennes à diverses distributions classiques, comme les suivantes ; normale, log-normale, de Weibull, exponentielle et Gamma. Nous rajoutons à celles-ci la distribution de densité  $f(x) = 2\lambda\pi x e^{-\lambda\pi x^2}$  pour  $x > 0$ , qui correspond à la distribution de la distance lorsque le processus décrivant les antennes est un processus de Poisson homogène de taux  $\lambda$ . Cette distribution est pertinente car de très nombreux articles [9, 8, 21] supposent une distribution poissonnienne des antennes.

Pourtant, tous les tests d'hypothèses Smirnov-Kolmogorov [25] rejettent la concordance entre la distribution empirique et ces distributions théoriques. Aussi, afin de sélectionner la distribution optimale, avons-nous calculé la norme  $L^1$  entre la distribution empirique et chacune d'entre elles<sup>15</sup>. Il s'ensuit que la loi log-normale de paramètres (6.02, 1.06) offre la meilleure approximation avec une erreur au moins deux fois inférieure aux autres distributions. La figure 7 illustre cette grande proximité entre la loi log-normale et la distribution empirique.

Étonnamment, la distribution issue du modèle poissonnien offre les pires résultats, avec une erreur significative (environ 7 fois supérieure à celle de la loi log-normale). Ceci prouve que le modèle poissonnien n'est pas le plus adapté pour modéliser la distance ménage-antenne.

Il nous semble important de rappeler qu'il s'agit ici de la distribution ménage-antenne, qui est différente de la distribution carré Insee-antenne, dont nous avons pu vérifier qu'elle est au mieux approximée par une loi Gamma.

#### 4.1.5 Première conclusion

La littérature nous confirme que la majorité des distributions de distances entre lieux d'habitation et un service donné suivent une telle distribution log-normale [23]. Ce constat, qui ne peut alimenter de causalité, a l'avantage de nous rappeler que les services supposés immatériels du numérique et les services matériels (la station-service ou l'épicerie de quartier) ont le même comportement statistique.

Il n'est pas aisé d'expliquer le résultat relatif aux ménages très pauvres (c'est-à-dire dont le revenu fiscal est inférieur à 7500 Euros, soit l'équivalent mensuel de 625 Euros), plus proches des émetteurs de téléphonie mobile que les autres ménages. Déjà, la disparition de nombreux ménages vivant en des zones rurales ou semi-désertes nous incite à la prudence. Tout au plus, pouvons-nous remarquer

---

14. Sachant que cette distance optimale varie fort selon l'espace : dans une vieille ville, une antenne ne porte pas à plus de 400 mètres, quand elle peut avoir en pleine campagne un rayonnement de plus de 10 km.

15. Elle mesure une distance entre théorie et empirie, qui correspond à l'intégrale de la valeur absolue entre les deux distributions.

que les débats sur les questions sanitaires relatifs à la téléphonie mobile n'ont pas alimenté de controverse intégrant la spatialité; et qu'à l'inverse des associations de préservation du patrimoine ont un périmètre local et peuvent exprimer, de la part de groupes sociaux aisés, la volonté de préserver une esthétique paysagère quand les plus pauvres restent relativement discrets en matière d'exigences environnementales. Cette hypothèse reste controversée : en Nouvelle Zélande, le prix de l'immobilier peut chuter de 20% du fait de la proximité d'un mât téléphonique. À Hamburg, il peut chuter de 2,3%; encore faut-il que les habitations soient à moins de 100 m d'un support autonome<sup>16</sup> et situées en périphérie [6].

Une autre hypothèse s'appuie sur le fait que les ménages pauvres habitent en des zones denses (cf. <https://www.insee.fr/fr/statistiques/4508514>) et que ces zones sont celles où les antennes sont les plus nombreuses.

Comme pour le cas de la loi log-normale, nous nous limiterons à un constat, en évitant les conclusions sociologiques ou politiques.

## 4.2 Proximité des jeunes aux antennes

Notre seconde hypothèse de départ était que les opérateurs aient pu privilégier les plus grands consommateurs d'appels, de SMS et de bande passante : les jeunes adultes et les adolescents autonomes. Il n'en est rien (peut-être du fait de l'existence de l'Arcep), comme le prouve le graphique 8, établi sur la base d'une distinction entre les 15–25 ans et les plus de 25 ans. Dans les faits, les « jeunes » ne sont même pas plus proches des antennes que les plus de 65 ans (dont l'usage des téléphones portable semble modéré), comme le montre le même graphique. Nous ne commenterons pas plus ce fait, en nous contentant de confirmer la non-pertinence de la classe d'âge (ou d'une discrimination générationnelle) en matière d'accès téléphonique ou de stratégie des opérateurs.

## 4.3 Couverture des villes

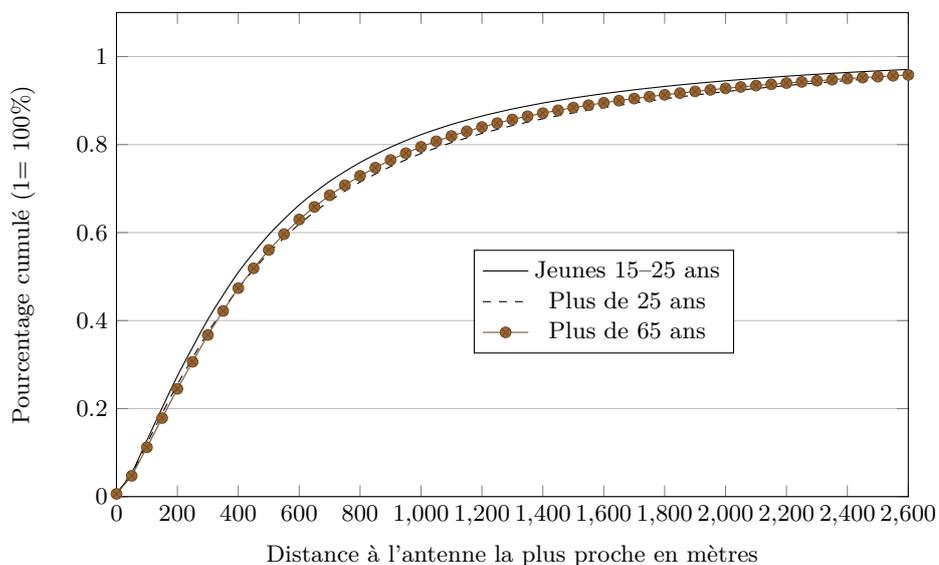
Ces premiers travaux nous ont incité à donner une dimension diachronique à l'enquête. Pour signaler cette dynamique temporelle avec un exemple, nous avons étudié l'évolution de l'équipement des communes en émetteurs 4G (dans cette partie, nous distinguons émetteurs, porteurs d'une « technologie », et antennes).

### 4.3.1 Taille des communes

Nous avons pour cela calculé le taux de couverture atteint (en supposant une couverture de 100% en juin 2020) mois après mois pour l'ensemble des communes en métropole. La table 2 présente la durée moyenne pour les communes selon leur taille pour atteindre un taux de couverture de 25%, 50% et 75%.

---

16. Comme un mât. Les auteurs précisent que cette faible différence ne vaut pas dans les quartiers denses où les antennes sont installées en haut des immeubles. Ils remarquent que les analyses relatives à la Nouvelle Zélande négligent de nombreux paramètres.



**Figure 8** – Histogrammes cumulés de la distance à la plus proche antenne des 15-25 ans (jeunes), plus de 25 ans et plus de 65 ans. 4 points par carré. En faisant l’hypothèse que les moins de 15 ans n’ont pas l’autonomie de leurs aînés en matière de pratiques téléphoniques, ils n’ont pas été pris en compte. Abscisse : distance à x mètres d’(au moins) une antenne. On ne retrouve pas les différences significatives rencontrées entre ménages pauvres et non pauvres.

Taille des communes (en habitants)	Taux de couverture (émetteurs 4G)		
	25%	50%	75%
De 2 000 à 10 000	4,9 ans	6 ans	7 ans
De 10 000 à 50 000	4,4 ans	5,8 ans	6,9 ans
Plus de 50 000	4,2 ans	5,7 ans	6,8 ans
<b>Plus de 2 000</b>	4,8 ans	5,9 ans	7 ans

**Table 2** – Durée moyenne pour qu’une commune atteigne un taux de couverture de 25%, 50% ou 75%. Pour simplifier la lecture, les mois sont exprimés en fractions décimales d’années : 1 mois = 0,08 an ; 6 mois = 0,5 an, etc.

Catégorie de communes	Taux de couverture (émetteurs 4G)		
	25%	50%	75%
8 villes les plus peuplées	2,9 ans	5,7 ans	6,8 ans
8 villes les plus riches	4,8 ans	5,9 ans	7 ans
8 villes les plus pauvres	3,8 ans	5,3 ans	6,75 ans
8 villes les plus jeunes	3,3 ans	6,25 ans	7,25 ans
8 villes les plus âgées	5,9 ans	6,6 ans	7,3 ans
Rappel villes de plus de 2000 hab.	4,8 ans	5,9 ans	7 ans

**Table 3** – Durée moyenne pour atteindre un taux de couverture de 25, 50 ou 75% pour des catégories de communes. Pour simplifier la lecture, les mois sont exprimés en fractions décimales d’années : 7 mois = 0,6 an, etc.

Il apparaît qu’à partir du lancement du déploiement de la 4G (fin 2011), il aura fallu en moyenne 4 ans et 10 mois pour que les communes atteignent une couverture de 25% ; 5 ans et 11 mois pour atteindre les 50% et 7 ans pour atteindre les 75%. Le rythme de déploiement est assez rapide (25% de plus tous les 13 mois) et assez régulier, une fois passée la période initiale de ce déploiement (le seuil 0–25%).

Dans ce tableau 2, les villages ne sont pas pris en compte : d’une part puisque ils accueillent de nombreuses antennes (cf. carte 2) et d’autre part car les contraintes du bâti en matière de rayonnement des antennes ne sont pas les mêmes que dans les villes (cf. note 14). Ce même tableau aussi montre que les communes de petite taille ont atteint leur taux de couverture de 25%, 50% et 100% en moyenne 9 mois plus tard que les communes de plus grande taille. La différence reste assez faible, ce qui confirme la fragilité de la notion de fracture numérique [14], fut-elle géographique.

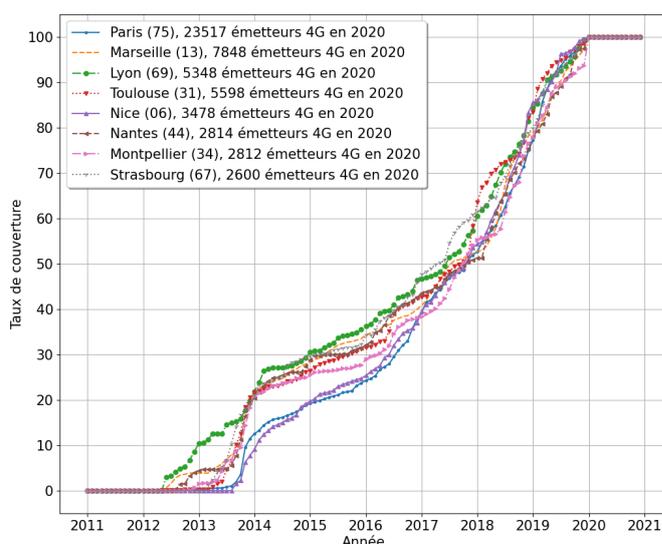
En calculant les écart-types à 25% (écart-type de 11 mois), 50% (9 mois) et 75% de couverture (7 mois), nous pouvons assurer qu’une ville est atypique si elle atteint ce seuil de 25% en moins de 3 ans et 11 mois (moyenne - écart-type) ou en plus de 5 ans et 9 mois (moyenne + écart-type). C’est le cas de Lyon (ville pilote 4G pour Bouygues dès 2012) qui atteint ce taux de couverture à 25% 3 ans et 2 mois après janvier 2011.

### 4.3.2 Quelques cas emblématiques

Nous complétons ce comportement « moyen » de l’évolution de la couverture 4G pour les communes importantes par quelques exemples précis afin de vérifier l’influence éventuelle de certains facteurs quantitatifs comme leur richesse, leur taux de pauvreté, et l’âge de leurs habitants<sup>17</sup>.

Nous considérons tout d’abord les 8 villes les plus peuplées. L’évolution de

17. Cf. <http://www.linternaute.com/ville/> et <http://www.linternaute.com/ville/classement/villes/richeesse>, dont les données proviennent de la Direction générale des finances publiques du Ministère de l’Économie et de l’Insee.



**Figure 9** – Évolution mensuelle du déploiement de la 4G pour les 8 villes les plus peuplées entre 2012 et 2020.

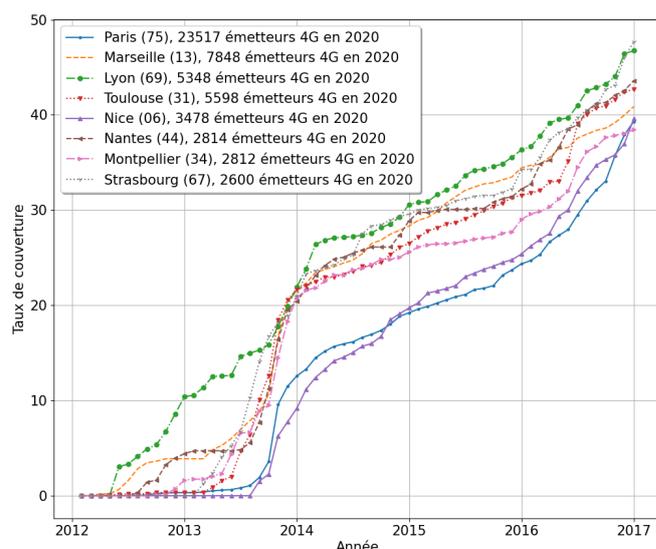
leur taux de couverture 4G est présentée dans le tableau 3. La comparaison de ce tableau avec le tableau 2 montre que les grandes villes ont bénéficié d'un déploiement initial plus rapide mais que cet écart s'est ensuite estompé. Cet écart est presque de 2 ans au seuil des 25% de couverture. Les figures 9 et 10 permettent de représenter ces résultats avec plus de détails. Nous avons remarqué que Lyon a bénéficié d'un déploiement plus rapide que les autres communes. Nous voyons que Marseille, choisie comme ville-pilote par Orange, n'est pas pour autant avantagée : son écart aux autres villes est bien plus faible que celui relatif à Lyon (cf. figure 10). Enfin, nous observons également que le déploiement a été le plus rapide sur deux périodes : entre 2013 et 2014 et entre 2018 et 2020.

Le tableau 3 indique également les évolutions de la couverture 4G selon les autres critères précités. En plus des communes les plus peuplées, les 8 communes les plus jeunes et les 8 les plus pauvres ont généralement atteint plus rapidement ce taux de couverture de 25% que les 8 communes les plus riches et les 8 les plus âgées. Toutefois ces écarts s'atténuent avec le temps et deviennent plus réduits quand sont atteints les 75% de couverture (aux environs de l'été 2018).

### 4.3.3 Bilan

Le fait que les villes les plus pauvres soient légèrement en avance (souvent hors de la fourchette définie par la moyenne et l'écart-type) confirme le résultat relatif aux ménages pauvres.

Par ailleurs, l'accélération déjà remarquée de la couverture des villes entre



**Figure 10** – Évolution mensuelle du déploiement de la 4G pour les 8 villes les plus peuplées. Zoom sur la période 2012–2017.

2018 et 2020 signale l’efficacité de l’Arcep et de ses derniers programmes<sup>18</sup>, intitulés « zones blanches centres-bourgs » et « new deal ». Les opérateurs avaient pour obligation juridique de garantir une couverture homogène du territoire et de basculer la grande majorité des émetteurs en 4G avant la fin 2020. Cette dernière régulation a certainement eu un impact fort sur le déploiement des réseaux mobiles, comme nous le voyons avec l’homogénéisation tardive des distributions statistiques. Il faudrait néanmoins vérifier l’effet de ces contraintes en étendant notre étude à d’autres pays.

## 5 Conclusion

Nous avons montré que les zones de ménages pauvres sont mieux couvertes que celles des non-pauvres ; second résultat contre-intuitif, la distribution des uns comme des autres obéit approximativement à une loi log-normale et s’éloigne significativement d’une loi de Poisson.

Par ailleurs, cette disparité socio-spatiale n’a pas d’analogie générationnel, puisque les distances aux antennes des 15–25 ans, des plus de 25 ans et des plus de 65 ans ont des distributions identiques.

L’étude historique de la progression de la couverture 4G des villes montre que les plus grandes ont bénéficié d’un taux de couverture plus élevé que les autres dans les quatre premières années, mais que cet avantage s’est ensuite réduit. Entre

18. Entre juillet et novembre 2018; cf. <https://www.arcep.fr/cartes-et-donnees/tableau-de-bord-du-new-deal-mobile.html>.

2018 et 2020, l'ensemble des communes de France a vu son taux de recouvrement téléphonique s'accroître et s'uniformiser, ce qui peut être mis au crédit de l'Arcep.

Cette enquête a eu aussi des retombées imprévues : la découverte des limites du *big data* (8 millions de Français manquent à l'appel), le masquage partiel des quartiers (très) riches et très pauvres. Elle induit aussi des questionnements méthodologiques, en nous invitant à la vigilance quant à la qualité des données mises à disposition du public tout en essayant d'en tirer le meilleur parti pour comprendre le monde contemporain.

## Références

- [1] F. Baccelli and B. Błaszczyszyn. On a coverage process ranging from the boolean model to the poisson voronoi tessellation with applications to wireless communications. *Advances in Applied Probability*, 33(2) :293–323, 2001.
- [2] F. Baccelli and B. Błaszczyszyn. Stochastic geometry and wireless networks. Volume I : Theory. *Foundations and Trends® in Networking*, 3(3–4) :249–449, 2010.
- [3] F. Baccelli, M. Klein, M. Lebourges, and S. Zuyev. Géométrie aléatoire et architecture de réseaux. In *Annales des télécommunications*, volume 51, pages 158–179. Springer, 1996.
- [4] F. Baccelli and S. Zuyev. Poisson-voronoi spanning trees with applications to the optimization of communication networks. *Operations Research*, 47(4) :619–631, 1999.
- [5] C. Boldrini and A. Passarella. Hcmm : Modelling spatial and temporal properties of human mobility driven by users' social relationships. *Computer Communications*, 33(9) :1056–1074, 2010.
- [6] S. Brandt and W. Maennig. Perceived externalities of cell phone base stations : The case of property prices in Hamburg, Germany. *Environment and Planning A : Economy and Space*, 44 (2) :396–410, 2012.
- [7] A. Busson and I. Lahsen-Cherif. Impact of resource blocks allocation strategies on downlink interference and sir distributions in lte networks : a stochastic geometry approach. *Wireless Communications and Mobile Computing*, 2018, 2018.
- [8] B. Błaszczyszyn and M. K. Karray. Spatial distribution of the SINR in Poisson cellular networks with sector antennas. *IEEE Transactions on Wireless Communications*, 15(1) :581–593, 2016.
- [9] Q. Cui, X. Yu, Y. Wang, and M. Haenggi. The SIR Meta Distribution in Poisson Cellular Networks With Base Station Cooperation. *IEEE Transactions on Communications*, 66(3) :1234–1249, 2018.
- [10] O. Dousse, F. Baccelli, and P. Thiran. Impact of interferences on connectivity in ad hoc networks. *IEEE/ACM Transactions on networking*, 13(2) :425–436, 2005.
- [11] P. Dubreuil. L'Arcep et la cartographie. *Études Digitales*, Numéro 10 :155–159, 2022.

- [12] H. ElSawy, A. Sultan-Salem, M.-S. Alouini, and M. Z. Win. Modeling and analysis of cellular networks using stochastic geometry : A tutorial. *IEEE Communications Surveys & Tutorials*, 19(1) :167–203, 2016.
- [13] M. Franceschetti and R. W. Meester. *Random networks for communication*. Cambridge University Press, 2007.
- [14] É. Guichard. Le mythe de la fracture numérique. In É. Guichard, editor, *Regards croisés sur l'internet*, pages 69–100. Presses de l'Enssib, Villeurbanne, 2011. Preprint :<http://barthes.enssib.fr/articles/Guichard-mythe-fracture-num.pdf>.
- [15] Y. Hmamouche, M. Benjillali, S. Saoudi, H. Yanikomeroğlu, and M. Di Renzo. New trends in stochastic geometry for wireless networks : A tutorial and survey. *Proceedings of the IEEE*, 109(7) :1200–1252, 2021.
- [16] B. Latour. Pensée retenue, pensée distribuée. In C. Jacob, editor, *Lieux de Savoir*. Espaces et communautés, pages 605–615. Albin Michel, Paris, 2007. Vol. 1.
- [17] D. Lee, S. Zhou, X. Zhong, Z. Niu, X. Zhou, and H. Zhang. Spatial modeling of the traffic density in cellular networks. *IEEE Wireless Communications*, 21(1) :80–88, 2014.
- [18] W. Lu, M. Di Renzo, and A. Busson. Stochastic geometry modeling and analysis of the error probability of two-tier cellular networks. In *EuCNC 2014*, 2014.
- [19] X. Lu, M. Salehi, M. Haenggi, E. Hossain, and H. Jiang. Stochastic geometry analysis of spatial-temporal performance in wireless networks : A tutorial. *IEEE Communications Surveys & Tutorials*, 2021.
- [20] I. Nakata and N. Miyoshi. Spatial stochastic models for analysis of heterogeneous cellular networks with repulsively deployed base stations. *Performance Evaluation*, 78 :7–17, 2014.
- [21] S. Parthasarathy and R. K. Ganti. Coverage analysis in downlink Poisson cellular network with  $\kappa - \mu$  shadowed fading. *IEEE Wireless Communications Letters*, 6(1) :10–13, 2017.
- [22] P. Pirozmand, G. Wu, B. Jedari, and F. Xia. Human mobility in opportunistic networks : Characteristics, models and prediction methods. *Journal of Network and Computer Applications*, 42 :45–58, 2014.
- [23] D. Pumain. Lois d'échelle et mesure des inégalités en géographie. *Revue européenne des sciences sociales*, XLV-138, 2007.
- [24] D. Pumain. *La géographie et les données massives*, pages 30–31. CNRS Editions, 2017. URL : <https://halshs.archives-ouvertes.fr/halshs-02008930/document>.
- [25] V. K. Rohatgi and A. K. M. Ehsanes Saleh. *Introduction to probability and statistics, 3rd Edition*. Wiley, 2015.
- [26] J. Tang, G. Chen, J. P. Coon, and D. E. Simmons. Distance distributions for matérn cluster processes with application to network performance analysis. In *2017 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE, 2017.

- [27] Y. Wang and Q. Zhu. Modeling and analysis of small cells based on clustered stochastic geometry. *IEEE Communications Letters*, 21(3) :576–579, 2016.

## Table des matières

<b>1</b>	<b>Projet et hypothèses</b>	<b>2</b>
<b>2</b>	<b>Sources</b>	<b>2</b>
2.1	Téléphonie mobile . . . . .	2
2.1.1	Installations radioélectriques . . . . .	2
2.1.2	Mise en pratique . . . . .	4
2.2	Socio-démographie de la France . . . . .	4
2.2.1	Variables de type démographique . . . . .	5
2.2.2	Variables de type économique . . . . .	5
2.3	Mise en correspondance des deux sources . . . . .	5
2.4	Remarques . . . . .	6
<b>3</b>	<b>Premiers résultats</b>	<b>7</b>
3.1	Limites du <i>big data</i> . . . . .	7
3.2	Des classes sociales estompées . . . . .	7
3.3	Point méthodologique . . . . .	8
3.4	Repérer les ménages pauvres . . . . .	11
<b>4</b>	<b>Apports de la modélisation</b>	<b>11</b>
4.1	La distance des ménages aux antennes . . . . .	11
4.1.1	État de l’art . . . . .	11
4.1.2	Notre méthode . . . . .	12
4.1.3	Distance aux antennes : ménages pauvres et autres . . . . .	12
4.1.4	Extrapolation de la distribution empirique . . . . .	13
4.1.5	Première conclusion . . . . .	15
4.2	Proximité des jeunes aux antennes . . . . .	16
4.3	Couverture des villes . . . . .	16
4.3.1	Taille des communes . . . . .	16
4.3.2	Quelques cas emblématiques . . . . .	18
4.3.3	Bilan . . . . .	19
<b>5</b>	<b>Conclusion</b>	<b>20</b>