



HAL
open science

Désagrégation spatiale des données de mobilité du recensement de la population appliquée à l'Ile-de-France

Cyril Pivano

► **To cite this version:**

Cyril Pivano. Désagrégation spatiale des données de mobilité du recensement de la population appliquée à l'Ile-de-France. Architecture, aménagement de l'espace. Université Paris-Est, 2016. Français. NNT : 2016PESC1185 . tel-01738731

HAL Id: tel-01738731

<https://pastel.hal.science/tel-01738731v1>

Submitted on 20 Mar 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Thèse de doctorat d'Université Paris-Est

Spécialité : TRANSPORT

Cyril Pivano

Désagrégation spatiale des données de mobilité du recensement de la population appliquée à l'Île-de-France

Thèse dirigée par :
Jean-Paul Hubert, IFSTTAR
Olivier Bonin, IFSTTAR

Soutenue le 20 octobre 2016

Jury

Brigitte Baccaïni (CGET)
Dominique Badariotti (Université de Strasbourg, rapporteur)
Arnaud Banos (CNRS, rapporteur)
Eric Cornélis (Université de Namur)
Christophe Terrier (Honoraire INSEE)



« À Paris, j'ai vu au moins quarante Canebière »

Marius (acte I, scène III)

Remerciements :

Cette thèse est le fruit de nombreuses rencontres. À mon égard, la plus fructueuse est certainement la rencontre avec Jean-Paul Hubert et Olivier Bonin qui m'ont fait confiance en me confiant ce sujet de doctorat.

La fonction du directeur de thèse est noble, il sème le futur dans le présent. Je ne remercierai jamais assez Jean-Paul Hubert d'avoir été un aussi bon jardinier. Il a su me laisser croire libre et indépendant ; cependant les idées qui ont poussées ne pouvaient suivre qu'une direction : celle d'une thèse soutenable.

Olivier Bonin est également un artisan de ces recherches et les formules de politesses viennent à me manquer. Il est le tuteur m'ayant accompagné durant ces trois années de recherches, bien plus, il est le botaniste venant chaque jour observer si la graine germe, se développe et fleurit bien. Et pour cela, il a dans de nombreux cas coupé court à toutes métaphores fallacieuses. Il est la personne qui a su m'écouter lorsque mes questions de recherche n'étaient pas clairement définies. Je tiens à le remercier pour la clairvoyance de ses réponses.

Je souhaite également remercier les financeurs de cette recherche : le Ministère de l'Environnement, de l'Énergie et de la Mer et le Carnot VITRES commanditaires du projet collaboratif Belgrand qui ont chacun financé une moitié de thèse. Cette recherche a aussi bénéficié d'une aide de l'État gérée par l'Agence nationale de la recherche au titre du programme Investissements d'avenir portant la référence ANR-10-EQPX-17 (Centre d'accès sécurisé aux données – CASD).

Je tiens ensuite à remercier, par avance, le jury de cette thèse pour le temps passé sur mes travaux scientifiques, et la pertinence de leurs remarques.

Une thèse est le fruit de nombreuses rencontres et la qualité des fruits dépend aussi du terreau. Je souhaite remercier toutes les personnes qui m'ont aidé, inspiré, supporté, et parfois même fait respirer.

Arrivé au Laboratoire Ville, Mobilité, Transport aux mois les plus profonds de l'hiver, la phase d'insertion et d'enracinement a été facile. Les deux premières personnes à m'avoir souri sont ni plus ni moins les figures emblématiques du LVMT Marianne et Mariane. Ces années se sont ensuite vite écoulées en l'agréable compagnie de toute une cohorte de doctorants, avec

d'abord Benoît, Vincent, Jeff, Éloïse Julie, Julie, Virginie, François, Anne puis les grands successeurs Teddy, Aude, Emmanuel, Xavier, Mathias, Maelis, Matthieu, Delphine, Youssef, Joséphine, Manon et Lydia. Je les remercie tous de m'avoir supporté dans les moments difficiles et même les autres.

Je consacre une mention toute particulière à Laurent et Laurent en raison d'incessantes intrusions de bureau pour des motifs scientifiques de la plus haute importance. J'espère avoir encore de nombreuses raisons d'apparaître.

Je tiens aussi à remercier Alexis et également à m'excuser auprès de lui, je lui avais promis d'être cité bien plus haut. Il a bien le droit à une phrase entière pour ses premiers pas vers la libéralisation des données.

Bien d'autres personnes sont à citer dans ces remerciements : Samuel, mon ancien collègue de bureau à qui je dois mes débuts, mes premières pousses. Philippe, quel bonheur d'avoir un collègue qui connaît Veynes, de le voir, je me rappelais mon pays. Natalia, merci d'avoir pris un air toujours aussi convaincu lorsque tu prononçais tes phrases de soutien. Sophie, merci d'être présente. Merci encore à Shaoqing, Xiaoyan, Gaëlle, ..

Le LVMT a encore beaucoup de personnalité qui ont directement ou indirectement aidé à la réalisation de cette thèse. Aux racines de mes travaux, il y a Caroline Gallez, Anne Aguiléra, Sandrine Wenglenski, Florent Le Néchet, Emre Korsu, Pierre Zembri, Anne Jarrigeon, Nicolas Coulombel, Gwendal Simon, Laetitia Dablanc, Fabien Leurent et bien entendu mon actuel collègue de bureau Pierre Frankhauser.

Je souhaite également remercier Sandrine, Sandrine et Virgine pour leur indéfectible appui au quotidien.

Enfin, je n'ai pas été tout à fait exact, la première rencontre chaleureuse à mon arrivée est bien entendu Sophie Cambon-Grau. Elle ferme la marche des remerciements consacrés aux membres du LVMT pour sa bienveillance naturelle.

Je dois ensuite remercier tous les membres du projet « Science et SAGE » (SCITESAGE) qui ont cru en moi une première fois. Merci à Nicolas Bécu. Merci à Nicolas Flipo. « *Le calcul sur maille a encore de l'avenir* » a-t-il dit lorsque je m'apprêtais à candidater à cette thèse, j'étais très loin de comprendre tout ce que cela pouvait signifier. Le hasard a fait que j'ai pu le remercier de vive voix le lendemain de ces lignes. Merci également à Amandine de Coninck, José-Frédéric Deroubaix, Catherine Carré, Jean-Paul Haghe, Jean-Marie Mouchel, Gaëlle Tallec et Céline Le Pichon.

Puis, je souhaite remercier l'ensemble du collectif MAPS. C'est dans ce réseau thématique de modélisation multi-agents appliquée aux phénomènes spatialisés que la question de l'accès aux données pour effectuer des simulations à grandes échelles a éclo en moi. C'est aussi là que j'ai rencontré Nicolas Bécu (que je remercie deux fois), Arnaud Banos, Patrick Taillendier, Diego Moreno, Marion Almaric, Nicolas Marillaux, Frédéric Amblard, Frédéric Rousseau et bien d'autres.

En remontant encore plus loin je souhaite remercier mes encadrants de mes années universitaires. Merci à Frédéric Audard, Sébastien Oliveau et salut à la mémoire de Jean-Luc Bonnefoy. Effectivement, les pousses d'idées que l'on s'accapare ne proviennent pas de nulle part. Merci à eux de m'avoir appris ce qu'un modèle de transport distribué signifiait.

Je tiens à remercier mes anciens collègues de promotion tous devenus docteur : Joan Perez, Marion Borderon et tout particulièrement Camille Perchoux. Elle est la personne qui a effectué le travail ingrat de lecture de premier jet et de formulation de remarques pertinentes, sans trop me brusquer.

Je souhaite aussi remercier ma famille pour son soutien et surtout d'avoir supporté mon absence. Merci Catherine, Jean-Roch, Céline et Raphaël et qu'on le veuille ou non, merci à Pascal. Il est évident que la première mériterait autant de lignes que ce manuscrit.

Enfin, je souhaite terminer en remerciant tous ceux que j'ai abandonnés, mais que je compte bien retrouver : le philosophe Aixois que je connais en personne (Blondel), Thomas, Zozette, Beber, Nicolas, Manu, Romain, Tanguy et en premier lieu Juliette.

Table des matières

Introduction générale	13
Chapitre 1 : Des données pour répondre aux changements de paradigme des politiques de transport	29
1.1 Le changement de paradigme des politiques de mobilité et ses conséquences sur le besoin de données en matière de mobilité.....	33
1.1.1 Genèse de la mobilité.....	34
1.1.1.1 Mobilité générale et mobilité individuelle.....	34
1.1.1.2 L'urbanisation en zones et la solution automobile au problème de l'accessibilité.....	41
1.1.2 Les nouveaux objectifs des politiques de transport : un changement de paradigme.....	48
1.1.2.1 Comprendre l'objectif du facteur 4.....	48
1.1.2.2 La réduction de la part modale de la voiture dans l'évaluation des projets de transport.....	57
1.1.2.3 L'appel aux données pour connaître et agir.....	62
1.2 Les données sur la mobilité dans la perspective des nouveaux besoins des politiques de transport.....	66
1.2.1 Les attributs et supports géographiques souhaités pour les données de déplacements dans le cadre du développement durable.....	66
1.2.2 Les données de déplacement produites par la statistique publique.....	72
1.2.2.1 Les Enquêtes Déplacements standard Certu.....	72
1.2.2.2 Les données de mobilité du recensement de la population.....	76
1.3 Élargir l'éventail des données pour répondre aux nouvelles questions.....	79
1.3.1 Les données de mobilité à l'ère du big data.....	79
1.3.2 Le couplage de données.....	89
1.3.3 Comment produire des couples origines-destinations domicile-travail synthétiques avec des localisations fines ?.....	92
Conclusion du premier chapitre.....	94
Chapitre 2 : Les fonctionnalités des supports d'enquêtes	97
2.1 Les trois fonctions des supports spatiaux des données de mobilité.....	100
2.1.1 Les supports d'analyse.....	100
2.1.1.1 Un support zonal obtenu par agrégation de zones de la base de données pour la description et la communication.....	100
2.1.1.2 Un support linéaire pour l'analyse du trafic : le passage par des données modélisées.....	104
2.1.2 Les supports des bases de données de mobilité diffusées par les producteurs.....	110
2.1.2.1 La production de données suffisamment représentatives pour être diffusées.....	110

2.1.2.2 La solution vers laquelle les producteurs se tournent est un support régulier constitué de petites mailles.....	113
2.1.3 Les référentiels de localisation.....	118
2.2 Les supports avec des données faiblement échantillonnées : le cas des Enquêtes Ménages-Déplacements.....	123
2.2.1 L'alignement des supports utilisés sur la géographie des IRIS de l'Insee.....	123
2.2.2 L'agrégation des données d'enquêtes à faible taux d'échantillonnage pour l'analyse.....	129
2.3 Les supports spatiaux et l'échantillonnage du recensement rénové de la population.....	132
2.3.1 Échantillonnage des communes de moins de 10 000 habitants.....	134
2.3.2 Échantillonnage des communes de plus de 10 000 habitants.....	135
2.3.3 La précision du recensement.....	140
Conclusion du chapitre 2.....	143
Chapitre 3 : Clonage et ventilations spatiales sous contraintes d'observations statistiques.....	145
3.1 Le clonage des observations du recensement.....	150
3.2 La ventilation spatiale sous contraintes.....	152
3.2.1 L'inférence écologique – extension aux données bi-localisées.....	155
3.2.2 Les données de contrôle pour ventiler les lieux de résidence.....	163
3.2.3 Les données de contrôle pour ventiler les lieux de travail.....	172
3.2.4 Les descriptions des trajets des déplacements comme variables de contrôle.....	176
Conclusion du chapitre 3.....	180
Chapitre 4 : Le processus de désagrégation spatiale des localisations résidentielles et d'emplois appliqué à la région Île-de-France.....	183
4.1 Appariement des fichiers détails Individus-CANTVILLE et MOBPRO.....	186
4.1.1 Appariement des observations de ménages actifs occupés des communes de moins de 10 000 habitants.....	190
4.1.2 Appariement des observations de ménages actifs occupés des communes de 10 000 habitants et plus.....	195
4.1.3 Évaluation de l'appariement avec les données du CASD.....	198
4.2 Ventilation des ménages aux lieux de résidence.....	202
4.2.1 Ventilation du support IRIS-Communes au carroyage de 200m de l'Insee.....	202
4.2.2 Évaluation de la ventilation sous contrainte des lieux de résidence des ménages.....	205
4.3 Ventilations des lieux d'emploi.....	216
4.3.1 Les secteurs d'activité.....	216
4.3.2 Les effectifs de salariés.....	218
4.3.3 Les distances en marche à pied.....	219
4.3.4 Évaluation de la ventilation des lieux de travail.....	220

Conclusion du chapitre 4.....	228
Chapitre 5 : Exploitations des données de mobilité désagrégées du recensement sur le support carroyé de 1km.....	231
5. 1 Analyse des distances domicile-travail à vol d'oiseau selon le choix du support	235
5.1.1 Les distances domicile-travail à vol d'oiseau.....	235
5.1.2 Les distances infra-communales.....	237
5.1.3 Les déplacements entre communes limitrophes.....	239
5.2 Cartographie des lieux générant de grands et de petits déplacements.....	241
5.3 Cartographie des carreaux d'attraction et d'émission.....	246
5.4 Zooms géographiques et discriminations socio-économiques.....	250
5.4.1 La zone aéroportuaire d'Orly et de Rungis.....	250
5.4.2 La Plaine-Saint-Denis.....	255
5.4.3 Vincennes, Saint-Mandé et le quartier du Bas-Montreuil.....	259
5.5 Les modes de transport.....	265
Conclusion du chapitre 5.....	276
Conclusion générale.....	277
Références bibliographiques.....	287
Références internet.....	300

Introduction générale

La désagrégation spatiale des données de mobilité du recensement est une méthodologie de production de données décrivant plus finement les localisations résidentielles et d'emplois des observations décrites dans les fichiers du recensement. La thèse proposée est que la désagrégation de données de mobilité du recensement apporte des éléments de connaissances supplémentaires, et répond ainsi, en partie, aux besoins de données dans le contexte du changement de paradigme des politiques de transport.

Le réchauffement climatique nous oblige à penser autrement les politiques de transport. L'essor économique d'hier s'est appuyé sur des moyens de transport rapides et relativement peu coûteux (et des inégalités de coûts de main d'œuvre). Maintenant, des voix s'élèvent progressivement et s'accordent à stigmatiser l'utilisation individuelle d'une voiture à carburants fossiles. Plus largement, le paradigme des politiques de transport évolue. Les besoins en données aussi.

Les anciennes politiques de transport demandaient aux professionnels de calibrer les infrastructures. Il convenait d'estimer le nombre d'utilisateurs induit par la construction de nouveaux équipements. Aujourd'hui, en plus des gains de temps, des gains économiques, de la quantification des trafics induits, de la réduction des points noirs de trafic, le changement de paradigme des politiques de transport considère les effets environnementaux et de santé publique ; ce qui nécessite une adaptation de l'appareil statistique publique. Par exemple, les politiques ont besoin de connaître les comportements de mobilité actuels pour inciter les individus à utiliser des modes de transport moins performants mais moins émetteurs de gaz à effet de serre (GES). Ce changement de paradigme impose quelque peu de découpler la croissance économique et la performance des moyens de transport. Dans ce contexte de volonté politique de diminution des émissions de GES liées aux transports, le besoin en données sur la mobilité se fait plus grand. Les données doivent décrire des objets plus divers. Notamment, elles peuvent servir à décrire les profils socio-économiques des individus pour mieux comprendre leur propension à utiliser systématiquement une voiture quand ils se déplacent, et ainsi proposer des aménagements de transport favorisant les reports modaux.

Un grand effort est actuellement fait pour produire plus de données, soit directement, soit indirectement en menant des politiques de diffusion de données (open data) ou en exploitant les nouvelles sources de données (big data). Et, parmi les objets de connaissance que les données tentent de décrire le thème de la mobilité est un peu particulier. C'est un objet insaisissable par essence. Localiser un déplacement dans l'espace est périlleux. Le décrire aussi. Pourtant les enjeux d'aujourd'hui nécessitent de disposer d'informations sur les déplacements des individus plus précis que jusqu'alors.

De manière générale, les données sont issues de processus de production plus ou moins complexes. Les données ont bien souvent des formats différents, ce qui ne facilite pas la comparaison entre elles. Dans le contexte du réchauffement climatique les données doivent décrire de nombreux objets de connaissances. Et, les supports géographiques des données doivent permettre la comparaison entre elles afin de comprendre les interactions entre les différents thèmes à traiter par les politiques environnementales. Les producteurs sont invités à diffuser les données en facilitant leurs combinaisons avec d'autres séries de données (interopérabilité). La transposition politique de cette nouvelle exigence a comme cadre législatif la directive européenne INSPIRE (INSPIRE, Directive 2007¹).

L'harmonisation des données proposée par la directive européenne nécessite un format interopérable. La volonté politique est de diffuser les données environnementales à l'aide d'un support géographique carroyé ayant un référentiel commun pour tous les pays membres de l'Union européenne. De cette manière, il est espéré répondre aux besoins des utilisateurs se servant des données pour effectuer des analyses. La thèse s'associe à cette volonté. Nous proposons une méthodologie de production de données sur la demande de déplacement sur un support carroyé. Plus exactement, nous allons désagréger les couples origines-destinations domicile-travail du recensement pour pouvoir les diffuser sur un support répondant aux prérogatives européennes.

Les données du recensement de la population apparaissent comme la première source de données à traiter, après examen des autres sources de données disponibles sur la mobilité. Elles renseignent sur de nombreuses caractéristiques, qualifient un grand nombre d'individus, et ce, avec un rafraîchissement temporel de 20 % chaque année. Cette richesse de description

1 INSPIRE. Directive 2007/2/CE du parlement européen et du conseil du 14 mars 2007. Journal officiel n°108. 25 avril 2007.p.14 [en ligne] (Consultée le 13/05/2016) < <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:FR:PDF>>

permet d'effectuer des analyses descriptives et peut également être mobilisée comme données d'entrées dans des chaînes de modélisation. Cette fonction sera d'autant mieux remplie si la description des localisations est suffisamment précise. Les données désagrégées du recensement sur un support répondant aux prérogatives européennes et de résolution fine pourront être combinées avec d'autres données.

La thèse consiste donc à effectuer un traitement statistique affinant les localisations des données de mobilité domicile-travail du recensement de la population sur un autre support. Le fichier détail anonymisé intitulé mobilité professionnelle (MOBPRO) diffusé par l'Insee renseigne sur les communes de résidence et d'emploi des actifs occupés. Nous allons voir que cette précision n'est pas suffisante. Le changement de paradigme demande plus de données pour analyser des territoires divers et pas nécessairement des données appuyées sur les limites administratives. En disposant de données finement localisées, elles pourront être ré-agrégées à façon et décrire des territoires variés. Le présent manuscrit a vocation à expliquer la méthodologie de désagrégation spatiale des données de mobilité du recensement. Pour affiner les localisations nous utilisons des données dites de contrôle ne renseignant pas directement sur la mobilité. Nous nous servons de données de contrôle pour construire une information sur la structure spatiale de la demande de mobilité domicile-travail (couples origines-destinations domicile-travail). Le processus de désagrégation est appliqué à l'ensemble des ménages dont au moins un actif occupé habite et travaille en Île-de-France.

En travaillant avec les données de mobilité du recensement nous traitons seulement de la mobilité professionnelle. Les déplacements domicile-travail sont néanmoins structurants du reste de la mobilité et nous supposons qu'ils sont à mettre en première ligne dans le contexte du changement de paradigme des politiques de transport.

La place des déplacements domicile-travail au sein de l'éveil écologique

Depuis quelques décennies, le monde prend conscience que le développement industriel et économique toujours croissant n'est pas durable (Meadows, 1972). En ce qui concerne l'aspect écologique, les concentrations atmosphériques de gaz à effet de serre (dioxyde de carbone, méthane et protoxyde d'azote) ont augmenté pour atteindre des niveaux sans précédent depuis au moins 800 000 ans (GIEC, 2013). Les politiques proposées à l'issue des grandes conférences internationales sur le climat s'accordent sur le fait qu'il faut réduire les émissions de gaz à effet de serre (GES) émis par les activités humaines. Dans le cadre du

document final issu de la conférence des parties (COP) de Copenhague en 2009, tous les principaux pays du monde se sont mis d'accord sur cet objectif.

Le G8 a également adopté en 2009 la recommandation du GIEC visant à diviser par 2 les émissions mondiales de GES entre 1990 (année de référence) et 2050. Cet objectif traduit l'effort que les pays industrialisés doivent accomplir ensemble pour que les émissions actuelles soient divisées par 2 au niveau de la planète, les ramenant ainsi de plus de 7 milliards de tonnes d'équivalent carbone à 3,7 milliards de tonnes d'équivalent carbone, ce qui correspond à la capacité maximale annuelle d'absorption par les océans et la biosphère.

Pour ne pas dépasser une concentration de 450 parties pour million (ppm) de CO₂ équivalent dans l'atmosphère, les émissions ne devront pas dépasser 0,6 tonne de CO₂ équivalent par habitant, ce qui pour la France exige un facteur de réduction environ égal à 4 (« Facteur 4 ») (CGEDD, 2013).

En France, le Grenelle de l'Environnement a officialisé cet engagement de réduction par 4 dès 2007. En 2013, les émissions de GES rejetés par la France se sont élevées à 497,8 millions de tonnes équivalent CO₂. À cette date la France représente 1,2 % des émissions mondiales alors qu'elle contribue à hauteur de 4,2 % au PIB mondial. Elle a diminué ses émissions depuis 1990 de plus de 10 % alors que dans le cadre du protocole de Kyoto l'objectif était de ne pas les augmenter (MEDDE, Emissions de la France²).

Le transport est le premier secteur d'activité émetteur de GES en participant pour un tiers des émissions en France (MEDDE, Emissions de la France). L'estimation des taux d'émissions dépend de nombreux paramètres (MEDDE, Le CO₂ et le transport³). Pour estimer l'importance des émissions de GES selon les différents types de transport, l'ADEME donne des chiffres sur les taux de consommation d'énergie fossile. Le transport tout confondu absorbe 70 % des consommations énergétiques dont 80 % par le transport routier. Et, les voitures représentent plus de 60 % des consommations du transport routier (ADEME, les chiffres clefs⁴).

En reprenant la terminologie du rapport issu du Grenelle de l'Environnement, nous devons

2 MEDDE. Emissions de la France [en ligne] (page consultée le 15/06/2016)

<<http://www.developpement-durable.gouv.fr/Emissions-de-la-France,33791.html>>

3 MEDDE, Le CO₂ et le transport [en ligne] (consultée le 29/06/2016) <http://www.developpement-durable.gouv.fr/IMG/pdf/Guide_Information_CO2-2.pdf>

4 ADEME, (2014). Les chiffres clefs. [en ligne] (consultée le 01/02/2015) <www.ademe.fr/chiffres-cles-climat-air-energie-2014>

aller vers un changement de paradigme (Tuot, 2007). Les anciennes politiques de transport souhaitaient améliorer les conditions de transport et d'accessibilité en s'appuyant sur la technologie automobile. Aujourd'hui, le changement de paradigme impose de ne plus faciliter les déplacements s'ils sont sources de GES. Les éléments de langage ont évolué, des études de transport nous sommes passés aux études de mobilité. Dans la littérature anglo-saxonne on parle de *mobility-turn* (Cresswell, 2009). Alors que les études classiques de transport s'attachent à quantifier le trafic, à estimer l'influence d'une route sur l'ensemble du réseau routier, une approche s'appuyant sur le concept de mobilité est plus globale.

La mobilité peut être définie comme la façon d'habiter un territoire en y pratiquant des activités hors de son domicile, à une certaine distance et à une certaine fréquence. Parmi ces activités, la mobilité liée au travail est essentielle. Le travail permet de disposer de revenus et trivialement de pouvoir payer un logement et satisfaire les besoins essentiels. L'Enquête Nationale Transport et Déplacements (ENTD) effectuée en 2008 indique que le motif domicile-travail représente près de 27 % des motifs de déplacements et 20 % quand il s'agit d'un lieu fixe et habituel. L'enquête estime que 72 % des déplacements domicile-travail sont effectués en voiture personnelle. La mobilité professionnelle est donc un élément central dans la tentative de réduire les émissions de GES : trois quarts d'un cinquième des déplacements totaux sont des déplacements domicile-travail effectués en voiture.

Connaître pour agir

L'ADEME donne des pistes pour agir (ADEME, le changement climatique⁵). L'action peut se porter sur trois leviers : la technologie, les changements de comportements et l'aménagement urbain. En plus de faire évoluer le parc automobile vers des véhicules plus propres, l'agence préconise de diminuer l'usage de la voiture par l'utilisation d'autres modes de transport : les transports actifs ou doux (marche et vélo) et les transports en commun. L'agence, pour le développement durable, insiste sur l'accessibilité des transports partagés (transports en commun, covoiturage, autopartage, transport à la demande, véhicules en libre-service) et des solutions pratiques pour les utiliser et pour passer d'un mode de transport à l'autre (horaires correspondants, stations regroupées...). Pour faciliter l'usage des transports

5 ADEME, Le changement climatique : comprendre ses causes et ses conséquences pour mieux agir [en ligne] (consultée le 25/05/2016) <<https://ile-de-france.ademe.fr/sites/default/files/files/DI/Changement-climatique/guide-pratique-changement-climatique.pdf>>

partagés, l'information en temps réel des usagers doit aussi être développée. Une troisième piste est envisagée pour diminuer les émissions de GES : la diminution du recours aux transports mécanisés est évidemment le meilleur moyen de réduire les émissions. Pour favoriser l'utilisation des modes doux le rapprochement des services du quotidien doit donc être facilité.

Pour mener à bien ces politiques, les législateurs ont un rôle crucial. Ils doivent, par exemple, instaurer des normes quant au taux maximum qu'un véhicule peut émettre en GES par kilomètre. Ils doivent également favoriser l'accessibilité aux transports partagés. Permettre la libéralisation des données de transport. Et la tâche la plus ardue est certainement de réorganiser les transports en agissant sur les localisations résidentielles et d'emplois afin de trouver un équilibre emploi-habitat favorisant les modes actifs.

Parmi ces objectifs, les données que cette recherche produit concernent plus particulièrement les politiques de transport favorisant l'usage des transports en commun et les politiques d'aménagement urbain devant avantager les modes actifs. Dans les deux cas, ces politiques nécessitent des connaissances rapportées à l'espace. L'information géographique constitue un savoir stratégique pour évaluer la durabilité des territoires. Le changement de paradigme des politiques de transport nécessite donc des données géolocalisées pour mener l'action.

De manière générale, les actions à mener pour s'adapter au réchauffement climatique doivent être territorialisées. Le rapport du Grenelle est clair sur le sujet : *« L'action locale concertée, à condition qu'elle intervienne à l'échelon territorial pertinent et que soient clarifiées les compétences des intervenants, fait l'objet d'un consensus. Seule une approche territoriale, dont la dimension variera selon les problématiques – depuis le quartier pour articuler les modes de déplacement, d'habitat et de consommation, jusqu'à l'inter-région pour estimer la cohérence de la répartition spatiale des activités au regard des espaces et de leur usage comme de leur contenu – peut répondre aux défis posés »* (Tuot, 2007, p.8)

Le rapport insiste sur la transparence. *« La loi doit passer et poser le principe qu'en matière environnementale tout est public et communicable, sans aucune exception, en dehors des trois restrictions classiques : ne pas porter atteinte au secret nécessaire à la performance économique de nos entreprises, ne pas porter atteinte au secret de la vie privée des personnes, ne pas mettre en danger la sécurité nationale ; même dans ces cas, l'accès à des informations doit être possible »* (Tuot, 2007, p.32)

La directive européenne INSPIRE vise à établir une infrastructure d'information géographique dans l'Union européenne pour favoriser la protection de l'environnement. « L'ordonnance du 21 octobre 2010 a transposé dans le droit français la directive européenne Inspire, qui, pour favoriser la protection de l'environnement, impose aux autorités publiques, d'une part de publier sur Internet leurs données environnementales géographiques, d'autre part de les partager entre elles. Ces dispositions ont pour objectif de faciliter la réalisation d'études et donc la conception, le suivi et l'évaluation des politiques environnementales, en favorisant la prise de décision dans un cadre démocratique, avec un bon niveau d'information de tous les acteurs et du grand public » (CGDD, 2011, p.1). La loi française suit les prérogatives de la directive et impose la diffusion de données relatives à l'environnement ; 34 thèmes sont énoncés (Inspire, Data Specification⁶). Citons par exemple, les réseaux de transport, l'hydrographie, les sites protégés, l'altitude, l'occupation du sol, l'ortho-imagerie, la géologie, les bâtiments, les lieux de production et les sites industriels... De nombreux objets d'études sont à décrire pour répondre aux enjeux climatiques. Même si la directive n'impose pas de créer ces données lorsqu'elles n'existent pas ou ne sont pas disponibles sous forme numérique, nous remarquons tout de même que dans la liste des 34 thèmes, il ne figure pas de données sur la demande de déplacements.

Pourtant une demande sur l'accès à des données renseignant sur les déplacements des individus est exprimée par les politiques, les administrations et les professionnels de la recherche scientifique. Par exemple, l'instruction du Gouvernement du 16 juin 2014 relative à l'évaluation des projets de transport présentant les nouveaux objectifs des évaluations des projets demande d'évaluer, avant et après, les taux d'utilisation des transports en commun dans les périmètres d'études concernés par l'aménagement. Toutefois les données sur la demande de déplacements sont lacunaires et ne permettent pas de répondre pleinement à cette demande. Les enquêtes de déplacements de l'appareil statistique national sont les enquêtes nationales transports et déplacements (ENTD), les Enquêtes Ménages-Déplacements (EMD) et dans une moindre mesure, les enquêtes de mobilité professionnelle du recensement de la population (MOBPRO). Le conseil national de l'information statistique (CNIS) exprime également un besoin de rénovation des données de déplacements (Cuillier, 2009). Un appel à projet du PREDIT-4, lui-même issu du Grenelle de l'Environnement, vise également à

6 Inspire, Data Specification [en ligne] (Consultée le 13/06/2016)
<<http://inspire.ec.europa.eu/index.cfm/pageid/2/list/7>>

améliorer la connaissance des pratiques de mobilités des individus. La thèse est également financée par un projet collaboratif du ministère de l'environnement qui s'inscrit dans cette demande politique qui réclame de meilleures données statistiques sur les comportements de mobilité des individus.

Face à cette demande, cette thèse propose d'examiner de plus près le rôle des supports géographiques des données afin de pouvoir améliorer les données existantes. Une donnée de déplacement est caractérisée par deux supports spatiaux zonaux et des attributs sur les déplacements. Un support zonal est une discrétisation d'un référentiel continu. Il est défini entre autres par sa résolution (surface moyenne des zones élémentaires composant le support) et sa couverture. Un support zonal permet de localiser des observations statistiques. Plus le support a une résolution fine, plus les localisations des individus sont précises. Outre les supports, les données de déplacements sont caractérisées par leurs attributs, ils décrivent les caractéristiques des déplacements (les motifs, les modes de transport, les distances, les durées, les trajets composant les déplacements) et celles des individus réalisant ces déplacements. Les variables caractérisant les individus permettent de construire des profils socio-économiques (âge, sexe, revenu, catégorie socioprofessionnelle, composition du ménage, équipement automobile). Les besoins exprimés par le changement de paradigme demandé par le Grenelle de l'Environnement nécessitent des données fines spatialement (des supports fins) et socialement (une description riche des attributs) pour mieux observer, analyser, et comprendre la mobilité. Par exemple, des données localisant précisément les origines-destinations des déplacements permettent de mesurer l'accessibilité aux modes de transport alternatifs à la voiture personnelle. Les supports spatiaux doivent avoir une résolution fine. Et, nous avons besoin d'une description socio-économique détaillée des individus pour comprendre les pratiques de mobilité et les choix des modes de transport de manière à pouvoir évaluer des potentiels de reports modaux.

Les données servent à décrire les individus mais pas uniquement. Elles servent à analyser des territoires, ce qui intéresse bien plus les politiques. Les données de déplacements doivent permettre de connaître la mobilité sur l'ensemble du territoire national et en ayant pour chaque territoire (finement découpé) un nombre d'observations suffisant pour comparer les comportements d'individus ayant des profils socio-économiques différents, mais ayant une ou des localisations similaires. Ces trois critères de représentativité (individuelle, spatiale et sociale) et ce, avec une précision géographique fine, demandent d'effectuer des enquêtes

interrogeant de très nombreux ménages. Les coûts de ces enquêtes peuvent vite devenir très élevés. Aussi, plus les données sont fines, plus la description des individus concerne leur vie privée et plus nous sommes confrontés aux contraintes légales du respect de l'anonymat. Nous proposons alors le couplage de données comme alternative de production.

Les problématiques de production de données de déplacement répondant au Grenelle de l'Environnement

Plusieurs solutions sont envisageables pour répondre pleinement au besoin en données du changement de paradigme des politiques de transport. Nous en énumérons plusieurs tout au long du manuscrit. Nous choisissons ici de désagréger les données de mobilité du recensement sur un support géographique carroyé pour d'une part, améliorer la connaissance sur la mobilité et d'autre part, permettre l'interopérabilité avec d'autres sources de données. Le recensement apporterait ainsi une précieuse information de par sa qualité d'échantillonnage et de par ses nombreuses variables socio-économiques.

Nous proposons d'effectuer une désagrégation spatiale des données de mobilité du recensement de la population. Le fichier de données librement diffusé informant sur la mobilité professionnelle est le fichier MOBPRO. Il décrit les communes de résidence et d'emploi des actifs occupés de plus de 15 ans. Le taux d'échantillon du fichier MOBPRO est de 25 %. Un quart de la population des actifs occupés résidant en France est contenu dans le fichier de données (8 339 649 observations représentant 26 703 768 actifs occupés). Le travail de la thèse est de proposer une méthodologie pour disposer de données sur l'ensemble du territoire en affinant les localisations des observations MOBPRO et nous l'appliquerons à l'Île-de-France.

Pour l'ensemble des actifs décrit dans MOBPRO, nous allons attribuer une localisation à l'intérieur de leur commune de résidence et une localisation à l'intérieur de leur commune d'emploi. Nous disons que nous désagrégeons les données du recensement car notre objectif est de répartir les observations dans tout l'espace géographique. Pour cela, nous allons devoir multiplier les observations puis les localiser plus finement. Pour les localiser, nous nous servons de données dites de contrôle.

Plusieurs leviers méthodologiques apparaissent:

- l'augmentation artificielle du nombre d'observations d'enquête ;

- la ventilation spatiale de données bi-localisées sous contraintes de données de contrôle ;
- les tirages de Monte-Carlo pour la ventilation spatiale ;
- l'agrégation et la ré-agrégation de support ;
- le couplage des données domicile-travail du recensement avec d'autres données de mobilité.

L'augmentation artificielle du nombre d'observations d'enquête :

L'objectif est de disposer d'éléments de connaissances sur l'accessibilité des territoires avec une résolution fine. Un grand nombre d'observations est nécessaire pour les répartir dans tous les lieux où se situent les origines et destinations des déplacements. Pour disposer d'un plus grand nombre d'observations, nous nous inspirons des méthodes de génération de populations synthétiques en proposant le concept de clonage d'observations d'enquêtes. Il s'agit de considérer les poids de sondage du recensement comme des coefficients multipliant les observations d'enquêtes. Nous espérons disposer de suffisamment d'observations statistiques pour recouvrir l'ensemble de l'espace géographique.

La ventilation spatiale de données bi-localisées :

L'espace géographique à recouvrir d'observations est défini ici comme les lieux générateurs de déplacements domicile-travail. Plus largement, nous disposons de données sur les lieux générateurs de trafic qui se matérialisent par une construction bâtie. Les données décrivant les bâtiments permettent d'estimer des effectifs de déplacement. En connaissant leur fonction, résidentielle ou économique, nous pouvons également orienter les flux, *i.e* distinguer les motifs. Cette connaissance sur les bâtiments est souvent elle-même peu précise géographiquement. Nous pouvons alors nous appuyer sur des données diffusées sur des supports zonaux. Nous parlons plus généralement de données de contrôle permettant de répartir ponctuellement les observations dans l'espace. Cette opération est nommée ventilation statistique. La ventilation spatiale de données bi-localisées est une succession de ventilations statistiques utilisant des données de contrôle différentes. Parmi les informations disponibles sur les appariements des couples origines-destinations domicile-travail, nous allons utiliser plusieurs données de contrôle ponctuelles et parfois des données renseignant sur les distances à parcourir. Plus largement, la méthode de ventilation spatiale de données bi-localisées s'apparente aux méthodes d'inférence écologique. Ces méthodes utilisent des tirages de Monte-Carlo.

Les tirages de Monte-Carlo :

La ventilation spatiale utilise des méthodes de tirages de Monte-Carlo pondérés par des variables de contrôle. Ces techniques s'appuient sur des tirages aléatoires répétés pour évaluer une quantité et/ou approcher la solution d'un système déterministe. Par exemple, dans une commune nous avons trois établissements économiques. L'effectif des salariés est une variable de contrôle que nous connaissons. Nous allons nous appuyer sur cette information auxiliaire (données de contrôle). Pour placer un actif à l'intérieur de sa commune de travail, nous devons choisir un établissement économique parmi plusieurs alternatives. Nous pouvons supposer que les individus ont des probabilités différentes de travailler dans chaque établissement. Un tirage de Monte-Carlo consiste à placer un actif dans un de ces établissements en fonction des probabilités construites à l'aide des variables attributaires des données de contrôle (effectif des salariés par exemple). Bien sûr nous ne savons jamais si les tirages au sort ont sélectionné les établissements économiques réellement fréquentés par les actifs. Toutefois, plus nous devons localiser d'actifs dans une zone, et donc plus nous effectuons de tirages au sort, plus la distribution des effectifs par établissements se rapproche de la réalité. Les tirages de Monte-Carlo permettent au moins de respecter le nombre de déplacements générés par les activités économiques. Nous pouvons utiliser la même méthode pour les lieux de résidence, à condition de disposer de données de contrôle. La question est de savoir si en procédant par tirages au sort successif des origines et des destinations, est-ce que nous reproduisons des couples origines-destinations représentatifs ? Quelle est l'incidence sur les indicateurs de la mobilité que l'on peut produire à partir de cette méthode ? En ventilant statistiquement les origines-destinations des actifs, retrouverons-nous des distances de déplacements cohérentes, retranscrivons-nous les contraintes d'accès aux systèmes de transport alternatifs à la voiture ?

Pour amoindrir l'incertitude inhérente à la stochasticité des tirages de Monte-Carlo, les données de contrôle pondérant les tirages peuvent être multipliées aux origines et aux destinations. Aussi nous pouvons ré-agréger quelque peu les résultats.

L'agrégation et la ré-agrégation :

La demande en données finement localisées n'a pas pour but d'être utilisée à une échelle microscopique. La production de données fines a vocation à être ré-agrégée pour pouvoir construire des supports géographiques d'analyses selon des besoins divers. Nous proposons de désagréger les origines-destinations sur des supports les plus fins possibles selon la

disponibilité des données de contrôle, puis nous ré-agrégeons les résultats sur un support régulier dont les mailles constituant le support sont des carreaux de 1km de côté. De cette manière, nous espérons gagner en représentativité tout en étant descendu en échelle. Plus nous ré-agrégeons, plus nous nous rapprochons des distributions avant désagrégation. Au demeurant, le choix du support carroyé de 1km permet de répondre aux objectifs d'harmonisation des données (INSPIRE).

Le couplage des données domicile-travail du recensement avec d'autres données de mobilité :
Enfin, la réponse que la désagrégation spatiale des données du recensement apporte ne permet pas de répondre complètement aux objectifs du Grenelle. Nous allons désagréger le fichier des mobilités professionnelles du recensement qui informe seulement sur le motif domicile-travail. Nous disposons au final d'une image partielle de la mobilité. Toutefois nous allons voir que le choix de ré-agréger sur un carroyage de 1km permet d'imaginer des solutions. La mise à disposition de données sur un support carroyé répond à la problématique d'harmonisation des données de la directive INSPIRE. L'idée est de produire des données de mobilité domicile-travail sur un support commun à d'autres types de données pour faciliter les analyses croisées entre des données décrivant des objets différents. Le travail de désagrégation des données du recensement peut être vu comme une étape préalable pour coupler différentes sources de données de mobilité tout en disposant d'une richesse de description des profils socio-économiques des individus. Pour faciliter l'interopérabilité des données nous proposons d'affiner les localisations du recensement sur un carroyage respectant les normes INSPIRE.

Nous remarquons que toutes ces questions sont en étroite relation avec un même outil conceptuel : les supports de l'information géographiques. Les problématiques sont développées au long des chapitres en introduisant systématiquement les contraintes des supports zonaux de l'information géographique.

Annnonce du plan

Le premier chapitre expose plus amplement les raisons de la demande de données exprimée par le Grenelle et le concept de couplage de données que nous proposons. Dans le contexte du changement de paradigme imposé par le Grenelle les politiques doivent partir de l'état présent pour mettre en place des infrastructures accompagnant les changements de comportements. Nous commençons alors par effectuer un travail descriptif des logiques de

localisations résidentielles et des activités économiques en surlignant le rôle catalyseur de l'automobile. Ce travail permet de décrire les déterminants individuels de la mobilité et ceux à une échelle supérieure. La mobilité des individus est structurée par l'agencement des lieux de résidence et des activités économiques et est facilitée par les infrastructures de transport. Les données de contrôle que nous utilisons ne renseignent pas directement sur les déplacements mais renseignent sur la structure de ces lieux générateurs de mobilité.

Ensuite, nous présentons quelques politiques et recherches scientifiques allant dans le sens du changement de paradigme. De cette manière, nous énumérons un certain nombre de demandes d'accès à des données de meilleure qualité. Les données actuelles ne sont pas assez précises pour effectuer des analyses descriptives de la mobilité. Nous effectuons donc un panorama des données de mobilité disponibles actuellement.

Ce travail permet d'argumenter notre choix de travailler avec le recensement de la population. À l'aide de cette vue d'ensemble des données nous mettons en exergue les forces et faiblesses des données du recensement pour effectuer des analyses descriptives. Mais les autres sources de données de mobilité ont également des inconvénients.

C'est alors que nous présentons l'idée de couplage de données. L'opération de couplage proposé consiste à combiner des données diffusées sur différents supports géographiques. Nous proposons de répondre aux enjeux du Grenelle en approfondissant la notion de support géographique des données. C'est ainsi que notre choix de traiter du recensement de la population prend une dimension supplémentaire. D'une part, la désagrégation des données de mobilité du recensement sert à affiner les localisations des couples origines-destinations domicile-travail, ce qui permet de faire des analyses descriptives des flux professionnels générés par les territoires aussi petits soient-ils. D'autre part, la richesse des profils socio-économiques qu'apporte le recensement désagrégé sur un support suffisamment fin va permettre de coupler les données du recensement avec les autres sources de données de mobilité. En effet, seul le recensement de la population offre une représentativité statistique sur l'ensemble du territoire national, ce qui nous fait supposer que le recensement doit être la source de données primaires sur lesquelles les autres sources de données pourront être combinées.

Notre proposition d'amélioration des données dans le contexte du changement de paradigme des données de mobilité suggère donc de travailler sur la résolution des supports des données. Le deuxième chapitre analyse les différentes fonctions qu'un support de données doit remplir.

Tout au long du processus de collecte, de diffusion et d'analyse, les supports géographiques jouent des rôles différents et imposent certaines contraintes de construction. Les fonctionnalités des supports des données d'enquête interagissent entre elles et conditionnent la qualité des données. Pour s'affranchir des contraintes des supports et favoriser l'interopérabilité des données, les nouvelles politiques en matière de production de données recommandent de les diffuser sur un seul et même support. Un support carroyé. Nous exposons dans un second temps les méthodes de construction des données d'enquêtes standardisées selon les normes en cours (standard Certu). Une troisième section décrit les méthodes de collecte du recensement. Cette description est réalisée en se focalisant sur la notion de poids de sondage. Les poids de sondage donnent une indication sur la représentativité statistique des données et nous verrons qu'ils sont dépendants des supports géographiques.

Le chapitre 3 présente les deux étapes théoriques devant être réalisées avant la mise en application concrète de la désagrégation spatiale. Pour transférer les données de mobilité vers un support de diffusion carroyé fin, nous avons besoin d'une enquête à fort taux d'échantillonnage, en plus d'une richesse des profils socio-économiques. La taille de l'échantillon et la richesse des variables attributaires décrivant les individus sont les critères prédéterminant la qualité de la désagrégation spatiale. Selon la taille de l'échantillon nous allons pouvoir ventiler les observations dans tout l'espace occupé par les données de contrôle et gagner en représentativité spatiale. Et selon la description des caractéristiques individuelles, nous allons pouvoir effectuer des correspondances entre les données à désagréger et les données de contrôle et obtenir, nous l'espérons, une information nouvelle sur la distribution des origines-destinations. Pour augmenter la taille des échantillons d'enquête, nous proposons de cloner les observations du recensement pour disposer d'un grand nombre d'observations. Cette opération conceptuelle consiste à détourner quelque peu l'utilisation traditionnelle des poids de sondage. Les poids de sondage des enquêtes sont utilisés ici comme coefficients multipliant les observations d'enquête. Nous multiplions les observations à l'identique pour pouvoir les ventiler dans l'espace des données de contrôle.

Le chapitre continue en présentant la méthodologie générale de ventilation de données et nous précisons comment nous répondons à la problématique de ventilation de données bi-localisées. Nous allons utiliser des données de contrôle pour ventiler les observations statistiques aux lieux de résidence. Et, en disposant d'un profil socio-économique permettant

d'inférer sur la localisation des emplois des actifs, nous ventilons dans un second temps les actifs occupés dans des établissements économiques.

Pour affiner les localisations nous devons donc sélectionner les variables de contrôle infra-communales (parmi celles disponibles) permettant de ventiler au mieux les couples origines-destinations à l'intérieur des communes. Nous présentons en fin de chapitre les données de contrôle que nous utilisons dans le processus de désagrégation appliqué à l'Île-de-France.

Le chapitre 4 présente les étapes du processus de désagrégation que nous avons effectué. Le processus de désagrégation spatiale appliqué à l'Île-de-France est constitué de trois étapes. La première consiste à appairer les données de mobilité du recensement avec les autres fichiers de données du recensement. Ainsi, nous allons gagner en précision géographique pour une partie des observations. Mais nous allons surtout obtenir une description plus riche que dans le fichier des mobilités professionnelles diffusé par l'Insee (MOBPRO). La deuxième étape consiste à ventiler spatialement les lieux de résidence des données du recensement sur le carroyage de 200m diffusé par l'Insee et répondant aux prérogatives INSPIRE. La dernière étape s'attache à ventiler les lieux de travail des actifs occupés décrits dans les fichiers de l'Insee à l'adresse postale. Pour chacune de ces étapes nous présentons des éléments d'évaluation.

Le dernier chapitre expose les apports de ce travail de désagrégation spatiale pour effectuer des analyses descriptives. Nous reprenons les thématiques qui auront été abordées au long du chapitre 1 et 2 et cartographions les résultats de la désagrégation. Les thèmes abordés sont les informations sur les déplacements de petites distances (déplacements infra et trans-communaux), la localisation des lieux émettant de grands et petits déplacements, la possibilité d'observer des territoires infra-communaux et/ou à cheval sur les limites administratives, la possibilité de disposer d'indicateurs d'accessibilité sur un support carroyé.

Au dernier chapitre, nous allons plus simplement nous concentrer sur l'enrichissement des analyses descriptives qu'apportent les données désagrégées de mobilité du recensement. La conclusion générale sera l'occasion de revenir sur les perspectives que ce travail de désagrégation permet d'envisager, notamment le couplage avec d'autres sources de données de mobilité.

Chapitre 1 : Des données pour répondre aux changements de paradigme des politiques de transport

Chapitre 1 : Des données pour répondre aux changements de paradigme des politiques de transport

Nous devons changer nos modes de vie pour faire face au réchauffement climatique. Parmi les habitudes à changer, les comportements de mobilité sont un enjeu majeur. Les recherches à mener dans la discipline du transport sont appelées à être renouvelées. L'ancien paradigme de recherche de l'augmentation de l'accessibilité générale au moindre coût, qui conduit à une utilisation prédominante et croissante de l'automobile, qu'on appellera le « tout routier », doit être abandonné pour celui du développement durable. C'est-à-dire une gestion plus contrôlée de l'amélioration de l'accessibilité, qui ne s'appuie plus seulement sur les modes les plus rapides, qui favorisent la croissance des villes en périphérie, mais aussi sur les modes lents (Wiel, 1991).

L'objet de ce chapitre est de montrer que ce changement de paradigme nécessite un travail de recherche sur les données de mobilité. L'accessibilité aux services, aux emplois, aux aménités est inégal dans l'espace géographique (Gottman, 1966) et l'on peut proposer que l'objectif général d'une politique de mobilité est de réduire cette inégalité. Pour mener à bien ce changement de paradigme les nouvelles politiques doivent disposer de données précises pour observer, localiser et promouvoir des changements de comportements de mobilité à moyen et long termes. Mais cela suppose aussi de comprendre les logiques spatiales qui ont produit ces distributions de ménages et d'activités. Il faut en effet respecter ces logiques, qui ont leurs rationalités économiques et sociales, et ne plus chercher à résoudre le problème de l'accès des ménages aux activités par le seul développement de l'automobile individuelle.

La première sous-partie (1.1) expose les logiques spatiales des répartitions et des mouvements des lieux de résidence et des activités économiques. Les localisations des lieux de résidence et des activités économiques sont respectivement soumises à des forces centrifuges et centripètes. Les établissements économiques se concentrent dans l'espace et les localisations résidentielles ont plus tendance à se diffuser. Les individus ont besoin de se déplacer pour travailler, se récréer, et habiter dans un lieu agréable. Les différentes logiques de production de ces lieux de résidence et d'activité ont donc engendré une demande de déplacements. L'offre de transport, collectif et individuel, a permis la réalisation de ces déplacements. La demande de déplacements n'est donc pas aléatoire.

Dans un premier temps nous présentons quelques éléments sur la formation de la mobilité.

Nous présentons la mécanique des lieux de résidence issue de la littérature : la concurrence foncière et la ségrégation socio-spatiale. Pour une description plus complète de la mobilité, la présentation de ces forces localisantes oblige à présenter les forces de la mécanique économique.

Ensuite, nous introduisons les nouvelles problématiques du changement de paradigme des politiques de transport. Outre la nécessité de réduire les inégalités d'accessibilité, les nouveaux enjeux climatiques ajoutent aux prérogatives des politiques de transport les questions environnementales. Nous présentons quelques recherches allant de ce sens. Les recherches présentées sont rassemblées sous le terme de politiques de facteur 4. Elles amènent toutes au constat suivant : il faut davantage des données décrivant les individus dans les territoires (1.2). En effet, en disposant des données sur les individus nous pouvons rassembler les informations pour étudier des lieux précis (des zones). Ce chapitre cherche donc à savoir en quoi il est nécessaire de disposer de plus de données sur les individus et sur les territoires. Dans le contexte du développement durable, les données sur les individus, sur leurs comportements sont particulièrement importantes pour améliorer l'efficacité énergétique des réseaux de transport en permettant de mieux cibler les territoires trop dépendants à la voiture personnelle. Il faut bien des données sur la présence des individus sur les lieux où il est souhaité agir. Problème de taille, les individus sont mobiles. Disposer de données sur la mobilité des individus est donc un impératif pour mieux connaître les territoires, c'est-à-dire connaître la présence des individus dans l'espace et dans le temps.

De quelles données dispose-t-on ? Que souhaite-on comme résolution ? Et comment produire les données nécessaires?

Pour répondre à ces questions, ce chapitre effectue un panorama des données de mobilité actuellement disponibles afin d'identifier les besoins en données. Deux types de données sont proposés : les données descriptives et les données explicatives (modélisées).

Les données officielles des enquêtes traitant des déplacements (les enquêtes nationales sur les Transports, les enquêtes Ménages-Déplacements) sont conçues pour l'étude d'un périmètre donné. Elles ne permettent pas de décrire avec précision les localisations internes à leur périmètre, donc la structure spatiale fine de cet espace. Elles sont donc peu utiles pour mener une approche spatiale. Néanmoins, elles apportent des éclairages sur les relations macroscopiques et il est possible de faire des séries statistiques temporelles. C'est à partir de

ces types d'enquêtes que des travaux prospectifs sont effectués. Mais ils sont fondés sur des relations observées sur la structure spatiale actuelle des aires urbaines, dans une dynamique de croissance par étalement. Ces travaux de préparation de l'avenir demandent donc plus de données pour élaborer des scénarios de structure et de croissance urbaine.

Les données actuellement produites résultent d'un arbitrage entre les besoins d'informations des services techniques pour conseiller les instances politiques et les moyens financiers disponibles. Le reproche est souvent exprimé que les enquêtes coûtent cher par rapport à l'utilisation qui en était faite (Cuillier, 2009), il paraît donc impossible d'augmenter le taux d'échantillonnage de ces enquêtes pour améliorer la précision.

Pour observer plus finement les déplacements des individus dans l'espace, il peut être envisagé a priori d'utiliser les données issues des questionnaires des réseaux de transport et de téléphonie (1.3). Cependant elles n'offrent pas de description fine des profils socio-économiques des individus, sauf à faire une enquête sur un échantillon de possesseurs de téléphone, ce qui soulève à nouveau un problème de coût. Pour reconstituer un profilage socio-économique des territoires et pour comprendre les comportements d'utilisation des réseaux, une autre voie est de procéder par modélisation grâce à des méthodes de couplage de données (Markoff et Shapiro, 1973), (Gotway et Young, 2002), (Roger, Pebesam, Gomez-Rubio, 2013). Ces méthodes sont toutefois limitées par la qualité de l'alignement des supports spatiaux et par le taux d'échantillonnage des bases de données qui sont couplées.

Ce chapitre démontre et conclut que pour faciliter les opérations de couplage de données de mobilité sur un support géographique d'échelle communale avec des données plus finement localisées un traitement sur les données du recensement rénové de population doit être effectué. Ce travail de descente d'échelle est nommé désagrégation spatiale des données du recensement.

1.1 Le changement de paradigme des politiques de mobilité et ses conséquences sur le besoin de données en matière de mobilité

L'analyse de la contribution des transports aux émissions de GES a fait prendre conscience que les comportements de mobilité se rangeaient sur une échelle de « durabilité » ou d'émissions de CO₂. Généralement le comportement reconnu comme le moins « durable » est celui de l'utilisation d'une voiture par un seul individu (autosolisme). L'utilisation des transports en commun, du vélo ou de la marche émet moins de CO₂. Selon l'ADEME (ADEME, Comparateur Eco-déplacements⁷) un trajet domicile-travail sur un an de 10km émet environ 869 kg de CO₂ en voiture personnelle, 559kg en bus, 125kg en TER, et 1kg6 en RER et métros. Mais la substitution de l'autosolisme par les modes plus durables ne dépend pas seulement de l'individu mais aussi de la forme urbaine et de sa structure de densité de résidence et d'emploi (Cervero,1997), (Frankhauser et al. 2007), (Frankhauser 2013), (Caruso et al. 2015).

La substitution des modes de transport peut être un objectif de politique de transport, mais une politique de transport vise d'abord à améliorer l'accessibilité des activités et des emplois pour les habitants et la réduction des inégalités. C'est pourquoi ce sous-chapitre commence par une section sur la formation de la demande de mobilité des individus et l'accessibilité générale d'un territoire qui conditionne la réalisation de cette demande (1.1.1). Selon nous c'est cette accessibilité générale qui est l'objet des politiques de transport. L'automobile a été un grand facteur d'amélioration de cette accessibilité générale, au point que bien des territoires en dépendent pour que leurs habitants puissent accéder aux activités dont ils ont besoin.

Mais, depuis la prise en compte des questions environnementales, les politiques de mobilité ont intégré l'objectif de réduire les émissions de gaz à effet de serre et cela représente un changement de paradigme qui remet en cause le besoin de données sur la mobilité (1.1.2). Dans l'ancien paradigme, la réduction des inégalités d'accès aux activités et à l'emploi s'opérait surtout grâce à l'accroissement des vitesses des déplacements permis par l'automobile en périphérie des villes, où plus de ménages pouvaient s'installer à moindre coût

⁷ ADEME, Comparateur Eco-déplacements [en ligne] consultée le 22/05/2016. <<http://quizz.ademe.fr/eco-deplacements/comparateur/>>

que dans le centre. Or, ce processus doit être freiné car il met en danger les conditions de vie des générations futures. La réduction des inégalités d'accès doit désormais se faire en utilisant autant que possible les ressources des modes moins polluants et en analysant de manière plus fine, les logiques de localisation de la résidence et des activités des ménages qui produisent la demande de mobilité. Cela a conduit à fixer des objectifs de réduction des émissions de gaz à effet de serre par le transport (politique du facteur 4) et à revoir l'évaluation des politiques et des projets de transport sur la base d'une information plus large.

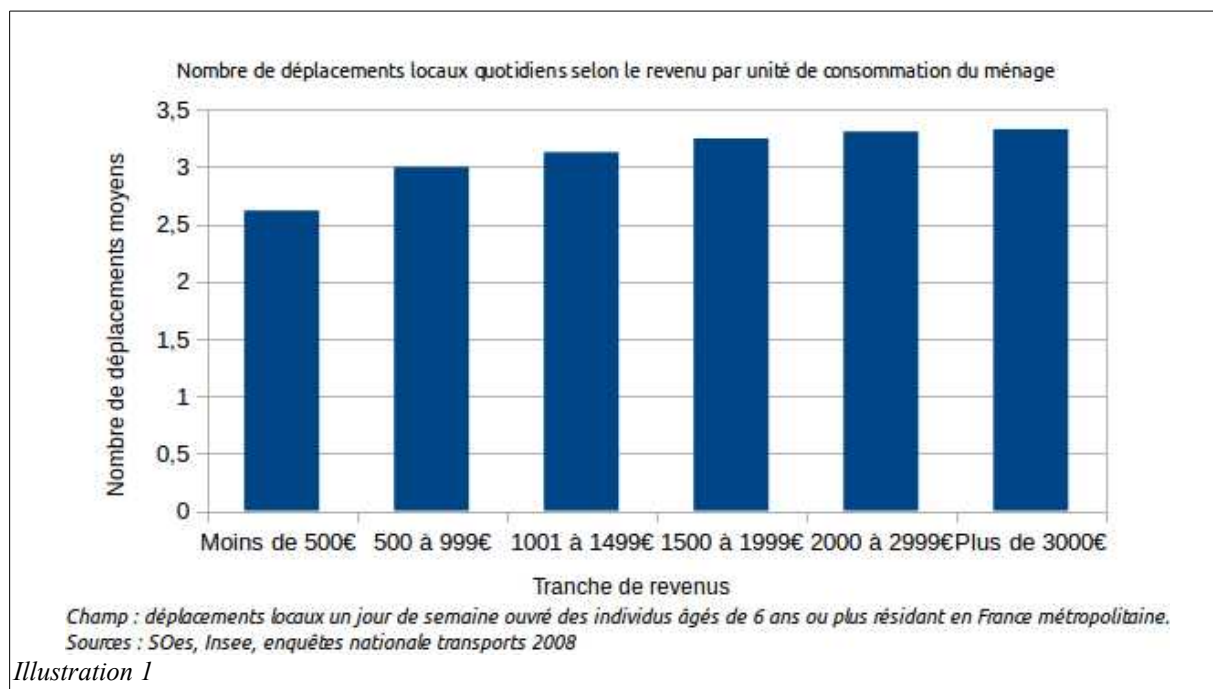
1.1.1 Genèse de la mobilité

La demande de la mobilité est à la fois issue de processus de choix individuel et est également formée par les opportunités offertes par les infrastructures de transport et les politiques d'urbanisme. Nous commençons par définir les problématiques individuelles d'accès à la mobilité. Puis nous montrons comment les anciennes politiques de transports ont cherché à répondre aux inégalités d'accessibilités.

1.1.1.1 Mobilité générale et mobilité individuelle

« En 2008, chaque jour de semaine ouvrée, les Français de 6 ans ou plus effectuent en moyenne 177 millions de déplacements locaux » (Armoogum, Hubert, Roux, 2010,p.5). Pris individuellement, un Français effectue entre 3,0 et 3,4 déplacements par jour. *« Ce sont 175 millions de décisions instantanées qu'il faut essayer de comprendre, d'analyser pour conduire des changements »* (ADEME, Blog Transport du futur⁸). Cette multitude de décisions de déplacement est habituellement subsumée sous le concept de mobilité générale. La mobilité générale est un phénomène collectif : c'est la somme des mobilités individuelle. Pour la comprendre, il faut connaître la façon dont les mobilités individuelles se combinent dans l'espace géographique et se renforcent. Au niveau individuel, la mobilité est aussi la capacité à se déplacer. Les analyses sociologiques ont démontré très tôt la relation entre la position dans l'échelle sociale et la volonté et capacité à se déplacer (Sorokin, 1927). Il existe une relation entre les revenus, les distances parcourues et le nombre d'activités réalisés (illustration 1).

8 ADEME: Blog Transport du futur [en ligne] (consultée le 04/04/2016)
<<http://transportsdufutur.ademe.fr/2013/06/appel-a-projet-mobilites-2013-connaître-pour-mieux-agir.html>>



Cette augmentation des distances parcourues selon les revenus s'explique en partie par des aspirations individuelles croissantes et rendue possible par les moyens de transport rapides et coûteux. Dans un contexte de croissance économique et urbaine, les services se diversifient mais selon des logiques de localisation qui ne sont pas forcément celles des ménages. Les individus souhaitent accéder à ces services en se déplaçant dans l'espace géographique. Cette demande est multiple et prend différents visages selon les catégories socio-économiques.

Pour analyser la mobilité générale, il est nécessaire de connaître les déterminants de la somme des mobilités. Les lieux de résidence des citoyens déterminent, en partie, la façon dont les mobilités s'additionnent. Outre, les attributs intrinsèques du logement qui sont plus ou moins fréquents selon les lieux, le choix du lieu de résidence doit répondre à un besoin de proximité à diverses aménités. Les individus doivent accéder à de multiples services et la localisation résidentielle doit permettre cet accès.

Le choix résidentiel dépend, entre autres, de l'accessibilité du lieu de résidence aux autres services comme les lieux d'emploi et les lieux de consommation de biens courants. La localisation résidentielle offre un certain panel d'opportunités. Et, pouvoir choisir dans ce panel d'opportunités dépend, en autres, de sa capacité à se déplacer. Cette capacité dépend des moyens économiques permettant d'arbitrer entre un logement de qualité et la proximité à

un éventail d'aménités urbaines, dans un contexte de concurrence aux ressources foncières amenant à une ségrégation socio-spatiale (illustration 2 et 3). Ainsi, en Île-de-France, les localisations résidentielles des ménages les moins aisés sont plus éloignées des services courants (Vallée al., 2015)

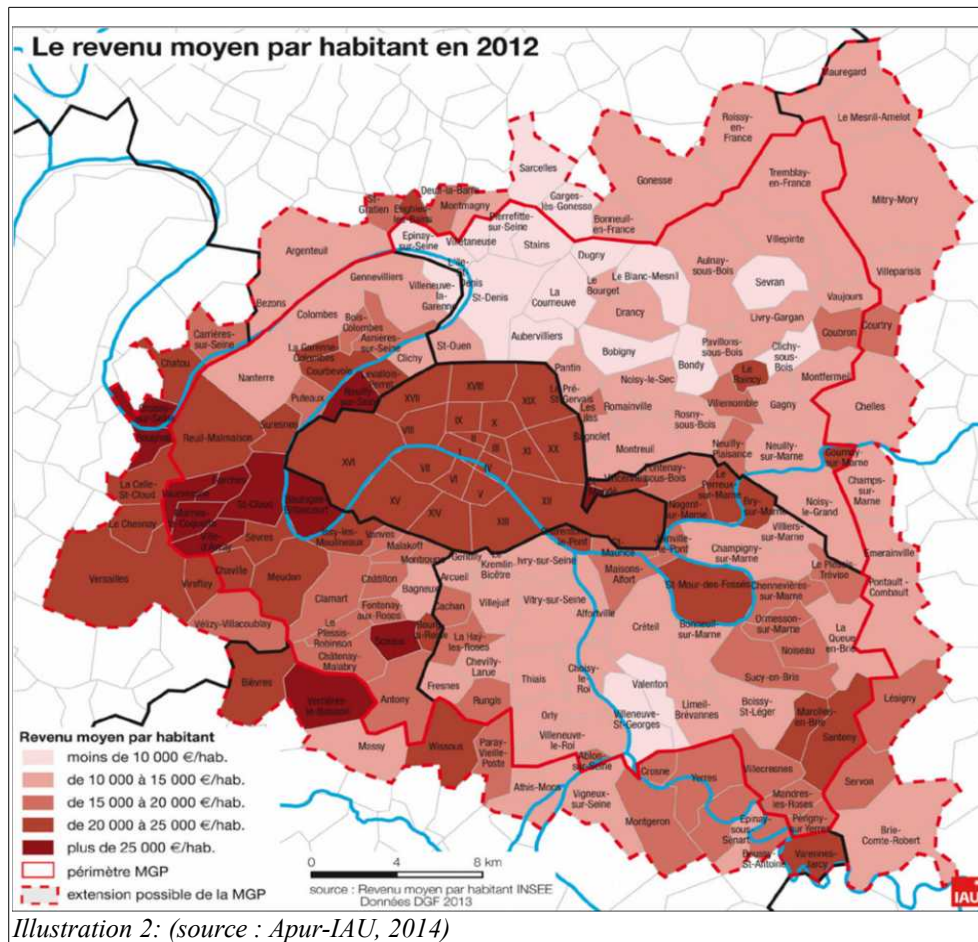
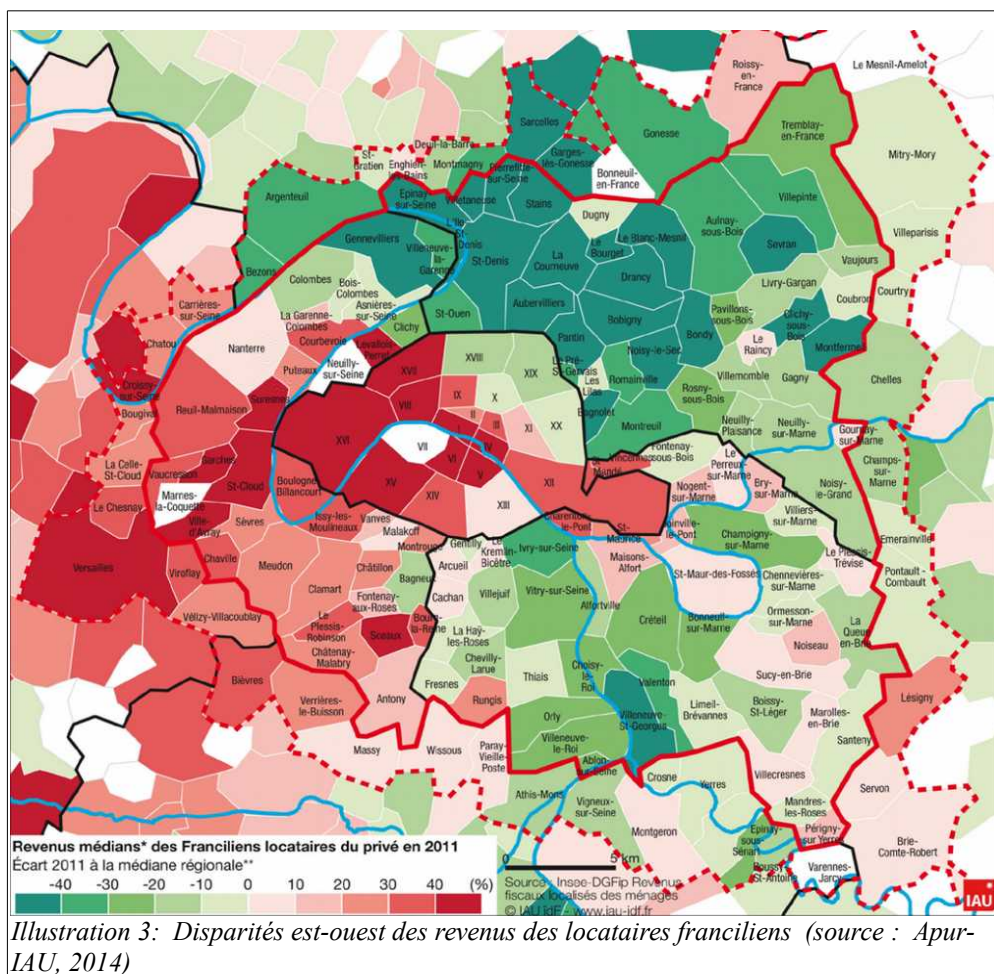
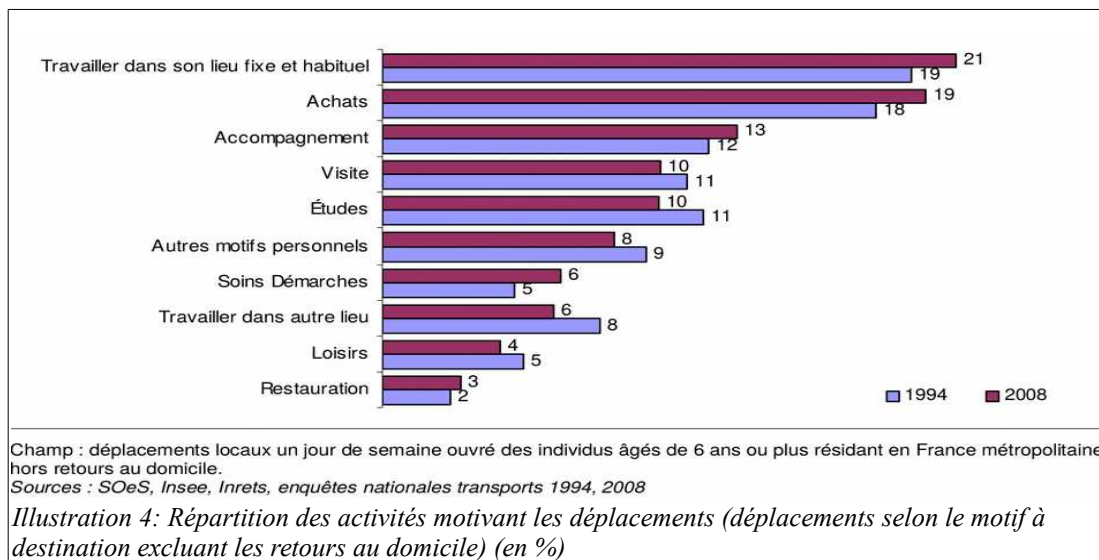


Illustration 2: (source : Apur-IAU, 2014)



Les choix résidentiels sont contraints par le gradient des prix fonciers qui est fortement différencié (illustration 3). Sous l'influence des mécanismes de l'offre et de la demande, les lieux offrant une meilleure qualité de vie sont plus chers à occuper, même si la qualité de vie n'explique qu'une faible partie des prix fonciers (Bonin et Hubert, 2015). La demande de logement des ménages ayant le plus de ressources économiques, peut donc conduire à faire augmenter les prix fonciers et empêcher les ménages plus modestes d'accéder au logement même dans les quartiers centraux mais traditionnellement populaires (phénomène de gentrification). Les évolutions sont différentes selon que l'accès au logement se fait par la location ou, ce qui est beaucoup plus récent, par l'accession à la propriété. Celle-ci engage des sommes bien plus importantes, des systèmes de crédit sur de longues périodes et peut donner prise à la spéculation. Par l'évolution des prix du foncier, les moins aisés sont renvoyés dans

des localisations moins convoitées ou plus périphériques (moindre accessibilité) et donc moins chères : «*la question des stratégies, résidentielles et autres, est directement liée à celle des ressources. Sans ressources, pas de stratégies. C'est par l'intermédiaire des ressources que stratégies individuelles et contraintes structurelles interagissent.*» (Bonvalet et Dureau, 1990, p.132). Les politiques de la mobilité doivent donc considérer ces mécanismes qui restreignent les choix de localisation résidentielle.



Continuons l'analyse de la demande de mobilité avec le motif du travail. Selon le rapport «*La mobilité des Français, panorama issu de l'enquête nationale transports et déplacements 2008*» commandité par le Commissariat Général du Développement Durable (CGDD, 2010), il y a en France, en 2008, 14,6 millions d'actifs ayant un lieu de travail fixe et régulier. De manière générale, ce motif représente 20 % (illustration 4) de l'ensemble des 177 millions de déplacements. En moyenne, les Français habitent à 11 km de leur lieu de travail. Cependant, la mesure des temps de trajet domicile-travail est difficile. Dans bien des cas les parcours ne sont pas directs. L'Enquête Nationale Transports Déplacements (ENTD) 2008 donne comme indication que les Français se déplacent en moyenne 55 minutes par jour, sachant qu'un déplacement dure en moyenne 17,9 minutes pour une distance de 25,2km. La littérature montre que les budgets-temps quotidiens moyens sont à peu près constants selon les agglomérations (Zahavi, 1980), mais les distances domicile-travail varient selon les déterminants socio-économiques (âge, CSP, sexe) (Bacaini, 1995) et selon les localisations (taille de l'agglomération, position dans l'aire urbaine) (Hubert et Delisle, 2010). Nous

pouvons ajouter que les choix de la localisation résidentielle sont dépendants du lieu d'activité professionnelle. Mais il n'est pas observé de volonté directe d'habiter à proximité de son lieu de travail (Korsu, 2006).

Les lieux de résidence doivent permettre l'accès à des services courants, marchands ou non. Toutefois, les actifs occupés peuvent aussi accéder à ces activités depuis leur lieu de travail. L'examen du chaînage des activités permet de distinguer des lieux structurants (lieu de résidence et lieu d'emploi), des lieux substituables des activités courantes ou des lieux exceptionnellement fréquentés, à mettre au second plan.

Ces deux localisations sont supposées rendre accessibles un certain nombre d'autres opportunités sous la contrainte du temps disponible pour les activités extérieures, domestiques et du transport. Cette contrainte du temps a été formalisée par les travaux de la Time-Geography (Hägerstrand, 1975). Autrement dit, les déplacements sont en partie structurés par les localisations résidentielles, et d'emplois pour les actifs. À partir des données de l'Enquête Globale Transports d'Île-de-France 1991–1992, Mouna Boulahbal explore l'effet structurant de la liaison domicile-travail. *« Il apparaît que le lieu de travail a un rôle polarisant fort sur la distribution spatiale des activités des individus, ce rôle pouvant être plus ou moins accentué selon la situation géographique de l'emploi et du domicile. En revanche, l'alignement de la distribution des activités autour du trajet domicile-travail est moins bien vérifié et, quand il est observé, il reste limité à une partie du trajet domicile-travail proche du lieu de résidence »* (Boulahbal, 2001). Il en ressort que les activités régulièrement exercées par les individus ayant un emploi ont tendance à être proches du lieu de résidence ou du lieu de travail, selon l'accessibilité des activités à ces lieux et selon le temps dont disposent les individus en ces lieux.

Les lieux de résidence et d'emplois ne déterminent pas à eux seuls les mobilités. Les besoins de déplacements et les arbitrages entre les différents lieux d'activités évoluent dans le temps sans que les individus changent nécessairement leur lieu de résidence ou d'emploi. En fonction du cycle de vie, les localisations à atteindre évoluent. Les choix de localisation résidentielle d'hier peuvent ne plus correspondre aux activités exercées aujourd'hui. (changement d'emploi, position du ménage). Les lieux changent également, que les ménages aient pu ou non anticiper certains de ces changements.

L'agrandissement des ménages par cohabitation et naissances entraîne souvent un

changement de résidence et peut provoquer un besoin de mobilité plus grand. Les ménages souhaitant habiter dans des logements plus grands s'éloignent des centres où ces logements sont rares, ainsi la proximité aux lieux d'emploi, aux commerces ou aux lieux de divertissement n'est plus la priorité. Mais les individus ne ré-adaptent pas pour autant l'ensemble de leurs programmes d'activités quotidiennes. Il est dangereux de vouloir systématiquement normaliser les comportements, d'autant plus que les tentatives d'optimisation de la mobilité peuvent suivre des logiques individuelles, de couples ou de familles qui visent des temporalités différentes.

Les choix résidentiels sont contraints par des conditions d'accessibilité. Parmi les critères de localisation des ménages avec enfants la possibilité d'accès à des écoles de qualité est bien souvent mise en avant. Toutefois, une localisation résidentielle ne permet pas forcément la proximité à la fois de l'école, du collège, du lycée et de l'enseignement supérieur. Nous observons tout de même à l'aide de l'ENTD (CGDD, 2010) que les déplacements domicile-étude sont en 2008 plus courts en distance (5,6 km) que les déplacements domicile-travail. Mais dans certains cas de ménages bi-actifs la congruence (adaptation réciproque) des programmes d'activités de chacun peut devenir vite complexe à mettre en place s'il faut accompagner les enfants à l'école.

Toutefois, l'approche centrée sur l'individu, ses besoins d'accessibilité et ses contraintes de temps permet peu de prendre en compte les évolutions des lieux. Une localisation résidentielle dure plus longtemps que les conditions de mobilité et les besoins de déplacements qui ont influencé ce choix de localisation. On peut alors aborder l'évolution des lieux par le concept d'accessibilité générale. L'idée est d'estimer une accessibilité moyenne à l'ensemble des services d'une agglomération ou d'un bassin de vie pour chaque zone découpant ce territoire. Par exemple, la distance moyenne à tous les emplois pour une zone de résidence peut être mesurée (Wenglenski, 2003). Cet indicateur permet de juger de la proximité des zones résidentielles au marché de l'emploi. Il peut être mesuré pour tous les autres types d'activités. En fait, il s'agit d'un indicateur utilisé depuis très longtemps pour estimer la rentabilité économique de projet de transport (Dupuit, 1844) ou pour évaluer la rentabilité économique des parcelles (Von Thünen, 1826). Le transport ayant un coût, les politiques d'aménagement doivent aussi avoir pour objectif d'aider les ménages, surtout les moins aisés, mais aussi les acteurs économiques, à minimiser les coûts de déplacement. Le suivi temporel de cet indicateur d'accessibilité moyenne et de sa différenciation spatiale

requiert un appareil d'observation basé sur une collecte de données stables et représentatives.

1.1.1.2 L'urbanisation en zones et la solution automobile au problème de l'accessibilité

La notion d'accessibilité générale qui exprime la proximité aux activités est relative à la performance des moyens de transports individuels et collectifs disponibles depuis un lieu. La production d'infrastructures de transport est une opération essentielle d'aménagement du territoire que la puissance publique doit contrôler⁹. La capacité à se déplacer est dépendante de l'offre de transport décidée par les politiques et dépend généralement de décisions d'aménagement passées. Les décideurs ont également d'autres prérogatives. Ils doivent jongler entre les logiques du transport et le développement économique. Cet ajustement est difficile et parfois contradictoire. Pour chaque lieu, les aménageurs peuvent estimer s'il est plus préférable d'y établir une fonction résidentielle ou une fonction économique. Il y a implicitement une concurrence foncière entre les fonctions urbaines alors que la séparation des fonctions produit une demande de déplacements élevée en termes de distance à franchir. Ce paradigme d'aménagement s'appuie sur le principe d'urbanisation en zones, et sur une accessibilité générale élevée, à condition d'être fondée sur la possession massive d'automobiles.

Le principe d'urbanisation en zones théorisé notamment dans la Charte d'Athènes (Le Corbusier, 1957) conduit à rassembler dans l'espace des activités de même fonction. Ce principe permet d'augmenter l'échelle des zones fonctionnelles et d'en maximiser la rentabilité économique. Parmi les critères qui définissent cette rentabilité et conduisent à affecter une zone à une fonction, la distance aux autres types d'établissements que les différentes fonctions ont besoin d'atteindre est un élément essentiel. La distance aux différentes demandes potentielles doit alors se comparer aux coûts de production. Nous parlons alors d'accessibilité. Et cette accessibilité doit être mesurée en prenant en compte les réseaux de transport présents ou futurs. Dupuy propose de parler d'urbanisme réticulaire (Dupuy, 1991). L'idée d'urbanisation en zone est de choisir les types d'activités des zones permettant la meilleure rentabilité économique de par son accessibilité. Les fonctions de même secteur sont regroupées dans l'espace amenant une spécialisation des territoires.

9 C'est un monopole politique dans la plupart des pays mais ça peut être délégué dans des partenariats public-privé

Pour comprendre cette spécialisation des territoires en contexte de croissance économique et démographique, il nous faut faire un rappel sur les stratégies de localisation des établissements économiques en rappelant le principe de spécialisation des établissements économiques. Succinctement, pour minimiser les coûts de production à la fin de la chaîne production-commercialisation des produits manufacturés, les décideurs économiques spécialisent les activités d'extraction, de transformation, de commercialisation, de recherche et développement et d'organisation de cette chaîne. Historiquement, à la suite de la révolution industrielle, les établissements manufacturiers s'implantent en ville et y font augmenter la population. De par leurs proximités spatiales, une concurrence apparaît entre eux. Toutefois, plus les établissements sont nombreux et concentrés, plus les entreprises ont intérêt économiquement à se spécialiser pour obtenir des économies d'échelle. Les entreprises perdent peu à peu leur polyvalence au profit d'une spécialisation accrue les rendant dépendantes du système urbain qu'elles ont elles même créés (économie d'urbanisation).

Pesant de plus en plus sur l'économie urbaine, ces entreprises impriment aux villes grandissantes une spécialisation économique que le pouvoir politique local cherche à renforcer, à partir de la fin du XXe siècle, en intervenant sur les services à la population ou aux entreprises, et sur l'aménagement du territoire. On parle alors de phénomène de compétitivité des territoires.

Les choix d'aménagements urbains des décideurs politiques doivent répondre aux besoins des agents économiques dans ce contexte de spécialisation accrue et de division internationale du travail. On met en place des politiques pour la formation, la garde des enfants, des politiques foncières, fiscales et de transport favorisant l'activité économique et nécessitant une coordination administrative supra locale. Les entreprises profitent de ces externalités et trouvent en plus des entreprises aux activités complémentaires avec lesquelles elles peuvent s'organiser. Ce rapprochement organisationnel nécessite une certaine proximité spatiale. Le besoin d'être proche spatialement dépend du niveau d'intégration et donc de dépendance des entreprises au sein de leurs systèmes de production. Pour un traitement plus complet nous pouvons nous référer à (Fujita, 1989), (Krugman, 1995), (McCann, 200), (Fujita, Krugman et Venables, 2001), (Fujita et Thisse, 2003), (O'Sullivan, 2007), (Polès, Shearmur, Terral, 2015).

Ces mécanismes concentrent les lieux d'emploi dans l'espace urbain, contrairement aux zones résidentielles (illustration 5). En conséquent, un réseau de transport doit permettre de

rapprocher les individus au marché de l'emploi tout en étant utile au développement des liaisons entre les établissements économiques.

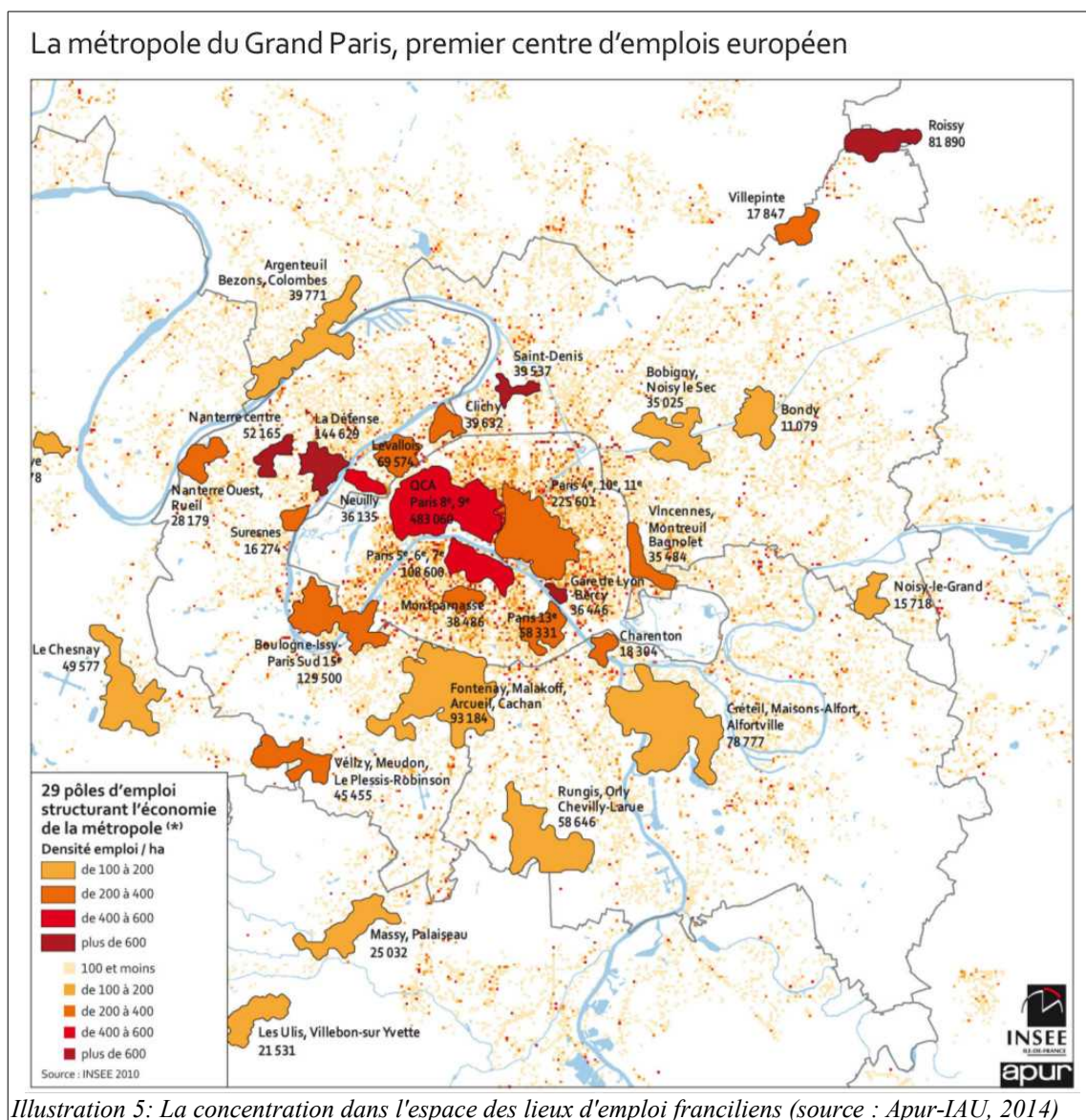


Illustration 5: La concentration dans l'espace des lieux d'emploi franciliens (source : Apur-IAU, 2014)

Pour prendre en compte les aspirations individuelles, le développement économique et la croissance urbaine, les politiques d'urbanisme après la seconde guerre mondiale ont répondu de diverses manières. Le développement des grands ensembles collectifs en bordure des agglomérations urbaines dans le cadre d'une politique nationale centralisée et le développement des lotissements pavillonnaires en périphéries à l'initiative d'acteurs privés ou locaux en sont deux exemples. Notons que dans le cas parisien des politiques de logements

collectifs en périphérie ont été menées afin de répondre également au surpeuplement urbain (DRIEA, présentation du SDRAUP¹⁰). La sous-partie 1.1.1 a présenté le processus de ségrégation socio-spatiale au lieu de résidence dans le cas d'un marché foncier privé, qui s'est appliqué directement aux lotissements pavillonnaires et indirectement aux grands ensembles en bloquant la sortie de ménages trop modestes pour accéder à un logement de meilleure qualité. De par la concurrence foncière, les politiques ont eu la tâche de construire des réseaux de transport permettant de relier les zones de résidence aux lieux d'emploi. Pour ce faire, les politiques se sont appuyées sur la technologie automobile.

En France, les politiques en faveur de l'automobile se sont surtout développées après la seconde guerre mondiale. La loi du 5 Juillet 1949 qui s'inscrit dans le cadre du plan Marshall pour la relance et la restructuration de l'économie française lance, en autres, un programme de développement du réseau autoroutier (Bezes, 2009). A l'intérieur des agglomérations, des infrastructures routières (sur goudron) ont été construites afin de connecter les zones résidentielles et d'emplois. Ces infrastructures ont été pensées pour maximiser l'accessibilité en espérant bien maximiser l'utilité économique des choix de localisations de l'ensemble des fonctions urbaines. De plus, afin de maximiser cette accessibilité, les infrastructures ont été construites d'après un modèle radioconcentrique, qui étale la dynamique de croissance urbaine vers la périphérie. Ce principe d'organisation des infrastructures a consisté à entourer les centres historiques de boulevards périphériques (les anciens remparts) puis de construire des radiales dans le maximum de directions possibles en fonction des contraintes de relief et des niveaux hiérarchiques des agglomérations à rejoindre.

L'automobile et le développement des infrastructures routières ont joué un rôle prédominant dans le processus d'étalement urbain. Premièrement, en reprenant les critères de Dupuy sur les réseaux, elle est un moyen de locomotion incroyablement performant de par sa plus grande « *ubiquité, instantanéité et immédiateté* » relativement aux autres modes de transport. (Dupuy, 1999). Cette technologie a favorisé l'établissement des populations plus loin des noyaux d'urbanisation historiques.

Deuxièmement, la technologie automobile est un produit de consommation sur lequel les États se sont appuyés pour le développement économique. L'exemple emblématique est le

10 DRIEA. Direction Regionale et interdépartementale de l'Équipement et de l'Aménagement d'Île-de-France. [en ligne] (page consultée le 28/03/2016) http://www.driea.ile-de-france.developpement-durable.gouv.fr/presentation-du-sdaurp-de-1965-a970.html#sommaire_3

principe économique du fordisme qui a permis un enrichissement général de la société (en valeur absolue). Ce principe a consisté en partie à payer d'avantage les ouvriers pour qu'ils consomment les biens qu'ils produisent à la chaîne. Les politiques françaises ont soutenu les fleurons industriels automobiles dans les moments difficiles¹¹. En 1965, le taux d'équipement des ménages en automobile atteint la barre symbolique des 50 % (Leroux, 1969).

Plus fondamentalement, il existe une corrélation très forte entre la croissance économique et celle du trafic routier, au point qu'on a parlé de « couplage ». « *La courbe de l'élasticité du taux de motorisation par rapport au PIB par habitant prend la forme d'un S : ce taux stagne lorsque les revenus sont faibles, augmente rapidement quand les revenus dépassent la moyenne et ralentit de nouveau lorsque les revenus atteignent des niveaux élevés* » (OCDE, 2014, p.22). L'automobile est le mode de transport le plus utilisé pour se déplacer. Cependant, sa progression conduit à un point de rupture avec ces avantages. En reprenant les termes de Dupuy, l'« ubiquité », « l'instantanéité » et l'« immédiateté » se sont transformés en « congestion », « asphyxie », et « paralysie ». L'utilisation massive de l'automobile engendre de nombreux désagréments. En présumant que la perte de vitesse a un coût, on peut chiffrer la congestion entre 0,1 % et 0,2 % des produits intérieurs brutes (PIB) nationaux des pays industrialisés (Palma et Zaouali, 2007), (Litman, 2009), (Orfeuil, 2008), (Parry et al. 2007) (Koning 2012).

Les politiques en faveur de l'automobile ont permis à une certaine partie de la population de choisir des localisations périphériques en les enrichissant (ou en facilitant l'obtention de prêts bancaires bon marché) et en les dotant de véhicules personnels. Il semblerait alors que le pari pris par des décideurs politiques soit gagnant. Le paradigme de l'accessibilité générale en voiture personnelle permet de connecter les lieux de résidence et d'emploi. Et elle permettrait de réduire les inégalités socio-spatiales. Une étude en Île-de-France propose la conclusion suivante : « *il n'est pas possible d'affirmer, au moins à l'échelle régionale et sur la base d'une gamme de vitesses issues de l'enquête Transport de 1994, que le réseau routier rapide a ajouté sa propre forme de discrimination à celle qui découle de la pression foncière et du prix du sol urbain. Le maillage de la voirie rapide est suffisamment fin pour ne pas avoir*

11 En 1975, L'état à joué un rôle de négociateur dans la prise de contrôle de Peugeot du groupe Citroën appartenant alors au groupe Michelin. La prime à la casse a été un dispositif international après la crise de 2008 pour soutenir les ventes d'automobiles (rapports de l'OCDE : <https://www.caim.info/revue-perspectives-economiques-de-l-ocde-2009-2-page-93.htm>)

créé un territoire à deux vitesses, si l'on peut dire, par ses conditions d'accès à la vitesse automobile » (Beaucire, 2001, p.346).

L'équipement en automobile est devenu une condition pour accéder aux aménités de la ville. 81 % des ménages sont équipés et 36 % des ménages disposent désormais d'au moins deux véhicules (CGDD, 2010). Le nombre moyen de véhicules par ménage continue donc d'augmenter ces dernières années. « *La voiture tend ainsi à devenir un équipement presque individuel, à l'instar du téléphone* » (CGDD, 2010, p.12).

Toutefois, le taux d'équipement varie en fonction de la localisation résidentielle. Pour une vision globale, reprenons le taux d'équipement selon la classification de l'espace français des espaces urbains de l'ENTD. « *Le nombre de véhicules par adulte varie de 0,8 dans les communes rurales à 0,7 dans les pôles ruraux et les pôles des aires urbaines de moins de 100 000 habitants ; il est de 0,6 dans les centres des aires urbaines de plus de 100 000 habitants et la banlieue parisienne, pour tomber à 0,3 dans Paris. Ainsi, plus les personnes résident près des grands centres urbains, moins elles sont équipées* » (CGDD, 2010, p.12).

Certains précurseurs ont remarqué que les ménages en périphérie sont dépendants de l'automobile (Dupuy, 1999). Dans une étude sur la dépendance automobile en grande couronne parisienne, B. Motte observe que les ménages non-motorisés développent certes, des stratégies de mobilité alternative, mais que de manière générale les ménages non-motorisés restent socialement marginalisés par leur non-participation au système automobile (Motte, 2006). Ils se localisent en priorité dans des lieux concentrant les services du quotidien. Les ménages motorisés mais aux revenus faibles sont également dépendants à l'automobile : « *alors que dans les communes les moins dépendantes de l'automobile leur équipement est particulièrement faible, ces ménages ont des niveaux de motorisation qui tendent à rejoindre ceux des CSP supérieures dans les communes les plus dépendantes. Ces différences très fortes entre les territoires traduisent selon nous l'effort de motorisation considérable auquel les ménages sont contraints pour faire face à une forte intensité de la dépendance automobile. En retour ils semblent contraints de limiter fortement leur mobilité en maîtrisant autant que faire se peut l'augmentation du kilométrage parcourus en voiture particulière.*»¹² (Motte, 2006, p.339)

Cette dépendance est particulièrement sensible chez les actifs occupés. L'ENTD 2008 étudie

¹² Notons également qu'une des conclusions critiques l'obligation de mesurer l'accès aux services avec comme unité de base la commune.

la possibilité de choisir son mode de transport pour aller travailler (CGDD, 2010). 73 % des actifs ayant un lieu fixe utilisent un mode privé (une voiture ou – marginalement – un deux-roues motorisé, comme conducteur ou passager). Parmi ces navetteurs, trois sur cinq déclarent qu'ils n'ont pas le choix de faire autrement (59 %). Ils en sont dépendants. Un sur huit a le choix et s'en sert de temps en temps (12 %). Les autres (29 %) ont le choix mais ne l'utilisent pas.

Le fait d'être dépendant de la voiture¹³ pour aller travailler dépend de la distance à parcourir. *« C'est dans la plage 1-3 km que le choix est maximum : environ deux actifs sur trois déclarent l'avoir et un tiers l'utilise. À partir de 10 km, ils ne sont plus qu'un sur trois avoir le choix, et seulement un sur dix choisissent un mode alternatif »* (CGDD, 2010, p.15). Cette proportion illustre l'absence d'alternative à la voiture.

Rappelons l'introduction générale du rapport du CGDD. *« Parmi les actifs ayant le choix, les raisons mises en avant pour ne pas choisir le mode alternatif à la voiture sont, par ordre d'importance, le temps de transport (40 %), l'inadaptation des horaires des transports en commun, le confort, la fatigue – essentiellement pour le vélo ou la marche à pied. Parmi les autres raisons citées (dans un cas sur huit en première raison) le fait d'utiliser sa voiture pour se rendre à son travail ou de devoir transporter un matériel encombrant est souvent invoqué, ainsi que les conditions météorologiques lorsque le vélo ou la marche sont le mode alternatif »* (CGDD, 2010, p.3).

D'autres raisons peuvent expliquer la préférence de certains à utiliser la voiture. L'automobile n'est pas seulement un instrument technologique, c'est un objet culturel ayant une valeur psychologique pouvant s'apparenter à la liberté, à l'indépendance et parfois au plaisir (Steg, 2003).

L'objectif des politiques environnementales de réduire la part de l'automobile dans les déplacements se heurte donc à de nombreux freins. A. Rocci donne une analyse sociologique des freins au changement de comportements de choix de mode de transports et note que : *« les individus développent des mécanismes d'auto-justification pour expliquer le décalage entre leurs discours portés par des valeurs collectives et leurs propres pratiques de mobilité »* (Rocci, 2007, p.544).

13 C'est ainsi que la question est libellée dans l'enquête.

Aujourd'hui avec les enjeux du développement durable, ces habitudes de mobilité, ces orientations d'aménagement et ces politiques de transport sont remises en cause. Il faut aussi trouver le moyen de « découpler » la croissance économique de celle des trafics routiers, ce qui est encore un autre défi, celui de la transition énergétique et de la croissance verte.

1.1.2 Les nouveaux objectifs des politiques de transport : un changement de paradigme

Alors que les politiques de transport ont, comme on vient de le voir, accompagné la croissance économique et la croissance démographique et spatiale des villes, la prise en compte des enjeux environnementaux est bien plus que le simple ajout d'un critère d'évaluation des projets. Il s'agit en effet de mettre en balance un intérêt immédiat pour l'économie des générations actives et les conditions de vie à long terme des générations futures. Il s'agit bien d'un changement de paradigme pour les politiques de transport, tel que le présente un rapport faisant le bilan de l'activité du groupe « Politiques de transport » du PREDIT-4 (PREDIT, Bilan thématiques¹⁴). Ce changement demande de nouveaux outils : outils scientifiques pour évaluer les objectifs de cette politique, dont l'élaboration a été encadrée par l'ADEME ou le PREDIT, à l'initiative d'appels à projets de recherche (1.1.2.1), outils techniques et légaux à définir par le gouvernement pour évaluer les projets (1.1.2.2) et, condition des deux précédents, outils d'information, c'est-à-dire de nouvelles données encadrées par la statistique publique et le Conseil National de l'Information Statistique (CNIS¹⁵) (1.1.2.3).

1.1.2.1 Comprendre l'objectif du facteur 4

Les nouvelles questions scientifiques ouverte par les perspectives du Grenelle portent sur l'évaluation environnementale et sur les déterminants des pratiques de mobilité. Pour ce qui relève de la problématique du changement climatique : *« le secteur des transports représente une part croissante de la consommation d'énergie finale en France (32 % en 2013 contre 29 % en 1990 et 18 % en 1970) et absorbe près de 70 % de la consommation de produits pétroliers. Il est le principal émetteur de CO₂ avec 36 % des émissions totales (hors*

14 PREDIT : Bilan thématique du PREDIT Groupe 6 « Politiques de transport » [en ligne]. (Consultée le 29/03/2016) <<http://www.predit.prd.fr/predit4/bilan-predit4-go6>>

15 Conseil national de l'information statistique

UTCF), devançant l'industrie et le secteur du bâtiment avec 23 % chacun. En 2013, la route représente plus de 80 % des consommations du secteur des transports [...] Les voitures représentent plus de 60 % des consommations du transport routier ; viennent ensuite les véhicules utilitaires légers avec 20 % et les camions avec 15 % ; les bus (et autocars) et les deux-roues ont un poids très faible (2,5 % et 1,5 % respectivement). » (ADEME, Les chiffres clefs, p.64¹⁶)

Selon le GIEC, les émissions humaines de gaz à effet de serre induisent un réchauffement global de la planète (GIEC, 2013). Pour y faire face, la France s'est engagée à réduire ses émissions de gaz à effet de serre :

- de 20% entre 1990 et 2020 en tant que pays-membre de l'Union européenne
- de 75% entre 1990 et 2050

L'objectif est de réduire par 4 les émissions (loi POPE du 13 juillet 2005 et travaux du « Grenelle »). Le Conseil Général de l'Environnement et du Développement Durable (CGEDD) a publié en février 2013 un rapport intitulé « *Le facteur 4 en France : la division par 4 des émissions de gaz à effet de serre à l'horizon 2050* », sous la direction de Jean-René Brunetière (CGEDD, 2013). Ce rapport apporte des éclairages sur les possibilités et les moyens d'atteindre les grands objectifs français permettant de faire face aux enjeux climatiques.

La réduction des émissions de gaz à effet de serre concerne les secteurs des transports, de l'industrie, de l'agriculture et de l'habitat. La spécificité du secteur du transport est que les baisses demandées ne peuvent pas faire l'objet seulement de la mise en place de nouvelles normes s'imposant aux consommateurs. La baisse requiert des changements dans les mobilités du quotidien, ce qui implique de modifier les enchaînements de programmes d'activités des individus. Les changements de technologie ne permettent pas, à eux seuls, d'atteindre les objectifs politiques de facteur 4.

En effet, les différentes études prospectives sont pessimistes. La diminution des gaz à effet de serre serait seulement de l'ordre d'un facteur de 2 ou 2,5. Le CGEDD s'appuie sur les rapports « *Prospective 2050* » du Conseil général des ponts et chaussées (CGPC, 2006), le rapport de la commission énergie du Conseil d'analyse stratégique (CAS, 2007), et le rapport du Programme de recherche et d'innovation dans les transports (PREDIT), du Laboratoire

16 ADEME. Les chiffres clefs. [en ligne] (Consultée le 01/02/2015) <www.ademe.fr/chiffres-cles-climat-air-energie-2014>

d'économie des transports (LET) et d'Enerdata (Crozet et al., 2008).

Les premières pistes pour abaisser la consommation de pétrole et les émissions de gaz à effet de serre sont de faire évoluer les technologies de transport en rendant les véhicules plus propres. Depuis les débuts du PREDIT, des financements sont concentrés sur la mise au point de technologies plus performantes énergétiquement. Pour favoriser l'utilisation de ces technologies propres l'État doit alors jouer un rôle d'incitateur.

« La France a opté pour la mise en place d'un système de bonus/malus écologique en 2008 qui ré-compense via un bonus, déduit du prix d'achat, les acquéreurs de voitures neuves émettant le moins de CO₂. À l'inverse, un malus est appliqué à tout véhicule émettant plus de 131 gCO₂/km. Ce dispositif de bonus / malus a eu un impact significatif sur l'évolution du marché et sur les émissions moyenne de CO₂ des véhicules particuliers. » (ADEME, Les chiffres clefs, p. 64).

Mais, selon le rapport « Prospective 2050 : *« De forts progrès techniques des véhicules économes et des carburants propres incités par une taxe carbone à un niveau approprié, permettraient une division par 2,5 des émissions françaises de CO₂ à l'horizon 2050 »* (CGEDD, 2013, p.16)

Signalons toutefois que le domaine des transports urbains apparaît moins problématique pour les auteurs de ce rapport.

- « Les émissions dues aux déplacements en centre-ville pèsent peu et les villes savent comment les réduire. La majorité des kilomètres parcourus en véhicule individuel et donc la majorité des émissions sont dus en fait à des trajets reliant deux points de la périphérie ou la périphérie au centre : en effet, la motorisation électrique est susceptible d'apporter de grands bouleversements dans ce créneau, à la fois dans les types de véhicules, dans la structure des services de déplacement et dans les régimes de production et d'utilisation de l'électricité. La «décarbonation» de ce créneau ne semble pas présenter de limite à long terme. Seul le rythme de déploiement de ces solutions est encore incertain» (CGEDD, 2013, p.18).

Ce rapport paraît optimiste étant donné le coût de ce changement technologique pour les finances des collectivités : ce ne serait qu'une question de temps pour que l'on assiste à un déploiement de solutions nouvelles. Mais pour que les individus utilisent ces nouvelles technologies de transport, peut-être faudrait-il qu'ils changent aussi leurs comportements de consommation.

La question de recherche suivante, également posée dans un appel du PREDIT, est d'identifier les leviers des changements de pratique de mobilité. Une première réponse est apportée par le projet *Scénarios de mobilité durable à l'échelle européenne sous contrainte du facteur 4 : horizon 2050* associant le LET et Enerdata dans une approche économique appliquant la méthode de Backcasting.

-« Cette méthode consiste à définir une situation de départ située dans le futur (soit 2050) et de travailler en aval [...] pour déterminer les scénarios pouvant répondre aux objectifs de la situation de départ dans le présent» (Lopez-Ruiz,2009).

Il s'agit alors de déterminer sur quels critères d'action jouer pour parvenir à une baisse des émissions de gaz à effet de serre. La recherche développe trois scénarios contrastés :

- Le premier, nommé « Pégase », s'appuie sur la poursuite des tendances actuelles. Il n'y a aucun changement de mode de vie et la baisse des émissions ne peut résulter que du progrès technologique.

- Le deuxième scénario, intitulé « Chronos », propose d'instaurer une augmentation du prix du transport routier. La route devenant trop chère pour certains déplacements (longue distance, marchandise) un besoin d'amélioration des autres systèmes de transport plus durables devenus plus compétitifs pour les usagers apparaît dans la simulation. Dans ce cas, la diminution des vitesses de déplacements est un facteur important dans les changements de mode de vie. Ce scénario implique une diminution de 20 % des vitesses du fait de l'emploi d'autres modes de transport que la voiture.

-Le troisième scénario, baptisé « Hestia »), accorde une initiative de plus à l'ensemble des acteurs en supposant qu'ils peuvent se relocaliser afin de réduire leurs coûts de transport.

Ces trois scénarios donnent ainsi de plus en plus de responsabilités aux ménages. Le premier (Pégase) demande une optimisation des systèmes de transports existants. Le deuxième (Chronos) ajoute l'hypothèse d'une forte taxation de la route et d'une tolérance à un ralentissement des vitesses de déplacement. Le troisième (Hestia) va encore plus loin en supposant une démarche éco-citoyenne proactive des ménages et des entreprises qui se relocaliseraient pour diminuer les distances parcourues.

Ces prévisions s'appuient sur un modèle macro-économique. Les contraintes locales et les relations transport-urbanisme sont sous-spécifiées (Lakshmanan et Anderson 2004). C'est suite

aux travaux du PREDIT-4 de 2008 que Lopez-Ruiz a effectué sa thèse en suivant une approche de modélisation moins agrégée des changements de comportements, basée sur un algorithme de fourmis pour générer plus finement les déplacements sur les réseaux (Lopez-Ruiz, 2008¹⁷). La question posée est intéressante : il nous faut observer plus finement spatialement et socialement les individus pour mieux savoir comment agir pour changer leurs comportements.

L'identification des leviers pour agir sur les pratiques et les comportements est une nouvelle étape dans ce changement de paradigme : il faut trouver de nouveaux leviers d'action sur les comportements, et mettre en place un dispositif de suivi et de contrôle des mobilités et de leurs impacts environnementaux.

Le club d'ingénierie prospective énergie et environnement (CLIP, 2012) analyse ainsi les dépenses énergétiques selon les profils des ménages, permettant de cibler les comportements. (présentation¹⁸) On y retrouve la relation classique entre les hauts revenus et les fortes consommations énergétiques. Ces travaux se basent sur l'Enquête Nationale Transports et Déplacements du SOeS et les enquêtes budgets des familles de l'Insee. Le principe est différent des travaux précédents. Il s'agit d'établir une situation de référence et de faire des projections selon divers scénarios. Quatre scénarios de changement de comportement sont proposés et conduisent à énoncer dix leviers d'action :

- « *Les leviers [...] qui semblent offrir une certaine efficacité dans la lutte contre le changement climatique, sont :*

- *La réduction des surfaces disponibles par personne: cela nous pousse alors à considérer que la tendance actuelle à l'augmentation des surfaces doit s'inverser. Cela passe par une adaptation des logements à la taille du ménage tout au long de la vie (il faut alors faciliter les mobilités résidentielles) et par la promotion de nouveaux modes de cohabitation. Les ménages retraités pourraient être les premiers concernés par des mesures dans ce sens.*

17 Malheureusement, il ne cite pas les données utilisées pour effectuer ces travaux de modélisation des changements de comportements.

18 Le CLIP a été créé en janvier 1993. Ses membres actuels sont : l'Ademe, le BRGM, le CEA, le Cirad, le Citepa, le CNRS, le CSTB, EDF, la Fnade, GDF, l'IFP, l'Ineris, l'INRA, l'Inrets, le GIE Recherche PSA Renault, l'ONE, la SNCF et le SNET. Participent au comité de coordination et d'orientation scientifique les ministères chargés de la Recherche, de l'Environnement, de l'Industrie, des Transports et du Logement, de l'Agriculture, ainsi que le Commissariat général du Plan.

- *Le maintien de consignes de température raisonnables dans les logements, malgré les travaux de réhabilitation réalisés : cette observation nous conforte dans l'idée que les utilisateurs d'un bâtiment réhabilité doivent réapprendre à s'en servir.*
- *La résistance face à l'augmentation de l'équipement électrique dans les ménages.*
- *La mutualisation des usages de manière générale.*
- *La restauration des liens sociaux de proximité.*
- *La réduction des distances quotidiennes parcourues par la réduction des portées des déplacements.*
- *L'usage quotidien des modes doux et des transports en commun.*
- *La réduction drastique des distances parcourues en mobilité exceptionnelle.*
- *La réduction des quantités de viande consommées*
- *La réduction forte du gaspillage alimentaire.» (CLIP, 2012,p.117)*

Les auteurs sont conscients de la forte pression qui serait appliquée aux individus s'ils devaient suivre ces dix points. On remarque que la moitié de ces mesures concerne la mobilité. Les auteurs s'interrogent également sur la façon de suivre l'évolution des comportements pour mieux diriger l'application des politiques dans le temps. Leur conclusion est la suivante :

- « Au terme de cette recherche, il nous semble primordial d'engager une réflexion sur la capacité à rendre « désirable » un monde où la mobilité des personnes serait fortement réduite et sur les conditions nécessaires à l'évolution des habitudes culturelles telles que le régime alimentaire ou les pratiques de consommation.

En effet, aucun levier ne pourra être laissé de côté si nous souhaitons maintenir le réchauffement global de la planète sous les 2 °C d'ici la fin du siècle. Les changements de modes de vie devront donc faire partie des solutions promues par les pouvoirs publics.

Par ailleurs, la qualité de l'exercice que nous venons de réaliser est, pour une part importante, dépendante des données que nous sommes parvenus à acquérir. Une piste d'amélioration de ce travail, notamment en vue d'un possible suivi dans le temps, serait alors d'adapter l'outil statistique à ces nouveaux besoins de connaissance et de description

des modes de vie des ménages français. Les difficultés rencontrées sont alors de deux ordres :

- un manque de données statistiques pour caractériser les populations en fonction de leurs « modes de vie » ;*
- un manque de données détaillées, centralisées et cohérentes sur les émissions de GES indirectes des consommations d'alimentation et de biens. Concernant la caractérisation des populations en fonction de leurs modes de vie, nous proposons ci-dessous un certain nombre d'éléments qui nous auraient permis, si nous avions pu en disposer, d'améliorer substantiellement cette étude :*
- La base « individus » de l'Insee, tirée du recensement de la population, caractérise chaque personne en fonction de sa CSP d'appartenance. En revanche, aucune information n'est disponible permettant de caractériser finement les ressources financières de ces personnes et des ménages qu'elles composent. L'intégration dans la base «individus» d'une donnée caractérisant les revenus des personnes constituerait une nette amélioration.*
- Les enquêtes réalisées sur les mobilités des personnes ne permettent pas de définir des habitudes de mobilité en fonction des modes de vie. En effet, les échantillons enquêtés deviennent trop faibles pour être exploités, dès que les populations observées sont caractérisées par 2 ou 3 paramètres : typologie urbaine du lieu d'habitation, âge des personnes, revenus, structure familiale. Un élargissement de l'échantillon enquêté serait alors une réelle plus-value pour le type d'étude que nous venons de mener.[...] » (CLIP, 2012, p.118).*

Les propositions semblent raisonnables d'un point de vue scientifique mais ne prennent pas en compte les conditions pratiques de production des données de qualité qui expliquent pourquoi ces objectifs sont en fait peu réalisables. Ce commentaire sur le manque de précision des données, et en particulier des données de mobilité revient dans d'autres travaux cherchant à évaluer l'impact social des changements de comportements requis par les objectifs environnementaux. Ainsi, nous trouvons le commentaire suivant dans l'article de Nicolas (Nicolas, 2012) proposant de mesurer la vulnérabilité des ménages par rapport à l'augmentation des coûts de transport automobiles, qu'elle résulte de la hausse du prix des matières premières ou de celle des taxes sur les carburants :

- « Pour construire statistiquement et analyser un tel taux d'effort budgétaire des ménages pour leur mobilité quotidienne, il est nécessaire de posséder à la fois des données socio-économiques fines sur les ménages (composition du ménage, activité, revenus, endettement), le contexte territorial dans lequel ils évoluent (localisation résidentielle, localisation des activités, offre de transport) et les pratiques de mobilité associées (budgets distances, modes utilisés). Or, il n'existe pas de recueil statistique spécifique permettant de fournir l'ensemble de ces données. Les travaux méthodologiques de GALLEZ (2000) et NICOLAS et al. (2001) proposent de contourner cette difficulté en utilisant de manière complémentaire les enquêtes Budget des Familles, les enquêtes Ménages-Déplacements locales et les Enquêtes Nationales Transport» (Nicolas, 2012, p.23).

En raison des données utilisées, l'essai de mise en lumière des ménages les plus vulnérables apparaît très agrégé dans les zones périphériques (illustration 6).

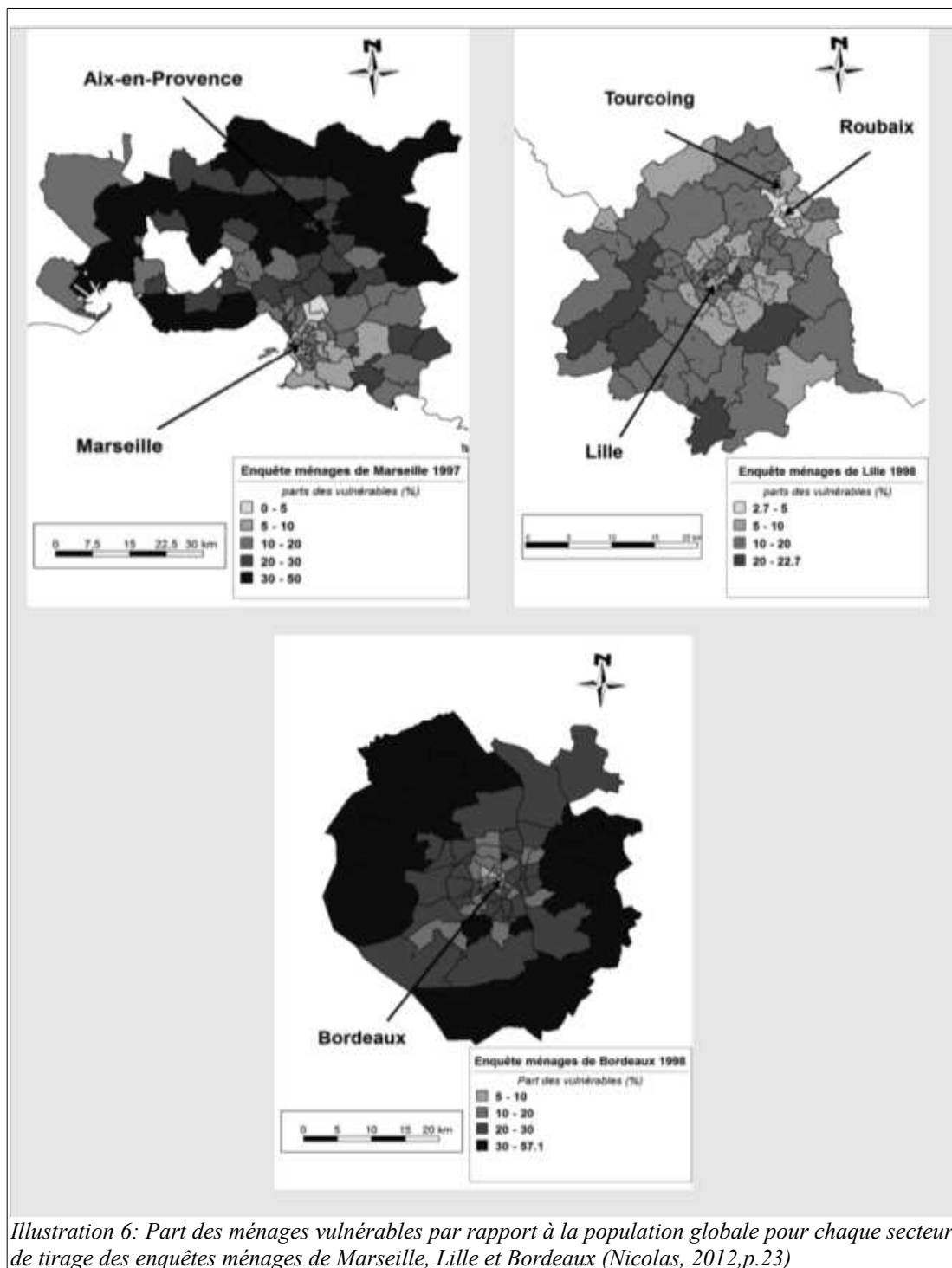


Illustration 6: Part des ménages vulnérables par rapport à la population globale pour chaque secteur de tirage des enquêtes ménages de Marseille, Lille et Bordeaux (Nicolas, 2012,p.23)

Toutefois, contrairement aux travaux de macro-économie du projet LET-Enerdata, ou aux projections basées sur l'ENTD 2008 du club des ingénieurs, ces travaux demandent une meilleure précision des localisations géographiques dans l'examen des changements de

comportement de mobilité à opérer. En effet, l'amélioration de la précision géographique des données d'enquêtes permet de mieux savoir où accompagner les politiques, par exemple en développant les systèmes de transports alternatifs à l'usage de la voiture particulière. En continuant notre raisonnement sur les changements de comportements, une fois les ménages vulnérables mieux localisés grâce à une donnée précise sur la demande de déplacement, les politiques de transport peuvent être mieux ciblées socialement et spatialement. Dans ce nouveau paradigme, les objectifs des politiques concernant les transports prennent le sens plus large de politiques de la mobilité, comme proposé par le Grenelle de l'Environnement, et intègre des critères bien plus variés qu'auparavant. On peut néanmoins s'interroger sur la possibilité de parvenir en même temps à prendre plus de critères en compte et à raisonner au niveau le plus fin dans la mesure où les phénomènes pris en compte font système. Ils ont donc tendance à s'ajuster les uns aux autres, ce qui peut rendre vaines les tentatives pour les prédire précisément.

1.1.2.2 La réduction de la part modale de la voiture dans l'évaluation des projets de transport

Pour aller vers la réduction des émissions de gaz à effet de serre, la loi de programmation relative à la mise en œuvre du Grenelle de l'Environnement du 3 août 2009 prévoit, en son article 13, la participation de l'État au développement de 1 500 kilomètres de voies nouvelles de transports collectifs en site propre (hors Île-de-France). Pour que ces infrastructures soient rentables, il faut les penser de manière à répondre correctement à la demande de transport dont nous venons d'identifier les contraintes et les besoins. Mais la rentabilité doit maintenant se comprendre dans le nouveau paradigme issu du Grenelle et intégrer la réduction de la part modale de la voiture émettrice de CO₂, ce que nous allons examiner maintenant.

La question des reports modaux nécessite d'analyser les multiples facettes de la mobilité et en particulier la concurrence entre les offres. Auparavant les politiques de transport s'intéressaient exclusivement à la distance-temps et à son coût. Aujourd'hui les enjeux environnementaux imposent de mesurer également l'efficacité des systèmes de transport (efficacité énergétique de la chaîne multi-modale, qualité de vie : temps d'attente, confort, confiance des individus accordée aux TC, pénibilité de la marche à pied).

L'instruction du Gouvernement du 16 juin 2014 relative à l'évaluation des projets de transport présente les nouveaux objectifs des évaluations des projets qui apparaissent comme un autre aspect du changement de paradigme.

« Le cadre général d'évaluation comporte trois volets :

– une analyse stratégique, définissant la situation existante, le scénario de référence, l'option de référence qui aurait prévalu sans le projet, les motifs à étudier l'éventualité d'agir, les objectifs du projet, les options de projet ;

– une analyse des effets des différentes options de projet, portant, de manière adaptée et proportionnée aux enjeux et effets envisageables du projet, sur les thèmes sociaux, environnementaux et économiques ;

– une synthèse, présentant les estimations sur le niveau d'atteinte des objectifs et sur les effets des différentes options de projet ; la synthèse peut être déclinée par territoire et par catégorie d'acteurs, en considérant au moins, à ce dernier titre, les usagers directs du projet de transport et les finances publiques.

Une note technique de la direction générale des infrastructures, des transports et de la mer (DGITM) présente la méthode pour appliquer le cadre général d'évaluation, qui est complétée de fiches-outils», (Bulletin officiel, Instruction du Gouvernement du 16 juin 2014, p.2¹⁹).

A l'aide des fiches-outils du référentiel d'évaluation des projets de transport cité dans l'instruction (Ministère de l'environnement, de l'énergie et de la mer, Les fiches outils du référentiel d'évaluation des projets de transport²⁰) nous remarquons que les effets à considérer aujourd'hui sont plus nombreux. Le ministère propose treize effets à étudier lors d'évaluation de projets. Précisons que ces fiches apportent des éléments méthodologiques pour évaluer ces effets. En ce qui concerne les données pour évaluer la demande de transport, le ministère insiste sur le fait de devoir estimer le nombre d'utilisateurs nouveaux que le projet d'aménagement induira. C'est-à-dire, pour une aire géographique donnée, il faut pouvoir connaître la part des déplacements effectués selon tel ou tel mode avant et après l'aménagement. Cet indicateur est nommé « part modale ». Cet indicateur est toujours à

19 Bulletin officiel. [En ligne] (Consultée le 28/03/2016) <http://www.bulletin-officiel.developpement-durable.gouv.fr/fiches/BO201412/met_20140012_0100_0047.pdf>

20 Ministère de l'environnement, de l'énergie et de la mer. Les fiches outils [En ligne] (Consultée le 28/03/2016) <<http://www.developpement-durable.gouv.fr/Les-fiches-outils-du-referentiel-d.html>>

mettre en relation avec sa zone géographique. Les moyens présentés dans les fiches pour calculer les parts modales sont les enquêtes et la modélisation de transport. La fiche présente ces méthodes de manière très sommaire. Les modèles de transport ont comme objectif de déterminer la part modale au sein de la zone avant et après l'aménagement. Pour calibrer les modèles sur la situation avant l'aménagement on se sert de données d'enquêtes de déplacement.

Notons que les fiches précisent un fait important sur la collecte des données: « *La collecte des données d'entrée des modèles est une étape qui nécessite une attention particulière puisqu'elle conditionne la qualité des résultats en sortie des modèles. Il faut toutefois noter que la collecte des données est généralement longue et fastidieuse* ». La fiche renvoie également à l'utilisation des données du recensement pour caler les modèles de transport.

Une fois le projet réalisé, il faut procéder à ce que le ministère appelle l'évaluation ex-post. La fiche-outil « *Données à recueillir pour l'évaluation ex-post (Octobre 2014)* » s'appuie sur une référence consacrée au transport sur site propre (CEREMA, 2015). Il est souhaité reproduire la chaîne de modélisation 3 à 5 ans après le projet de transport réalisé. La fiche insiste sur le fait d'utiliser les mêmes modèles qu'ex-ante pour des soucis de cohérence. Les auteurs (non précisés) ont cru utile d'écrire : « *Il est important de conserver le modèle et ses fichiers pour une réutilisation opérationnelle au moment de l'étude ex-post* ». Comme si, par retour d'expérience, on avait constaté que le stockage des données numériques pouvait défaillir.

Les fiches-outils du ministère traitent seulement des transports collectifs mais nous supposons que la démarche d'évaluation est la même pour les autres types d'infrastructures. En ce qui concerne les modes doux (marche à pied et vélo) qui sont par définition les plus durables et qui maintiennent en bonne santé (Morris et Hardman, 1997), leurs utilisations doivent être favorisées par un rapprochement «des lieux de vie», des lieux de résidence et d'emploi (illustration 7).

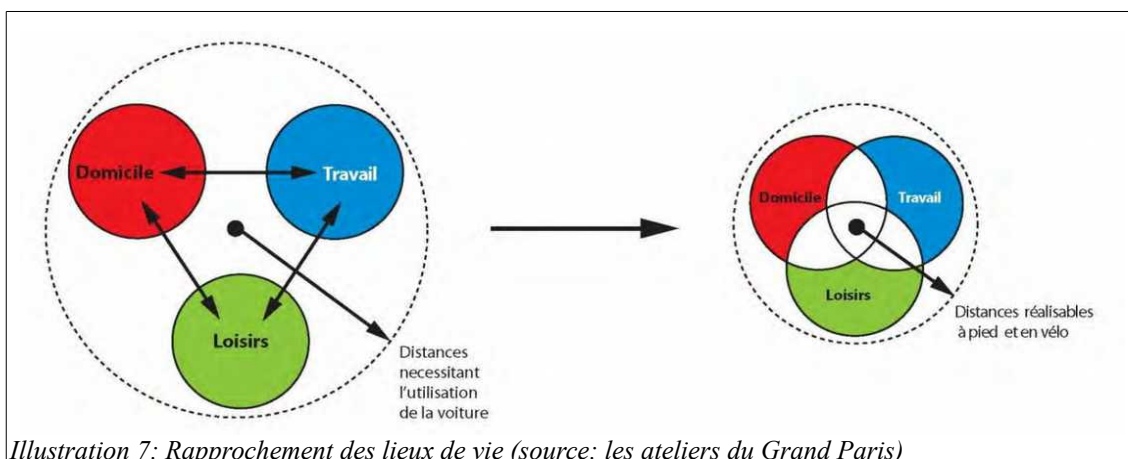


Illustration 7: Rapprochement des lieux de vie (source: les ateliers du Grand Paris)

Cet objectif peut toutefois être confronté à certaines conclusions mettant en garde sur l'hypothèse toute faite que les individus souhaitent habiter à proximité de leur lieu de travail (Korsu, 2006). Rappelons également que les localisations à atteindre sont multiples et doivent être en accord avec les besoins de chacun des membres des ménages. Il ne s'agit donc pas de connecter simplement les zones de résidence et économique à proximité par des infrastructures pédestres. Il faut faire de véritables circuits desservant au mieux l'espace optimisant l'accessibilité générale en marche à pied.

La démarche de modélisation et de simulation avant et après les aménagements de transports concerne également les aménagements pédestres, cyclables et aussi l'électro-mobilité, le co-voiturage et l'auto-partage qui font intervenir de nouveaux acteurs. La mise en place de ces nouveaux systèmes de transport doit être effectuée en ayant au préalable bien déterminé leurs usagers potentiels. Il faut les rentabiliser le plus rapidement possible et qu'ils soient accessibles à tous. Des éléments de connaissance sur la demande de déplacement sont donc nécessaires. Des données décrivant les aires géographiques desservies par les projets de transport doivent être disponibles pour mener à bien ces politiques de transport.

Les objectifs des évaluations ex-ante et ex-post des projets de transport donnent une place centrale à la décomposition modale des déplacements au sein des zones concernées par les projets de transport. Ces objectifs sont donc localisés et leurs indicateurs dépendent de la définition des zones d'étude. Or, comme on le verra plus loin, le périmètre choisi peut faire considérablement varier les indicateurs et leurs évolutions. Dans une autre fiche méthodologique le ministère présente les principes de définition d'une aire géographique pour

effectuer les mesures de mobilité (Ministère de l'environnement, de l'énergie et de la mer. Aires d'études²¹). Pour les projets de Transport en Commun Urbain (TCU), il convient de définir deux périmètres d'études: une aire d'étude restreinte et une aire élargie. Pour le premier, les quartiers politiques de la ville desservis par le projet de transport sont données en exemple. Le deuxième type de périmètre regroupe quant à lui l'ensemble des quartiers politiques de la ville qui pourront se rabattre sur le projet de TCU et qui bénéficieront alors d'une meilleure accessibilité aux aménités. Dans le cas de parcs relais créés en bout de lignes de TCU, il faut également s'intéresser aux territoires d'origine des usagers de ces parcs. Toutefois, nous ne pouvons pas connaître les territoires d'origine des usagers des éventuels parcs relais avant leur création. Ainsi, la définition d'une aire d'étude ex-ante est souvent difficile. Il faudra faire des hypothèses pour délimiter les zones d'études en sachant que selon les découpages choisis les effectifs dénombrés ne seront pas identiques, les résultats d'analyse de mobilité seront différents. Idéalement, plusieurs aires géographiques d'analyse sont à utiliser.

La possibilité de mesurer les parts modales avant et après sur une zone définie demande à faire attention à l'information zonale dont on dispose. Un modèle de transport dit agrégé n'informe que sur des évolutions de zones émettrices ou réceptrices de trafic. Dans ce cas, les données ne renseignent pas sur les phénomènes d'adaptation des ménages. En aucun cas nous savons si ce sont les ménages qui ont changé de mode de transport ou si les évolutions des parts modales des zones sont dues au renouvellement des ménages par mouvement naturel ou migratoire. Autrement dit nous ne savons pas si les gens se sont adaptés en changeant de mode de transport, en changeant de lieu de résidence par les mécanismes de la pression foncière, en disparaissant ou en arrivant avec l'intention de pratiquer une mobilité différente. Pour en savoir plus, il nous faut alors des données sur les zones avec, si possible, des données sur les migrations résidentielles. Les aires géographiques d'analyse doivent en outre être définissables sur tout le territoire national. Cela fait une grande quantité de données à obtenir.

21 Ministère de l'environnement, de l'énergie et de la mer. Aires d'études[En ligne] (Consultée le 28/03/2016) < http://www.developpement-durable.gouv.fr/IMG/pdf/III_-_Aires_d_etude_V_01_10_2014.pdf >

1.1.2.3 L'appel aux données pour connaître et agir

Parallèlement à cette demande pour plus de données, nous voyons se développer un mouvement de libéralisation des données sur la demande de transport au même titre que sur l'offre de transport (Jutand, 2015). Ce mouvement peut être désigné sous le terme « open data ». Illustrons cet appel à la libéralisation des données sur la demande de transport avec un autre rapport ministériel issu du Grenelle de l'Environnement de 2007, celui du comité opérationnel consacré à la recherche, composé de représentants de la recherche académique, des entreprises et de parlementaires qui se sont interrogés sur les recherches scientifiques à mener.

- « Pour être en mesure d'agir, la recherche doit mobiliser des données, des modèles, des outils et des compétences. Afin de répondre aux enjeux de la biodiversité, des risques sanitaires et du changement climatique, la recherche doit se traduire en moyens d'action et doit permettre le développement de stratégies innovantes. Ces stratégies doivent répondre à un double objectif : inverser les tendances actuelles (émission de gaz à effet de serre, pression sur l'environnement et augmentation des risques) et s'adapter aux contraintes de notre environnement (perturbations climatiques, disponibilité énergétique, limitation des ressources...). Les propositions du comité opérationnel Recherche concernent à la fois les méthodes et les stratégies innovantes. » (Guillou et Colombani, 2008, p.6)

La sous-section consacrée à la Ville, l'Urbanisme et l'Habitat a soulevé les points suivants :

- « Fédérer des équipes autour du thème de la modélisation environnementale de la ville, c'est-à-dire de la simulation du fonctionnement des zones urbaines à plusieurs échelles, en interaction avec les évolutions globales.*
- Développer des systèmes d'observation urbaine interdisciplinaires, permettant de fédérer les efforts, de capitaliser les données et de favoriser les échanges et donc les collaborations. Les moyens modernes (exemple: géolocalisation) devrait permettre de renouveler les processus de collecte de données, une fois mises en place des solutions compatibles avec le respect de la vie privée, à l'instar de ce qui est fait dans le domaine de la santé*
- Développer la modélisation comme outil d'aide à la concertation et à la décision, et comme outil d'évaluation des politiques publiques, etc.*
- Favoriser l'élaboration de scénarios de l'évolution des systèmes urbains par des travaux*

multidisciplinaires intégrant toutes les dimensions de l'évolution urbaine, afin d'étudier les effets sur les comportements des variations des prix (immobilier, énergie), les effets des infrastructures de transport (routes, métros, tramways, sites propres bus)

-l'importance de disposer de données pertinentes et régulièrement mises à jour sur les attentes et comportements des habitants ; une équipe d'une quinzaine de personnes devrait être spécifiquement mise en place à cet effet pour un coût annuel récurrent de 2M€.

-Le pôle de Marne la Vallée, qui se structure autour du PRES Paris-Est, du pôle de compétitivité « Advancity », du pôle scientifique et technique francilien du MEEDDAT. C'est au sein de ce pôle que pourraient se regrouper et se structurer une action majeure sur le thème «modélisation, bases de données, métrologie ».

Ce dernier point sur la métrologie mérite quelques explications supplémentaires. La métrologie est la science de la mesure. Dans l'esprit du comité de recherche, la métrologie s'applique de manière très concrète, par exemple au système d'éclairage urbain. En allant plus loin, la métrologie définit les principes et les méthodes permettant de garantir et maintenir la confiance envers les mesures résultant des processus de mesure. Il s'agit d'une science transversale qui s'applique dans tous les domaines où des mesures quantitatives sont effectuées. Cette référence à la métrologie signale donc que les données doivent être issues de méthodes de mesures scientifiques. Or, en reprenant l'exemple des données sur les parts modales, les effectifs utilisant les transports en commun seront différents selon la référence spatiale choisie pour l'aire géographique ciblée par le projet. Les données relatives sont des artefacts, il faut que celles-ci soient cohérentes (Desrosières, 2003), ce qui dépend des méthodes de construction. En reprenant la distinction de Cirillo et al. (2004), nous pouvons classer les données en deux types, celles issues des enquêtes de terrain (données descriptives) et celles issues de traitement de données post-enquêtes (données explicatives) qui relèvent d'un processus de modélisation.

Le plus souvent les données statistiques de mobilité sont construites à l'aide d'enquêtes de terrain qui constituent le meilleur moyen pour observer la complexité de la mobilité (Cirillo et al., 2004). Mais cela coûte vite très cher surtout si la rentabilité économique est demandée. Alors pour disposer de plus d'information, la construction de données explicatives à partir de sources d'information différentes peut être utile. La question de la métrologie interroge donc sur les limites de ces méthodes de collecte et d'enrichissement. Jusqu'à quel point pouvons-

nous effectuer des traitements post-enquêtes sur les données ? Le rapport du Grenelle de l'Environnement invite à développer la modélisation, à produire des données, tout en disposant de données pertinentes. Le rapport traite implicitement de données issues d'enquêtes. La question est de savoir utiliser ces deux méthodes à bon escient en précisant bien ce sur quoi les données doivent renseigner.

En 2013, le PREDIT-4²² lance l'appel à projet intitulé « *Connaître pour mieux agir* » qui développe cette question :

- « Ce projet s'inscrit dans les actions « Mobilités ». Il vise à améliorer nos connaissances sur les pratiques réelles pour mieux concevoir de nouvelles solutions de mobilité multimodales en identifiant les contraintes (notamment en termes d'organisation spatiale), en facilitant la réorganisation des activités des entreprises et des ménages. Ce projet a pour principal objectif de réaliser un prototype qui valide d'une part l'acceptabilité des acteurs dont les citoyens, et d'autre part les dispositifs technologiques, pour créer et mettre en forme de nouvelles connaissances sur les pratiques quotidiennes de mobilités. Il permettra également, sous certaines conditions, à des acteurs économiques (bureaux d'études, consultants, fournisseurs de services de mobilités) d'industrialiser le processus. Par ailleurs, le projet doit apporter des éclairages sur les points suivants :

- L'accès à de nouvelles connaissances sur nos pratiques de mobilités est-il actuellement techniquement possible, à quel coût ?*
- En quoi de nouvelles représentations, via des outils adaptés, des mobilités, modifient les pratiques pour les usagers et les mesures à mettre en œuvre pour les acteurs de la décision ?*
- Comment ce projet mobilise-t-il les différents interlocuteurs : citoyens, décideurs politiques locaux et nationaux, industries utilisatrices de solutions de transport et industries créatrices de solutions de transport ?*
- Quelles sont les modalités pour mesurer l'évolution des pratiques des citoyens en fonction du contexte, de l'offre et des contraintes ? » (PREDIT. Connaître pour mieux agir²³).*

La première question pose la question de la valeur ajoutée des données plus précises. Nous

22 Le PREDIT-4 a été également lancé par le Grenelle de l'Environnement.

23 PREDIT. Connaître pour mieux agir [en ligne] (Consultée le 03/03/2016)
<<http://www.predit.prd.fr/predit4/appel-a-proposition/44471>>

commençons par répondre à la quatrième question sur les modalités des données dont nous disposons pour mesurer l'évolution des pratiques des citoyens (partie 1.2). Puis, nous nous interrogerons sur les nouvelles approches pour produire des données sur les pratiques de déplacement (partie 1.3).

1.2 Les données sur la mobilité dans la perspective des nouveaux besoins des politiques de transport

Comme invité à faire l'appel à projet du PREDIT-4, Cette partie interroge les modalités de production des données de déplacements. L'analyse porte, dans un premier temps, sur le choix des attributs et des supports géographiques des données représentant les individus dans leurs territoires (partie 1.2.1). Dans un deuxième temps, les données actuellement disponibles sur la mobilité sont présentées (1.2.2).

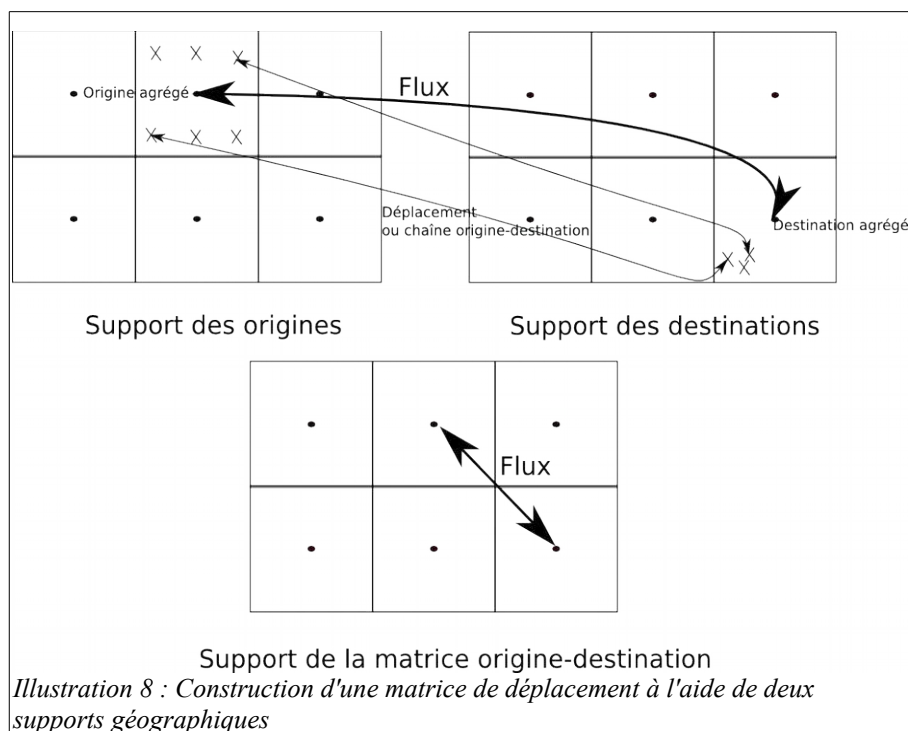
1.2.1 Les attributs et supports géographiques souhaités pour les données de déplacements dans le cadre du développement durable

La richesse des attributs décrivant les individus et leurs déplacements a un coût et est aussi limitée par les règles de respect de l'anonymat. Le modèle (l'archétype) des données souhaitées dans le cadre du développement durable demande néanmoins une résolution spatiale fine et un certain raffinement des variables socio-économiques collectées pour construire des profils socio-économiques pertinents. Les supports géographiques doivent être relativement précis et offrir de nombreux attributs. Nous examinons d'abord la contrainte due au choix du support.

Les besoins exprimés par les fiches outils du ministère demandent la délimitation d'aires géographiques d'analyse cohérentes, restreintes ou élargies. Or, la façon de répondre à cette question détermine en partie les exploitations qui seront faites des données car selon les périmètres choisis, les mesures des distances, des parts modales et des autres indicateurs vont varier. Dès lors, comment garantir la confiance envers les mesures à faire dans les projets d'évaluation de transport? Cela amène-t-il à reconsidérer l'approche des fiches outils ? Peut-on répondre aux besoins exprimés par un seul, éventuellement deux référentiels géographiques d'analyse, ou faut-il pouvoir redéfinir des périmètres en fonction de chaque type de problème ? Ces questions amènent à définir les outils conceptuels que sont les « supports géographiques » des données de déplacement.

Pour formaliser les déplacements d'une population dans une base de données, deux supports de l'information géographique sont utilisés.

Nous appelons support géographique une discrétisation zonale d'un référentiel continu. Il est défini par sa résolution (surface moyenne des zones élémentaires composant le support) et sa couverture²⁴. Ainsi les caractéristiques d'un support (maillage lorsque les zones le composant sont régulières ainsi que sa résolution) induisent le choix de l'échelle d'analyse. Une échelle peut être définie comme la dimension spatiale à laquelle les entités, les modèles ou les processus sont observés et caractérisés (Marceau, 1999). D'autres auteurs définissent la notion d'échelle par deux composants : le grain et l'ampleur (« extent ») (O'Neill et King, 1998). Le premier est l'unité spatiale minimale recueillant les observations. Le deuxième est bien la couverture totale de la zone des observations. Ces deux paramètres sont identiques à ceux utilisés pour définir un support géographique. On peut donc poser la définition d'une échelle est équivalente à celle d'un support utilisé pour découper un territoire.



La construction d'un support pour mettre en données et analyser des déplacements est une réunion de deux supports. Un premier pour les origines, un deuxième pour les destinations (illustration 8). La réunion ou produit de deux supports géographiques identiques pour

24 D'autres critères sont pris en compte pour définir un maillage. Parmi eux se trouve sa dimension (2D, 3D). Si nous raisonnons en 2 dimensions, le volume au sens de l'ampleur (étendue) du maillage est égal à son aire.

formaliser les déplacements est une matrice origine-destination. Pour compter les déplacements entre ces zones, nous construisons en effet un tableau à double entrées. Chaque case du tableau permet alors de comptabiliser le nombre de déplacements effectués d'une zone à une autre, appelé « flux ». Avec l'illustration 8, nous remarquons que ce formalisme modifie l'information sur la géométrie des déplacements parce qu'il réduit la précision des lieux d'origine et de destination.

Nous mettons là le doigt sur un élément important de cette thèse : l'agrégation des déplacements dans une matrice de flux peut biaiser les distances des déplacements enquêtés. Il faut se demander quelles en sont les conséquences sur la mesure des distances et sur les autres indicateurs qui en découlent. Pour produire ces indicateurs avec le minimum de biais dû au support, il faut connaître la localisation des extrémités des déplacements le plus précisément possible. Concrètement, il faut connaître les lieux de résidence et d'activités à l'adresse, voire plus précisément dans le cas de très grandes parcelles, comme les grands ensembles ou certains sites industriels. Cette adresse peut être collectée par des enquêtes de terrain mais leur vérification et leur traitement sont longs et très coûteux et leur exploitation risque de ne pas respecter la règle de l'anonymat.

La collecte de données sur les déplacements s'effectue néanmoins avec difficulté sur un support de grande résolution. Un déplacement est une action éphémère. Les déplacements sont produits en très grand nombre. Les enquêtes ne font donc que collecter un petit échantillon de déplacements. Les déplacements diffèrent par leur motif, le mode utilisé ou par le moment de leur réalisation, mais la plus grande différenciation des déplacements vient de la localisation des origines et destinations qui peuvent varier presque continûment sur tout un territoire.

Une enquête statistique collecte des déplacements. Cet ensemble collecté peut être vu comme des échantillons de la mobilité d'un échantillon d'individus. Cela permet de comprendre les logiques individuelles de la mobilité. Il peut aussi être vu comme un échantillon de l'accessibilité générale à l'intérieur d'un territoire.

Vouloir que ces échantillons soient représentatifs peut conduire à des stratégies de sondage différentes, plus ou moins coûteuses et compatibles avec les budgets des autorités organisatrices de transport. Avoir un échantillon représentatif de la mobilité d'un individu demande de l'observer pendant une longue durée, ce qui augmente le prix de l'enquête pour un individu. Il devient donc difficile de constituer un échantillon d'individus socialement et

spatialement représentatif de la population d'un territoire. La représentativité de l'accessibilité d'un territoire pourra s'analyser par une matrice entre zones.

Mais plus une zone est grande et plus les comportements de mobilité en son sein sont supposés différents. Par exemple, il existe un gradient d'utilisation des transports en commun selon la distance d'accès. L'utilisation d'un zonage amène à postuler que l'écart de distance d'accès à un même point d'accès entre les observations n'existe pas ou tout du moins n'est pas significatif. Pour observer le rayon d'attraction d'une station de transport en commun il faudrait découper l'espace avec plusieurs zones de suffisamment petites tailles pour ne pas faire d'hypothèse de construction trop forte.

Par construction, l'utilisation d'un support fin amène à devoir effectuer un tirage au sort bien réparti dans l'espace pour disposer d'observations pour l'ensemble des zones du support. Les variations dans le temps ont aussi de l'importance.

Les Enquêtes Ménages-Déplacements (voir ci-dessous) représentent un compromis entre ces trois objectifs de représentativité (spatiale, sociale et individuelle) et les budgets disponibles qui privilégient les territoires des grandes agglomérations. Toutefois, en reprenant les besoins exprimés par les fiches du ministère, nous devons tendre à caractériser la demande de déplacements sur tout le territoire. Nous devons connaître les modes de transport et les profils socio-économiques permettant de comprendre les logiques des choix contraints. Nous devons abandonner le formalisme de la matrice de déplacement et traiter de fichiers de mobilité représentatifs des logiques spatiales et sociales de la mobilité comme proposé par les Enquêtes Ménages-Déplacements. Dans le paradigme de la mobilité durable les éléments à distinguer sont donc nombreux. Pour pouvoir stocker l'ensemble des informations nécessaires il convient de disposer de fichier de déplacements pondérés. Chaque ligne du fichier décrit une observation. En colonnes, nous disposons des variables décrivant : les déplacements avec des variables de localisations se référant à un support géographique, les caractéristiques socio-économiques des individus et les poids de sondage relatifs au taux d'échantillonnage. Le couple des variables de localisation peut se référer à des supports différents, par exemple un support pour les zones centrales, un autre pour les zones périphériques. Un couple de deux variables de localisations constitue ce qu'il est communément appelé un couple origine-destination qui est la formalisation géométrique la plus simple d'un déplacement. On lui adjoint des données attributaires, généralement définies comme suit.

Les données attributaires décrivant les couples origine-destination

-Un motif : Un déplacement est un moyen d'accès et pas une fin en soi. A chaque déplacement correspond un motif. Les données de mobilité doivent renseigner sur le chaînage de ces déplacements. Autrement dit, nous devons connaître les schémas d'activités. Certaines de ces activités sont plus ou moins récurrentes dans le temps et structurent l'ensemble des autres motifs de déplacements. Parmi les motifs les plus récurrents, le motif domicile-travail fera consensus pour la population des actifs (Quetelard, 2010). Il est donc intéressant de disposer de données de déplacement sur ce motif et les aménageurs peuvent prévoir grâce à elles la demande en heure de pointe.

-Les modes de transport : Comme nous l'avons vu dans la présentation des évaluations ex-ante et ex-post, il est indispensable de disposer d'une information fine sur les parts modales des modes de transport. Dans le cadre de la mobilité durable, il faut savoir si l'utilisateur peut utiliser un mode de transport autre que la voiture personnelle. Le choix du système de transport est l'expression des contraintes d'accès aux différents lieux d'activités. Nous devons observer les distances d'accès aux points d'entrées des réseaux du système de transport (distance aux parkings, aux stations de transports en commun, aux stations d'auto-partage, aux co-voiturage). Cette connaissance des aires de rabattement nécessite de disposer d'un support suffisamment fin pour pouvoir différencier les points de génération de trafic (lieu de résidence et d'activité) et les points d'entrée aux systèmes de transport.

-La succession des trajets décrivant les itinéraires des couples origine-destination: La connaissance de la succession des variables de localisation formalisant la géométrie des trajets est en théorie possible. Cette succession de trajets permet de coder les itinéraires des déplacements, à la condition, comme précédemment, de disposer d'un support suffisamment fin pour pouvoir bien différencier les localisations. D'un point de vue pratique les enchaînements de petits trajets (notamment à pied) composant les déplacements sont difficilement observables et donc difficiles à mettre sous forme de données. Les fiches méthodologiques précisent que les enquêtes peuvent permettre de recourir à une modélisation de trafic capable de construire ces enchaînements de trajet. Selon les règles méthodologiques en vigueur, il n'est pas nécessaire de les enquêter car trop coûteux pour l'enquêteur (Certu, 2003). Dans bien des enquêtes, nous ne connaissons pas la succession spatiale des trajets composant les déplacements, ou seulement les principaux.

-Les profils socio-économiques : Comme nous l'avons vu le fait d'agréger spatialement les observations à l'aide d'un support de l'information géographique réduit la précision de localisation des individus. L'objet d'une analyse statistique est d'expliquer les différences de comportements de mobilité et de choix de systèmes de transport entre les individus d'une même zone. En disposant de nombreuses variables décrivant les individus nous pouvons alors essayer d'expliquer des logiques de comportements en construisant et en observant divers profils socio-économiques.

-Dates, heures de départ et d'arrivées : Pour comprendre les choix de mode de transport et plus globalement les chaînes d'activités il est indispensable de connaître les dates mais aussi les horaires permettant de mesurer les durées des schémas d'activités. Ces informations permettent de comprendre les choix de mode de transport et les phénomènes de congestion (d'heures de pointe) ainsi que les logiques d'emploi du temps des individus dans leur ménage.

Précisions que les variables attributaires caractérisant les couples origines destinations sont présentées séparément mais ces variables sont en interdépendance : le motif, l'heure, les possibilités d'accès aux différents systèmes de transport, le profil socio-économique de la personne, l'itinéraire à parcourir sont des éléments à considérer ensemble pour expliquer la mobilité des individus. De plus, nous devons disposer de l'ensemble de ces éléments sur un support spatial suffisamment fin pour pouvoir étudier les distances d'accès au transport en commun, mais nous pouvons recourir pour cela à de la modélisation qui peut s'appuyer sur d'autres données. Et bien entendu, il est impératif de connaître ces informations pour l'ensemble de la population du territoire national.

1.2.2 Les données de déplacement produites par la statistique publique

La statistique publique regroupe l'ensemble des services statistiques ministériels (Journal Officiel, Arrêté du 1^{er} Février 2016²⁵). Les données de déplacement traitant de la mobilité des individus, autre que l'ENTD, sont les Enquête-Déplacements standard Certu des services en charge du Transport et des données du recensement de la population (illustration 9).

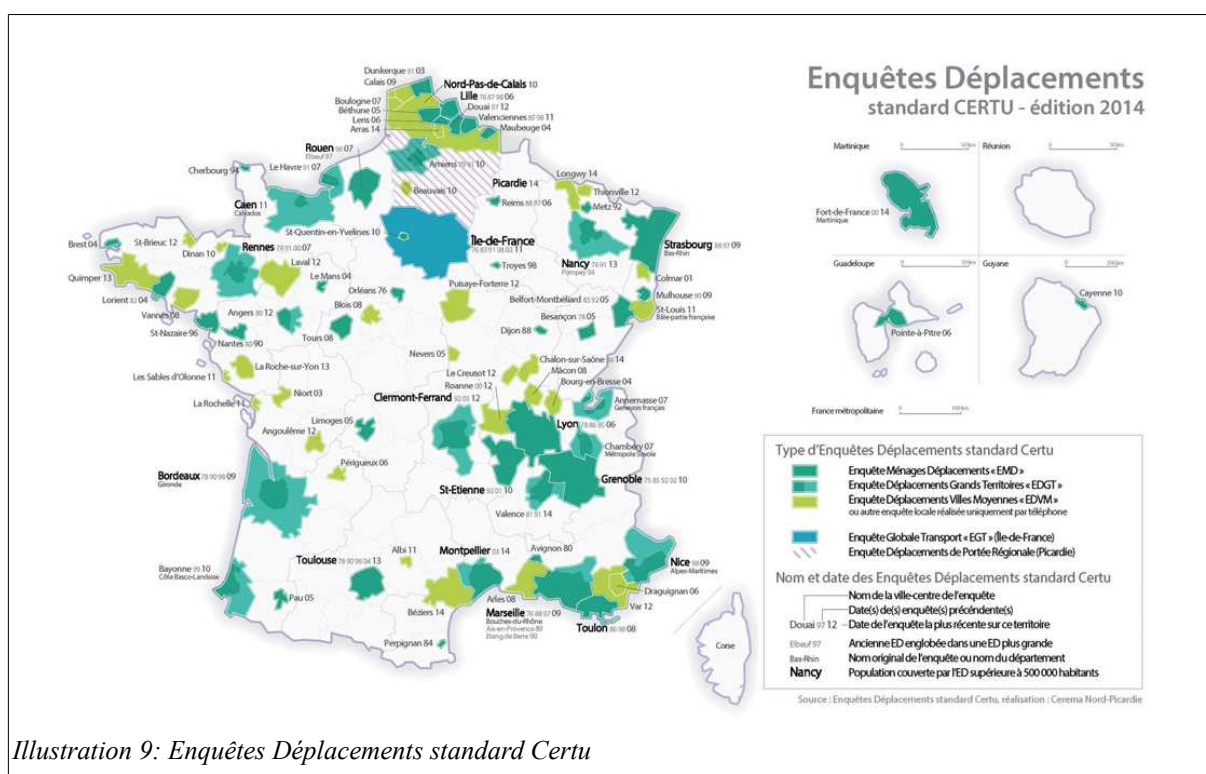


Illustration 9: Enquêtes Déplacements standard Certu

1.2.2.1 Les Enquêtes Déplacements standard Certu

Les enquêtes Déplacements standard Certu regroupent l'ensemble des méthodologies des enquêtes utilisées dans les projets d'évaluation de transports et/ou dans les plans de déplacements urbains. Le Centre d'études sur les réseaux, les transports, l'urbanisme et les

²⁵ Journal Officiel, Arrêté du 1^{er} Février 2016 [en ligne] (Consultée le 16/06/2016)
 <https://www.legifrance.gouv.fr/affichTexte.do;jsessionid=8EBD409457CDC7FA261A348D1DE36AA6.tpdila12v_2?cidTexte=JORFTEXT000032000138&dateTexte=&oldAction=rechJO&categorieLien=id&idJO=JORFCONT000031999976>

constructions publiques (Certu) était un service du ministère en charge des transports chargé de conduire des études dans le domaine des réseaux urbains, des transports, de l'urbanisme et des constructions publiques, pour le compte de l'État ou au bénéfice des collectivités locales, établissements publics ou entreprises chargés de missions de service public. Depuis le 1^{er} janvier 2014 le Certu a été intégré avec d'autres centres de recherches dans un établissement public administratif appelé CEREMA (Centre d'étude et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement).

Une des fonctions de cet établissement public est d'établir des guides méthodologiques de production d'enquêtes dédiées à l'analyse de la mobilité et du fonctionnement des systèmes ou infrastructures de transports. Lorsque une collectivité territoriale souhaite établir un plan de déplacement urbain ou évaluer un projet de transport, elle doit procéder à une enquête de terrain lui permettant de disposer d'une « photographie de la mobilité » des administrés. Ces données sont également utilisées pour calibrer les modèles de transport. En suivant le guide méthodologique dit « standard Certu » (Certu, 2008), les collectivités reçoivent une subvention de l'État couvrant environ 20% du coût de l'enquête. Les méthodologies d'enquête préconisées sont différentes selon les types d'espaces et les moyens économiques disponibles des collectivités : Enquêtes ménages déplacements EMD, pour les grandes agglomérations ; Enquêtes déplacements grands territoires (EDGT) et Enquêtes déplacements villes moyennes (EDVM). Elles ont suivi les procédures d'examen de la qualité statistique par le comité du Label du CNIS et font à ce titre partie du dispositif de la statistique publique. (CNIS, Fiches de présentation d'enquêtes statistiques pour examen d'opportunité²⁶)

« Depuis 1976, 98 EMD ont été réalisées dans cinquante-cinq agglomérations. L'enquête ménages déplacements est reconnue d'intérêt général par le Conseil national de l'information statistique (CNIS), ce qui atteste de son utilité pour la société. »(CEREMA, 2013, p.6)

-Les Enquêtes Ménages-Déplacements (EMD) :

« Les enquêtes sont réalisées au domicile des personnes enquêtées. Toutes les personnes âgées d'au moins 5 ans et habitant dans le logement, sont interrogées en

26 CNIS, Fiches de présentation d'enquêtes statistiques pour examen d'opportunité [en ligne] (Consultée le 16/06/2016)
<http://www.cnis.fr/files/content/sites/Cnis/files/Fichiers/commissions/territoires/2013/documents_preparatoires/DPR_2013_2e_reunion_COM_territoires_enquetes_opportunite.PDF>

personne. Elles recensent les caractéristiques sociodémographiques des ménages (localisation, logement, motorisation, etc.) et des personnes (âge, sexe, profession, etc.) ainsi que la description précise de tous leurs déplacements réalisés la veille du jour d'enquête. Les caractéristiques de ces déplacements (motifs, modes, zones géographiques, heures de départ et d'arrivée) sont recueillies sur la base de définitions précises.

Les enquêtes Ménages-Déplacements portent sur un échantillon représentatif des ménages de l'aire d'étude. La taille de l'échantillon est fixée de manière à assurer une fiabilité statistique minimum des résultats par secteurs. Il est recommandé d'enquêter au moins 1500 ménages sur l'ensemble de l'aire d'étude, 160 personnes dans chacun des secteurs de résidence choisis comme unité de tirage de l'échantillon sans jamais descendre au-dessous de 70 ménages » (CEREMA ,EMD, EDVM et EDGT : méthodes et guides²⁷).

-Les enquêtes Déplacements Villes Moyennes (EDVM) :

« Pour répondre aux besoins de connaissance des comportements de mobilité sur des territoires de plus en plus diversifiés, le Certu a élaboré, au début des années 2000, une méthode de recueil des données de mobilité adaptée aux villes moyennes. La méthode « enquête déplacements villes moyennes » (EDVM) concerne les agglomérations dont le pôle urbain ne dépasse pas 100 000 habitants, l'aire de l'enquête pouvant elle-même les dépasser afin d'intégrer un territoire plus large. Les enquêtes se déroulent par téléphone et les déplacements ne sont recensés que pour une ou deux personnes âgées de plus de 11 ans par ménage tiré au sort. Les questionnaires sont allégés par rapport à ceux de l'EMD. Le respect du « standard Certu » garantit la fiabilité et la comparabilité des résultats. Moins coûteuse qu'une EMD, une EDVM apporte logiquement des informations moins nombreuses » (CEREMA, EMD, EDVM et EDGT : méthodes et guides).

- Les enquêtes Déplacements Grands Territoires (EDGT) :

« Le Certu a adapté la méthode des EMD pour des périmètres élargis associant territoires denses et territoires périurbains, voire ruraux. On parle alors d'enquêtes déplacements grands territoires (EDGT) : à une EMD classique sur les territoires les plus denses, on associe une enquête téléphonique sur les territoires périurbains moins denses et éloignés des centres des agglomérations »(CEREMA ,EMD, EDVM et EDGT : méthodes et guides).

27 CEREMA, EMD, EDVM et EDGT : méthodes et guides [en ligne] (Consultée le 26/03/2016
<<http://www.territoires-ville.cerema.fr/emd-edvm-et-edgt-methodes-et-guides-a679.html>>

-L'enquête Globale Transport (EGT):

Parmi les Enquêtes Déplacements, nous distinguons l'Enquête Globale d'Île-de-France. Elle est basée sur les mêmes questionnaires et principes que l'EMD et donne une vue d'ensemble de la mobilité francilienne. L'enquête renseigne également sur l'équipement des ménages (voiture, scooter, vélo...) ainsi que sur la possession du permis de conduire et/ou d'une carte d'abonnement (transports en commun, vélib', autopartage...). Cette enquête a déjà été réalisée plusieurs fois par le passé : en 1975, 1983, 1991, 2001 et 2011, soit tous les dix ans, environ. Elle se distingue également par un support géographique différent des EMD, comme nous le verrons plus loin.

Forces et faiblesses du dispositif des EMD

Les données des EMD appliquent donc le modèle de données mis en avant par les fiches outils. Il n'y a pas à s'en étonner puisqu'elles occupent une position centrale dans le dispositif de suivi et d'analyse de la mobilité en France. C'est ce qui en fait leur force et a permis leur extension à des territoires plus variés et moins denses. L'application de ce modèle dépend toutefois d'un support de collecte, spécifié par la méthode Certu, qui semble aussi constituer un point faible du dispositif, et de son extension à ces territoires moins denses. Un rapport du CNIS (Cuillier, 2009) fait un constat critique et rappelle en effet des points à améliorer de ces enquêtes Ménages-Déplacements, dont certains sont liés au support géographique des EMD et plus fondamentalement au taux d'échantillonnage.

- « Les collectivités territoriales voient apparaître une exigence supplémentaire suite au Grenelle de l'Environnement. Or, les enquêtes EMD se caractérisent par leur coût élevé, leur manque de représentativité sur des territoires fins, la lourdeur et la lenteur de leur exploitation, la difficulté des comparaisons temporelles et l'absence de mise à disposition au-delà des commanditaires. De plus en plus, les décideurs souhaitent disposer d'informations fréquentes et rapides, mais avec un détail géographique plus fin » (Cuillier, 2009, p.5).

Pour répondre à ces insuffisances le CNIS propose une rénovation: *-« La réflexion autour de cette rénovation doit prendre en compte plusieurs préoccupations. Les enquêtes rénovées doivent en premier lieu s'intégrer à un dispositif global de connaissances des mobilités des ménages, comprenant notamment les campagnes annuelles de recensement et les enquêtes Transports ».* (Cuillier, 2009, p.5)

1.2.2.2 Les données de mobilité du recensement de la population

Les données du recensement sont la référence pour connaître la dynamique et la structure sociale et démographique des territoires. Elles ont un rôle technique pour le redressement de la non réponse dans la production des données statistiques. En effet, toutes les personnes contactées pour une enquête, comme par exemple les EMD, ne peuvent ou ne veulent pas répondre. Pour éviter que la structure de l'échantillon des répondants biaise les résultats (parce qu'ils sont par exemple plus âgés que la population réelle), cette structure est « redressée » par une pondération qui augmente la représentativité des classes où au moins des personnes qui ont répondu, de façon à retrouver les proportions des classes indiquées par le recensement. Dans les modèles, les enquêtes fournissent une structure relative (en%) de la demande et la valeur absolue de cette demande est donnée par les stocks d'individus du recensement, décomposés selon cette structure. Enfin, le recensement contient également une information sur le lieu de travail ou d'études et sur la façon d'aller au lieu de travail qui constitue un élément de connaissance important sur la mobilité. Il faut ajouter cependant que ces déplacements domicile-travail ne sont pas immédiatement traduisibles en mobilité (Hubert, Pistre, Madre, 2016).

La politique de diffusion de l'Insee du recensement applique les principes de la directive européenne INSPIRE sur la libre diffusion des informations de référence. L'Insee diffuse plusieurs fichiers dits « de détail anonymisés », avec une ligne par logement ou individu, mais sans permettre l'identification des répondants.

Tableau 1 : Les fichiers détail anonymisés du recensement rénové		
Nom des fichiers	Principaux avantages	Principaux inconvénients
Fichier Logement ordinaire	Des données sur l'ensemble des bulletins collectés et localisations des lieux de résidence aux IRIS-Communes	Décrit seulement les chefs de ménages
Fichier Individus localisé au canton-ville (INDCANTVILLE)	Décrit l'ensemble des individus des ménages	Localisation aux cantons de résidences pour les habitants des petites communes
Fichier Individus localisés à la région	Des modalités plus détaillées que INDCANTVILLE	Localisation à la région ou département et EPCI pour les départements et EPCI de plus de 700 000 habitants
Fichier Mobilité scolaire des individus	Des données sur les communes de scolarisation	Pas de variable socio-économique, pas de mode de transport
Fichier des Mobilités professionnelles (MOBPRO)	Des données sur les communes de travail	Localisations à la commune, moins de variables socio-économiques
Fichier des activités professionnelle des individus	Des modalités plus détaillées que MOBPRO	Zone d'emploi de travail
Fichier des migrations résidentielle (MIGCOM)	Plus disponible depuis 2007 ²⁸	Plus disponible
Source : Insee		

Le fichier détail appelé « fichier des mobilités professionnelles » (MOBPRO), issu de l'exploitation complémentaire, donne les informations suivantes. « *Chaque enregistrement du fichier correspond à un individu décrit selon les caractéristiques de ses déplacements pour aller travailler (déplacements domicile-travail), ses principales caractéristiques sociodémographiques, ainsi que celles du ménage auquel il appartient. Sont pris en compte, tous les individus actifs ayant un emploi âgés de 15 ans ou plus, recensés en France métropolitaine et dans les départements d'outre-mer (DOM). Le niveau géographique le plus fin pour localiser le lieu de résidence de l'individu et son lieu de travail est la commune (arrondissement municipal pour Paris, Lyon, Marseille).* » (Insee, Documentation du fichier

28 Ce fichier n'est plus disponible car la question sur le changement de résidence a été modifiée en 2009 pour être harmonisée au niveau européen. Pour l'instant il n'est pas communiqué la date de ré-apparition du fichier.

détail des mobilités professionnelles²⁹)

Toutefois, le lien entre les individus d'un même ménage n'est pas diffusé dans le fichier MOBPRO car il permet potentiellement d'identifier de nombreux ménages. Notre processus de désagrégation de désagrégation appliqué à l'Île-de-France a pourtant comme point de départ le fichier MOBPRO. Notons qu'il existe un fichier confidentiel complet du recensement accessible aux chercheurs par le Centre d'Accès Sécurisé aux Données (CASD) (Le Gléau et Royer, 2011). Nous reviendrons plusieurs fois sur le potentiel d'utilisation de ce fichier dans nos recherches.

De manière générale les données du recensement ne sont pas comparables à celles d'une enquête sur la mobilité des individus selon les standards Certu. Les EMD décrivent tous les déplacements et trajets réalisés un jour de semaine. Le fichier MOBPRO ne fait que renseigner sur l'adresse du lieu de travail et sur le mode de transport habituel, s'il y a déplacement. Seule la commune du lieu de travail est diffusée et nous ne savons pas si les recensés se rendent réellement sur le lieu déclaré dans le bulletin du recensement ni combien de fois par semaine. Toutefois ces données du recensement ont une couverture spatiale, un taux de sondage, et une fréquence d'actualisation meilleurs que ceux des EMD et, dans certains cas même, ont une résolution spatiale plus fine.

Comme invite le CNIS, il est sûrement judicieux de réfléchir à des méthodes permettant d'allier les avantages de ces deux types d'enquêtes mais au préalable il faut leur trouver un support géographique commun.

29 Insee, Documentation du fichier détail des mobilités professionnelles [en ligne] (Consultée le 16/06/2016) <http://www.insee.fr/fr/ppp/bases-de-donnees/fichiers_detail/RP2008/doc/contenu_RP2008_mobpro.pdf>

1.3 Élargir l'éventail des données pour répondre aux nouvelles questions

Il existe une source de données décrivant des déplacements avec une précision fine et en grand nombre : les données produites par les gestionnaires des réseaux de transport et de téléphonie que nous englobons sous l'appellation « Big Data ». Mais l'utilisation se heurte pour l'instant à l'impossibilité de les associer au référentiel socio-démographique. Ces données apportent néanmoins une représentation de l'accessibilité territoriale mais sans pouvoir la raccorder à des logiques individuelles.

1.3.1 Les données de mobilité à l'ère du big data.

En dehors des enquêtes statistiques, des données sur la demande de déplacements, avec une localisation fine des traces de déplacements, existent en grand nombre. Les nouvelles sources sont les comptages routiers automatiques, la billétique des systèmes de transport en commun, les traces de téléphonies mobiles, les données issues des applications mobiles, des sources issues des sites et réseaux sociaux internet géo-localisant les messages et photos. La géolocalisation des téléphones mobiles peut être réalisée par leur positionnement par rapport aux antennes des opérateurs, aux bornes Wi-Fi ou à la puce GPS intégrée aux téléphones récents.

En reprenant les attributs et les supports géographiques de données de mobilité que nous avons identifiés pour les besoins du développement durable nous pouvons analyser la pertinence de ces nouvelles sources de données de mobilité en accordant une attention particulière à la façon dont l'information géographique est produite.

Les données collectées par les opérateurs de téléphonie mobile :

Les données de déplacement issus des téléphones mobiles sont principalement issues de deux procédures de collecte, les traces GSM, les traces GPS et les bornes Wifi.

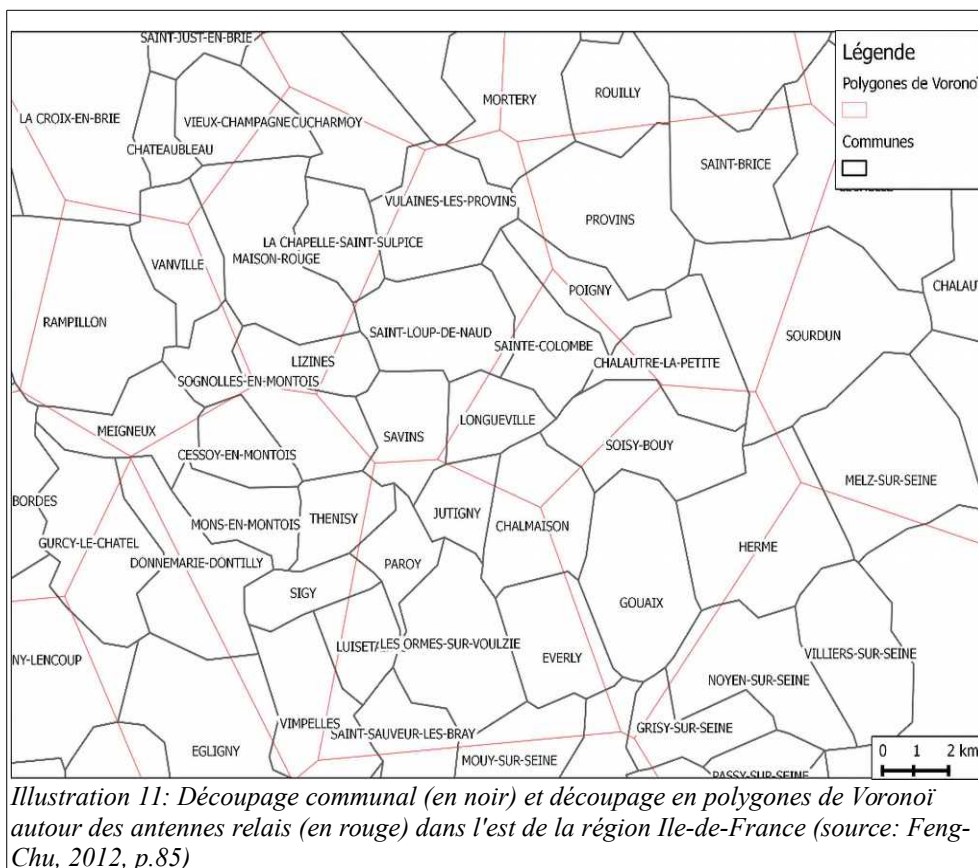
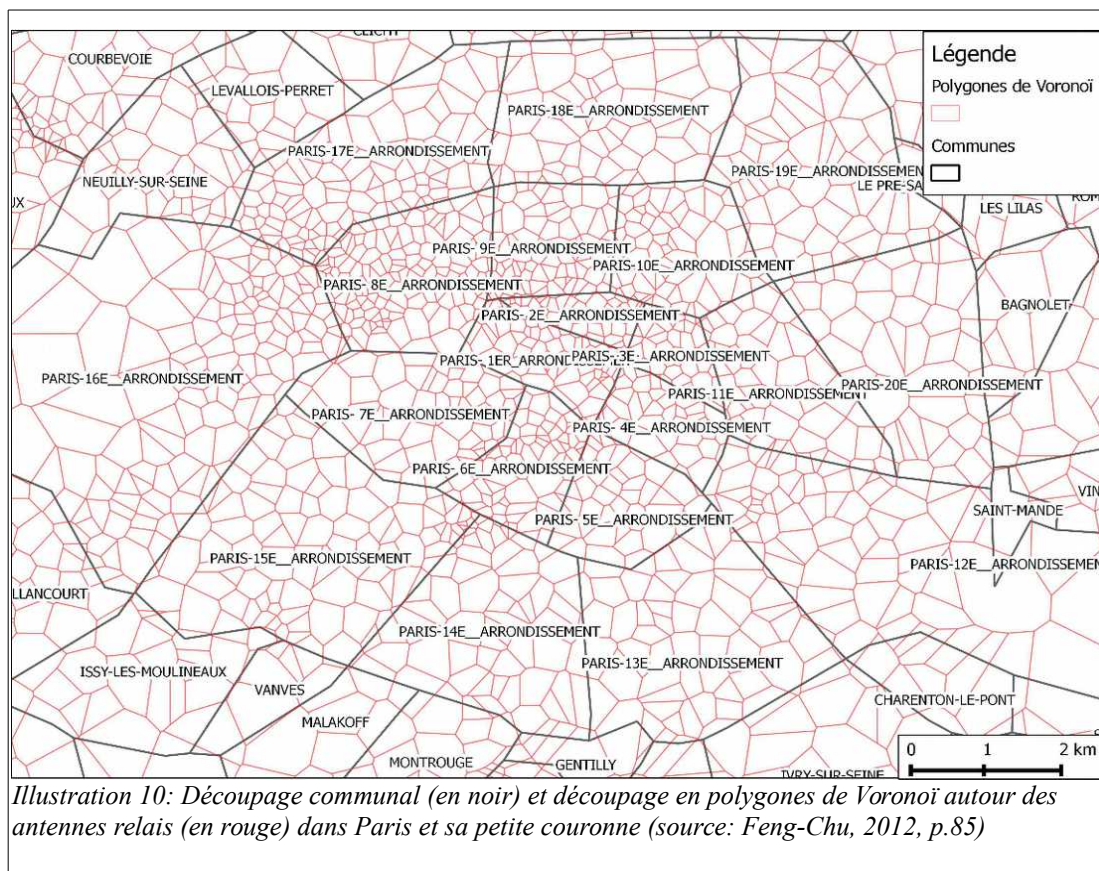
Les traces GSM sont produites lorsqu'un téléphone « capte du réseau » en recevant ou émettant un signal vers une antenne relais. Pour acheminer un appel téléphonique ou un sms les opérateurs téléphoniques doivent savoir près de quelles antennes relais sont les récepteurs pour leur envoyer les signaux. Les téléphones émettent également des traces dites

« fantômes » à l'allumage et au démarrage, et plusieurs fois par jour afin de gérer au mieux le réseau. Il faut éviter que les antennes-relais soient surchargées.

Plusieurs méthodes de géolocalisation peuvent alors être effectuées (Géolocalisation : votre portable vous suit à la trace³⁰). La plus simple est appelée : Cell ID (Identification de cellule / Numéro pour l'identification univoque de la station GSM auquel le téléphone est connecté.). Cette méthode ne nécessite pas de matériel supplémentaire, elle fonctionne si le téléphone mobile est dans une zone de couverture. Le téléphone est localisé à l'antenne relais. La finesse de la localisation dépend du parc d'antennes relais mis à disposition par l'opérateur (plus l'antenne est isolée, plus la zone de couverture est vaste et moins la localisation est précise). Selon la littérature technique la précision varie entre 100 et 700 mètres en zone urbaine. En zone rurale, cela peut aller jusqu'à 35 kilomètres.

Le site internet <https://www.antennesmobiles.fr/> donne la carte des antennes relais. Feng-Chu (Feng-Chu, 2012) propose une estimation de la portée des antennes en faisant des polygones de voronoï autour des antennes-relais (illustrations 10 et 11). Nous voyons ainsi les marges d'erreur des localisations collectées par la méthode de Cell-ID.

30 Géolocalisation : votre portable vous suit à la trace [en ligne] (Consultée le 15/05/2016)
<<http://reseau.echelon.free.fr/reseau.echelon/geolocalisation.htm>>



Des méthodes de géolocalisation plus fines existent pour les téléphones mobiles équipés du matériel adéquat (3G). Une autre méthode s'appuie sur le différentiel de temps (Enhanced Observed Time Difference (EOTD)). Le processeur du téléphone peut envoyer un signal à plusieurs stations environnantes que seule la plus proche renvoie. Le temps écoulé entre l'émission et la réception permet de calculer la localisation. Plusieurs méthodes basées sur le différentiel de temps existent. Le document de la thèse de Bouna Fall (2013³¹) citant page 16 Munoz et al.(2009³²) recense les divers niveaux de précision selon les différentes techniques.

Cell-ID : 100 m – 35 km

Angle Of Arrival (AOA) : 50 m – 150 m

Uplink Time Of Arrival (U-TDOA) : 50 m – 150 m

Enhanced Observed Time Difference (E-OTD) : 60 m – 200 m

Time Of Arrival (TOA) : 3 m – 10 m

L'avantage de ces méthodes est qu'elles utilisent des fréquences radios basses pouvant franchir les murs. Le fonctionnement continue donc d'être assuré à l'intérieur des bâtiments, des entrepôts, des parkings, des tunnels, s'ils ne sont pas trop profonds. En Île-de-France, certains quais de métros et de gares souterraines sont équipées de systèmes de retransmission de radiotéléphonie cellulaire. Chloé Millon (Millon, 2015) reconstitue à partir de ce dispositif les entrées et sorties des franciliens utilisant la ligne du RER A. Ces données peuvent fournir des informations de qualité de service, tels que des temps de trajet moyen de station à station, des fréquences, des charges en passager par véhicules. Elles peuvent aussi, dans certains cas, permettre l'étude des itinéraires (et des choix d'itinéraires) des voyageurs.

Citons également les travaux de thèse Fen-Chong (2012) reconstituant des flux de mobilité des Franciliens à partir de traces GSM. L'avantage de ces données de localisation est de fournir les horaires précisés.

L'accès à ces données est toutefois très encadré et restreint. Les données sont produites par les opérateurs de communication qui les commercialisent. Aussi ces données posent plusieurs difficultés de traitement dans les analyses de mobilité. Fen-Chong rappelle les imperfections

31 Fall B. (2012) « Evaluation des performances d'un système de localisation de véhicules de transports guidés fondé sur l'association d'une technique radio ULB et d'une technique de retournement temporel » Thèse de doctorat: Electronique, Valenciennes: Université de Valenciennes et du Hainaut-Cambresis,p.174 [en ligne] <https://tel.archives-ouvertes.fr/tel-00973788/document>

32 D. Munoz, F. Bouchereau, C. Vargas, R. Enriquez-Caldera (2009), "Position Location Techniques and Applications, " Academic Press, Burlington, p.275

des données GSM, notamment pour la méthode « Cell ID ».

- « *Le téléphone portable peut être enregistré comme localisé auprès d'une antenne relais A qui se trouve en réalité beaucoup plus loin que l'antenne relais B qui se situe plus près physiquement du téléphone.* » (Fen-Chong, 2012, p.84)

- « *Il peut s'écouler un intervalle de temps assez long entre deux enregistrements, puisque le seul type d'événements qui intervient à intervalles réguliers, est la mise à jour du réseau (qui recherche la localisation du téléphone portable) toutes les six heures environ. Il n'est donc pas possible de suivre de manière exhaustive toutes les localisations des individus et entre deux localisations successives. Si l'intervalle de temps qui s'est écoulé est long, on n'a aucune idée de la trajectoire réelle de l'individu.* N. Andrienko, G. Andrienko, H. Stange et al. (2012) qualifient ce type de données d'épisodiques, dans la mesure où elles ne fournissent pas d'information sur un mouvement continu mais sur des épisodes de ce mouvement». (Fen-Chong, 2012, p.88)

- « *Lorsqu'un téléphone portable se trouve en bordure du périmètre de zone de couverture de deux antennes-relais, il peut changer de station et donc de localisation. Ces changements de station de base créent des oscillations entre stations, expliquant que certaines observations présentent un nombre important de déplacements entre deux stations alors que dans la réalité le téléphone est immobile* ». (Fen-Chong, 2012, p.86)

De plus, ces données ne renseignent pas sur le ou les détenteurs des téléphones.

-« *En réalité, l'entité dont la localisation est enregistrée est un téléphone portable qui possède un identifiant. On ne dispose donc d'aucun renseignement sur son détenteur. C'est uniquement en recoupant les données de localisation avec les données de facturation qu'il est possible de retrouver l'identité du détenteur du téléphone portable.* » (Fen-Chong, 2012, p.87). Or une directive européenne impose de préserver l'anonymat des usagers dans les fichiers diffusés par les opérateurs. (Directive Européenne 2002/58/CE, qui encadre le secteur des Télécommunications, et notamment les données personnelles collectées par l'opérateur (Journal officiel de l'Union européenne, Protection des données dans le secteur des communications électroniques³³).

33 Journal officiel de l'Union européenne, Protection des données dans le secteur des communications électroniques [en ligne] (Consultée le 24/03/2016)
<http://europa.eu/legislation_summaries/information_society/legislative_framework/124120_fr.htm>

Ce dernier point est très regrettable pour les analyses de la mobilité durable. Les données dont nous avons besoin doivent intégrer des profils socio-économiques pour qu'il soit possible d'effectuer, par exemple, des mesures d'élasticités des comportements de reports modaux. Mais en plus, nous ne disposons généralement pas du mode de transport.

Les données GSM doivent être enrichies de données auxiliaires permettant de mieux définir les caractéristiques de déplacements. Y. Asakura et E. Hato ont proposé en 2003 (2003) un algorithme permettant de reconstituer les situations de déplacement et d'immobilité pour produire une information proche de celle collectée par des méthodes d'enquête plus classiques. Toutefois, ils ne construisent pas les profils socio-économiques. Ils se servent d'un échantillon de population décrivant leur déplacement dans un carnet (diary) et les compare avec ceux collecté par des GPS. Les auteurs reconstituent les déplacements et comparent la reconstitution des traces GPS avec les déclarations des sondés. L'idée d'utiliser des connaissances externes peut être développée. Le principe est de caractériser l'espace environnant des traces et d'inférer des caractéristiques de cet espace aux traces (caractéristiques socio-économiques, activités présentes pouvant être un motif, modes de transports utilisables) à partir de données dites alors de contrôle. Toutefois cette méthodologie pose le problème d'adéquation entre les différents supports. Sur les cartes des antennes-relais, par exemple, le découpage des polygones de voronoï et le support communal sur lequel nous disposons des données de référence sur les territoires ne correspondent pas. Ils sont dit non-alignés. Il convient de combiner des données ayant des supports incompatibles (Gotway et Young, 2002).

Ce problème de non-alignement est encore plus problématique lorsque on souhaite déterminer les modes de transports. Cela consiste à affecter les localisations sur le linéaire des réseaux de transport. El Mashri (2013) propose plusieurs solutions de clustering pour préciser les positions sur le réseau de transport. Une première étape consiste à regrouper les traces puis une seconde analyse utilise ces premiers regroupements pour les appliquer sur le réseau. El Mashri justifie cette démarche de clustering afin de faciliter la lecture des données de déplacements. De même la thèse de Fen-Chong ré-agrège les données de mobilité pour leur donner un sens, elle parle de données révélant une organisation spatio-temporelle.

Les données GPS de téléphonie mobile :

Les téléphones portables peuvent être géo-localisés à l'aide de la technologie GPS, lorsqu'ils sont équipés d'une puce GPS. Les méthodes de géo-localisation utilisent la trilatération ou la multilatération. En principe la localisation par GPS est plus précise mais elle consomme beaucoup d'énergie.

Contrairement aux méthodes par GSM, l'utilisateur doit autoriser volontairement le GPS pour que l'application mobile collectant les traces puisse fonctionner. Inconvénient ou avantage, les applications mobiles doivent avoir en paramètre un pas de temps de collecte des localisations. Pour suivre les déplacements finement il faut un pas de temps assez court. La précision des localisations à un pas de temps court rend possible l'identification du mode par l'analyse de la vitesse et des accélérations, avec quelques incertitudes cependant (Tao Pham, 2016). A chaque tentative de collecte de position, l'application met un certain temps à accrocher le signal GPS. Selon la finesse du pas de temps et le temps d'accroche du signal, les batteries de téléphones se déchargent plus ou moins rapidement. Dans bien des cas il n'est pas possible de laisser l'application mobile enregistrer tous les programmes d'activités réalisés en 24h (Nguyen-Luong et Allilo, 2014). Aussi, en admettant disposer de traces pour une population enquêtée en grand nombre, le nombre total de données à traiter peu devenir très grand. Mais supposons que dans l'ère du big data les méthodes de traitement soient au rendez-vous.

Toutefois, contrairement aux données GSM, le processus de géo-localisation devient très vite inopérant à l'intérieur des bâtiments, en zones dites couvertes. En revanche, il est possible de détenir des informations sur les profils socio-économiques des détenteurs des téléphones. Le processus de collecte se faisant par l'intermédiaire d'une application mobile les usagers doivent la télécharger au préalable. Ils sont donc conscients de leur enregistrement de parcours et peuvent si l'application le demande remplir un formulaire décrivant les usagers. Le programme de recherche intitulé *Mobi-lise*, projet financé par l'ADEME dans le cadre du programme "Mobilités : connaître pour mieux agir". *Mobi-lise* vise l'expérimentation de collecte de données par smartphone afin d'améliorer la connaissance de la mobilité par les contributions communautaires et ainsi compléter les enquêtes réalisées par les collectivités (comme les Enquêtes Ménages Déplacements). Bien entendu pour disposer d'une

représentativité socio-économique les répondants doivent être représentatifs de toutes les classes d'âge et/ou de leur catégorie socioprofessionnelle.

Les données d'accroches des bornes Wi-Fi :

L'entreprise Google collecte de nombreuses informations sur nos localisations. A chaque fois que l'on active le GPS d'un téléphone et ou que l'on se connecte à une borne Wi-Fi le service de Google enregistre les données de localisations. Ces données sont accessibles sur le service de Google Maps dans l'onglet « Vos trajets » (illustration 12).

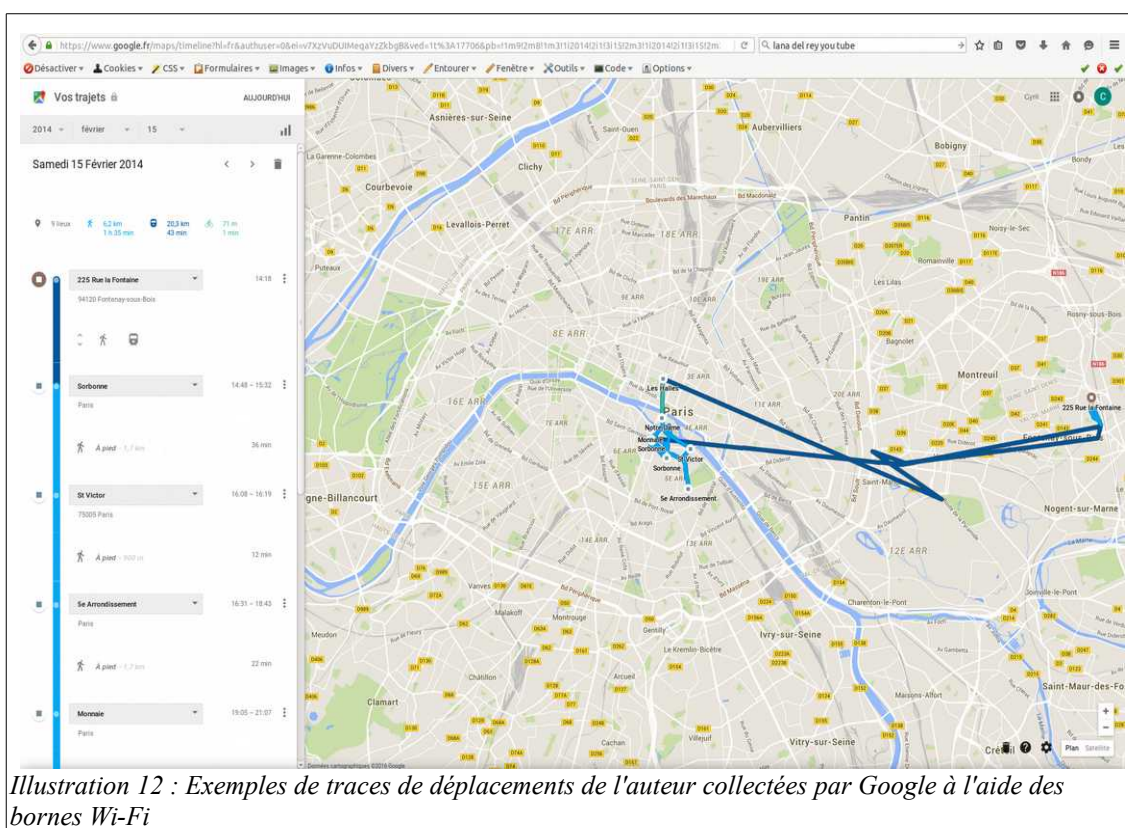


Illustration 12 : Exemples de traces de déplacements de l'auteur collectées par Google à l'aide des bornes Wi-Fi

« L'historique des positions permet au système de fournir des informations utiles en fonction des lieux où les usagers se sont rendus avec les appareils sur lesquels ils se sont connectés avec leur compte Google. Il est possible d'obtenir des prévisions relatives aux trajets domicile-travail fréquents ainsi que de meilleurs résultats de recherche. Les informations de localisation peuvent en outre être utilisées par les applications ou les services Google, notamment dans le cadre des annonces affichées » (Google. Support Google³⁴).

34 Google. Support Google [en ligne] (consultée le 28/03/2016)
<<https://support.google.com/accounts/answer/3118687?hl=fr>>

Une des utilisations faites par l'entreprise de ces données de localisations est de déterminer les zones congestionnées des réseaux et propose des itinéraires adaptés en fonction de l'état du réseau. *"L'historique des positions permet de créer une carte privée des lieux où vous vous rendez avec les appareils auxquels vous êtes connecté, afin notamment d'améliorer la recherche de cartes et le calcul d'itinéraires pour vos trajets domicile-travail"*, lit-on sur la page dédiée.

Rappelons que les traces de localisations de Google sont aussi dépendantes des temps entre les différents points de collecte d'information. C'est sûrement ce qui incite Google à demander d'enregistrer manuellement les adresses du domicile et de travail pour améliorer son service de prédiction de trafic.

En effet, il ne faut pas confondre les établissements auxquelles les téléphones se sont accrochés (bornes Wi-Fi) et les motifs de déplacements. Il se peut qu'un téléphone capte la borne Wi-Fi de l'établissement voisin.

Enfin, nous observons sur la carte que le service de Google donne une estimation du mode de transport. Toutefois la construction du mode de transport n'est pas encore infaillible. Sur la carte, il est dit qu'une partie des déplacements de la journée sont effectués en vélo. Dans la réalité, je peux vous dire que je n'utilise jamais le vélo. Nous retrouvons là les algorithmes permettant d'affecter les localisations sur un autre support pour enrichir les données attributaires décrivant les traces de déplacement collectées.

Données billettiques :

« Un système billettique traite, de façon automatisée, des données à caractère personnel sur les détenteurs de cartes. » (CEREMA, portail billettique³⁵). Il peut s'agir d'abonnement aux réseaux de transports en commun, en stations de vélos partagés, de véhicule de partage. Les données de billettique sont les traces de passage d'un usager en règle, dans un appareil de validation, fixe ou embarqué. Cet usager peut avoir un titre de transport anonyme ou personnalisé. Elles peuvent donc potentiellement tracer cet usager pour un trajet s'il est anonyme, tous ses trajets s'il ne l'est pas.

Un des projets du PREDIT-4 nommé Mobilletic (El Mahsri et al., 2015) analyse le potentiel

35 Cerema, portail billettique [en ligne] consulté le 24/03/2016 <<http://www.billettique.fr/spip.php?article205>>

des données billétiques des réseaux de transport en commun pour observer les déplacements multi-modaux. Ce projet s'est focalisé sur l'agglomération rennaise où un abonnement unique permet de se déplacer en bus, en tramway, en métro et en vélos libre service. De fait, le parcours des abonnés peut être retracé en collectant les traces de validation des abonnements, à condition que les usagers valident. Les traces de passage sont finement localisées et les heures de validation sont également disponibles et précises.

Les inconvénients de cette source de données sont qu'elle est limitée aux utilisateurs des réseaux publics de transport. Dans bien des cas, nous ne disposons que de la station d'origine des déplacements : *«Pour des systèmes comme le métro ou le bus, seules les données de l'origine du déplacement sont disponibles (l'usager valide son ticket uniquement à la montée)»* (Ifsttar, mobilletic³⁶). La destination peut être retrouvée en postulant que la prochaine validation indique la destination de l'ancien déplacement mais cela peut être sujet à des erreurs. La partie du déplacement à l'extérieur du réseau de transport est également inconnue. Enfin, les données billétiques ne contiennent pas de motifs de déplacements.

Mais le principal obstacle à une exploitation au niveau individuel est d'ordre légal. La législation encadrant la mise en place de système billétique est claire au sujet du respect de l'anonymat.

-« La possibilité de circuler de façon anonyme doit, dans tous les cas, être préservée : chaque responsable de traitement qui met à disposition des usagers des titres nominatifs de transport doit également prévoir de leur laisser le choix d'utiliser des titres de transport anonymes. [...] Outre l'utilisation des titres de transport, le dispositif billétique vise à assurer la réalisation d'analyses statistiques, la mesure de la qualité du fonctionnement du système, la détection de la fraude technologique et le suivi des impayés » » (Legifrance, 2016.³⁷)

Du fait de cette législation, les systèmes de billétique, en France, doivent être conçus pour rendre techniquement impossible le couplage des traces avec les données personnelles de l'abonné. Les données de localisation par la billétique doivent être anonymisées et doivent être analysées séparément de celles sur les abonnés (tarifs spéciaux, lieu de résidence, etc.). Les

36 Ifsttar. mobilletic [en ligne] (Consultée le 29/03/2016) < <http://mobilletic.ifsttar.fr/> >

37 Légifrance [en ligne] (Consultée le 12/02/2016)

<https://www.legifrance.gouv.fr/affichTexte.do;jsessionid=5CC613290E3003C2E829BB6E02DEE09A.tpdjo10v_3?cidTexte=JORFTEXT000019109412&dateTexte=>

praticiens et les chercheurs doivent déterminer les lieux d'origines et de destinations après coup. Et ils se servent également des données de référence institutionnelles pour tenter de retrouver les points de générations de trafic (Tamblay et al., 2015) ou construire les motifs de déplacements (Devillaine et al., 2012).

1.3.2 Le couplage de données

Le panorama sur les données des grandes institutions et de celles issues du big data fait ressortir que chacune des sources de données à ses avantages et inconvénients.

Les données du big data permettent de disposer d'une masse de données importantes et donc de représenter l'accessibilité d'un territoire. Les localisations sont précises quand le matériel est bien opérationnel. Mais nous ne disposons pas de l'ensemble des éléments de connaissances dont le changement de paradigme des politiques de transport a besoin. Les données de téléphonie mobile (GSM, GPS, Wi-Fi) ont l'avantage d'informer assez précisément sur les localisations. Une information sur les itinéraires peut être effectuée à partir de traitement sur les données que constituent ces traces de téléphonie mobile. Toutefois, on ne dispose pas d'information sur les motifs, les profils socio-économiques et les modes de transport. Les données de billettiques informent, quant à elles, sur les modes de transports, les heures et certains points de passage des déplacements. En revanche, nous ne connaissons pas les points de départs et d'arrivées des déplacements, ni les motifs, ni les profils socio-économiques. Plus précisément, il est possible de retracer les individus sur une période de temps assez longues pour identifier les individus, mais cela pose des problèmes de confidentialité. Et tout en ne respectant pas les contraintes d'anonymat, la représentativité statistique n'est pas assurée³⁸.

La littérature présentée lors du panorama des données de mobilité (Y. Asakura et E. Hato, Google, El Mashri, Tamblay et al, Devillaine et al.) montre des exemples pour reconstituer des informations et notamment des éléments attributaires des déplacements à partir de connaissances externes sur les réseaux de voiries ou les lignes de transport public. Pour tendre vers une description plus fine de la mobilité dans un territoire nous proposons de suivre cette voie et de coupler différentes sources de données. Des données concernant les déplacements, mais aussi des données sur les déterminants locaux de la mobilité, sur les

38 La pratique des smartphones n'est pas complètement diffusée par exemple.

caractéristiques de la population, de l'offre et des infrastructures de transport et de l'emploi peuvent être mobilisées. Il nous paraît important et dans la logique du Grenelle de prendre comme base un référentiel socio-économique, car on a besoin de comprendre les logiques des comportements. L'autre approche fondée sur une base de données de déplacements enrichie par des données socio-économiques paraît moins opérantes, en raison notamment du biais statistique introduit par la possession et l'utilisation d'un téléphone mobile. Les deux sources de données renseignant sur les profils socio-économiques que nous envisageons de mobiliser sont les EMD et le recensement.

Une première option consiste à extrapoler les Enquêtes Ménages-Déplacements. Celles-ci avaient pour vocation de produire un ensemble complet de données. Mais le taux d'échantillonnage et les fréquences d'actualisations sont faibles. La précision des descriptions est trop peu précise et limitée aux jours de semaine hors vacances. Les besoins de connaissance soulevés par le Grenelle de l'Environnement doivent s'appuyer sur un dispositif plus précis constitué à partir de bases de données assises sur des échantillons de grandes tailles. L'extrapolation d'EMD consiste à produire par la modélisation une base de données de déplacements pour l'ensemble de la population. La chaîne de modélisation à quatre étapes (génération, distribution, choix modal, affectation) réalise cette opération. Les deux premières étapes consistent à produire une matrice de déplacement en extrapolant les données EMD à l'aide du recensement et/ou d'autres sources de données. Toutefois, la chaîne de modélisation est critiquable. Le modèle à quatre étapes ne tient pas compte des enchaînements d'activités, et la congruence des activités des ménages n'est pas considérée. Il vise surtout à obtenir une représentation des flux sur le réseau en période de pointe (Bonnell, 200) (Ortuzar, 2011). Or les objectifs du Grenelle sont plus ambitieux.

Une autre modélisation empruntée par le modèle Mobisim consiste à utiliser les EMD pour définir des schémas d'activité, déterminer leur fréquence et les affecter à une population synthétique de ménages, construite à partir du recensement, pour simuler leur mobilité (Antony et Vuidel, 2010).

Néanmoins, le manque de précision des EMD, dû au faible taux d'échantillonnage, affecte aussi la représentativité des schémas d'activité selon les zones. Il est difficile de coupler directement les données des EMD à une autre source de données car les supports de l'information géographique ne sont pas identiques (problème de non-alignement des

supports³⁹).

Remarquons que même lorsqu'on donne un rôle central aux EMD, le recensement joue un rôle de référentiel qui, à notre avis, peut être développé. Le fichier sur les mobilités professionnelles domicile-travail à une meilleure fréquence d'actualisation et un taux d'échantillonnage bien plus élevé que les EMD. Toutefois cette source de données ne décrit qu'un seul motif (déplacement domicile-travail) et ne donne aucune précision sur les trajets et les horaires des déplacements. Il paraît alors pertinent d'envisager une stratégie d'enrichissement du recensement par d'autres sources.

Par exemple la carte des antennes-relais de GSM (illustrations 10 et 11), peut permettre d'associer les traces passant par ces antennes aux communes ou aux IRIS. L'analyse temporelle des traces permet en effet de découvrir avec une bonne probabilité l'antenne-relais du lieu de résidence. Dès lors nous pouvons augmenter les éléments de connaissance en couplant les données GSM avec les données du recensement. L'opération consiste à attribuer pour chaque antenne-relais des traces partant de ces antennes aux effectifs du recensement décomptés au lieu de résidence par le support communal. Ce travail de transfert de données d'un support à un autre est ce que nous pouvons appeler un problème de changement de support (COSP⁴⁰). Pour répondre au non-alignement des supports, il faut transférer les données sur un support plus fin, puis ré-agréger vers le support dont on souhaite disposer de l'information.

Effectivement, les supports des antennes relais (les polygones de voronoï des illustrations 10 et 11) et des données du recensement sont différents. Pourtant toutes les possibilités de couplage de ces bases de données dépendent de la possibilité d'accorder les supports géographiques par des méthodes appropriées ne dégradant pas l'information. Afin de faciliter les opérations de couplage **il est donc proposé dans cette thèse d'affiner les localisations du fichier des mobilités professionnelles du recensement.** Le traitement de données du fichier MOBPRO proposé ici a pour objet de répondre aux besoins d'analyses descriptives de la mobilité et à servir potentiellement de données primaires dans les traitements post-enquêtes permettant de couvrir d'autres formes de mobilité que celle issue du motif domicile-travail.

39 - « *In any one study, several different types of data may be collected at differing scales and resolutions, at different spatial locations, and in different dimensions. Many statistical issues are associated with combining such data for modeling and inference.* » (Extrait du résumé du livre de Gotway et al. (2002))

40 En effet, le passage d'une information de deux supports parfaitement alignés revient à transférer l'information entre deux supports identiques.

1.3.3 Comment produire des couples origines-destinations domicile-travail synthétiques avec des localisations fines ?

Les données décrites dans le fichier MOBPRO ont pour supports géographiques les communes de résidence et d'emploi. Nous proposons de ventiler les lieux de résidence et d'emploi à l'intérieur des communes pour produire un fichier « synthétique ». L'idée n'est pas de retrouver les localisations réelles des ménages enquêtés, mais de conserver les répartitions des données attributaires à l'intérieur des communes. A partir de l'échantillon du recensement nous allons redistribuer les observations dans tout le territoire des communes afin de disposer d'une exhaustivité spatiale. Dès lors nous pourrons effectuer des analyses de mobilité sur des territoires plus fins que la commune en choisissant le support le plus adapté au problème.

Les couples domicile-travail sont dits synthétiques parce qu'ils sont issues d'un processus de construction mathématique. Ce processus modélise à des degrés de complexité divers la répartition à grande échelle des lieux de résidence actifs, de leur lieu de travail et du choix du mode de transport, dont le fichier MOBPRO donne la répartition au niveau communal, avec la précision de son taux de sondage.

Ce travail de ventilation à une échelle plus fine rappelle les travaux sur l'inférence écologique (ecological inference) de Gary King (King, 1997). L'inférence écologique consiste à reconstruire des attributs individuels (ou à un niveau plus fin) à partir d'observations agrégées. La méthode peut amener à faire des erreurs ou des amalgames si l'on ne contrôle pas bien la relation entre les supports géographiques. L'exemple célèbre est de déduire faussement à partir d'effectif de criminels et d'étrangers, la proportion des criminels étrangers et nationaux. L'erreur écologique (ecological fallacy) a été bien identifiée (Visser, 1994) (Freedman, 1999) (Kousser, 2001).

Afin d'éviter ce type d'erreur Gary King propose de se servir des méthodes de l'analyse spatiale (Crawford et Young, 2004) et notamment la régression multiple géographiquement pondérée. Pour chaque zone à l'intérieur des communes nous allons appliquer des coefficients pondérateurs. Comme nous traitons de données de mobilité, donc de couples origines-destinations, nous devons effectuer une régression multiple pondérée géographiquement pour les origines et une pour les destinations, tout en préservant le lien d'interdépendance entre ces localisations.

Pour construire ces coefficients nous allons nous servir de données de contrôle externes au recensement. Nous allons tirer ces données de contrôle de la structure spatiale à grande échelle des lieux de résidence et d'emploi obtenues par d'autres sources. C'est l'utilisation de données de contrôle qui va permettre de modéliser l'organisation interne des communes.

Une des limites de ces méthodes est toutefois leur méthode de résolution. Nous utilisons des tirages de Monte-Carlo et il est difficile de valider les agencements de localisations produits. La solution proposée est de ré-agréger les couples origines-destinations synthétiques produits par cette ventilation spatiale à un niveau un peu plus grossier que les données de contrôle mais plus fins que les communes. Nous espérons ainsi retrouver la représentativité statistique et contrôler les risques d'erreur écologique. De cette manière, nous espérons construire une matrice de déplacement domicile-travail sur un support suffisamment fin pour effectuer des analyses de mobilité telles que le demande le nouveau paradigme des politiques de transport et de mobilité durable. Une question est de déterminer de quelle résolution spatiale nous avons besoin en matière de mobilité. Ceci fait l'objet du deuxième chapitre.

Conclusion du premier chapitre

Dans le contexte du développement durable, nous assistons à un changement de paradigme des politiques de transport. Ce changement de paradigme nécessite des données plus précises géographiquement et offrant une information riche sur les profils socio-économiques. En effet, la mobilité des individus évolue en fonction d'une dynamique diffusante des lieux de résidence et d'une dynamique focalisante des lieux de travail. La technologie des transports a permis à ces dynamiques de suivre pendant plusieurs décennies ces chemins opposés mais portés par la croissance économique. Or aujourd'hui, il apparaît depuis bientôt deux décennies que cette croissance n'est pas durable et que le transport contribue trop au changement climatique. Le changement de paradigme imposé par les politiques de facteur 4 nécessite l'abandon de l'amélioration de l'accessibilité par l'augmentation des vitesses pratiquées par les voitures. L'amélioration de l'accessibilité doit être ciblée sur les modes et une organisation des transports les moins consommateurs en énergie fossile. Les nouveaux objectifs de politiques de transport doivent également tenir compte des logiques de choix résidentiel et des stratégies de localisation des activités. Elles conduisent à formuler une demande nouvelle d'information fine sur la mobilité et les territoires à laquelle les dispositifs de la statistique publique ne peuvent pas répondre. Pour disposer de données fines sur la mobilité il faudrait beaucoup augmenter leur taux d'échantillonnage et ce n'est pas possible financièrement.

De nouvelles sources d'information issues des gestionnaires des infrastructures de transport et de télécommunication apparaissent. Leur principal inconvénient est qu'il leur manque une base socio-économique représentative. Elles apportent néanmoins des connaissances sur les variations temporelles. L'avenir paraît au croisement des sources de données. Nous proposons d'asseoir les opérations de couplage de données en s'appuyant sur une source de référence socio-économique pourvue d'un support géographique adaptable et modifiable à volonté, pas trop petit cependant pour conserver une certaine robustesse. Les données de mobilité du recensement offre une information sur les profils socio-économiques mais ont un support spatial trop grossier (le support communal). Nous proposons donc de construire une telle base en produisant un fichier de couples domicile-travail synthétiques localisées le plus finement possible grâce à des méthodes d'inférence écologique contrôlées par des sources à grande échelle.

De cette manière, nous proposons de répondre au besoin de données dans le contexte du développement durable non pas en produisant de nouvelles enquêtes de terrain mais en travaillant sur les supports des données d'enquêtes. Un support est un outil conceptuel dont la résolution conditionne la richesse des analyses. Notre travail de raffinement des données du recensement a pour objectif d'améliorer la finesse des analyses descriptives des territoires. Toutefois, les supports de données remplissent d'autres fonctions. Dans le prochain chapitre nous allons voir qu'effectivement les données sont produites, diffusées, et analysées à l'aide de plusieurs supports et dans certains cas un même support remplit toutes ces fonctionnalités. Nous allons donc observer en quoi les différentes fonctionnalités des supports interagissent entre elles et conditionnent les analyses descriptives.

Chapitre 2 : Les fonctionnalités des supports d'enquêtes

Chapitre 2 : Les fonctionnalités des supports d'enquêtes

Les nouveaux objectifs des politiques de transport, définis à la suite du Grenelle de l'Environnement, requièrent des données décrivant finement les profils socio-économiques et les localisations constituant la demande de mobilité des individus. La demande de mobilité est ici caractérisée par les couples de localisations que les individus doivent atteindre, nous traitons des couples origines-destinations domicile-travail. Ces informations sur les individus des ménages permettent d'effectuer des diagnostics territoriaux. L'exploitation des données de mobilité doit croiser différents critères sociaux, démographiques et spatiaux et a besoin pour cela de disposer d'échantillons importants avec des localisations a priori fines et des profils socio-économique riches.

Une tension existe pourtant entre les finesses de descriptions sociales et spatiales attendues. Pour décrire les zones constituant les supports de l'information géographique des territoires, il faut a priori compter les éléments individuels dans chaque zone et caractériser les observations. Plus les zones du support sont petites et plus les profils socio-économiques collectés sont riches plus les enquêtes sont coûteuses et longues. Les enquêtes par sondage permettent de réduire les coûts et les temps de collecte. Au sein de chaque zone un échantillon de population est tiré au sort et il est considéré comme représentatif des populations de chaque zone. De cette manière nous disposons de profils socio-économiques seulement en localisant de manière peu précise (sur des zones) les observations d'enquêtes. La qualité des données dépend en autres de cette relation entre la taille des zones et du nombre d'observations composant les échantillons de chacune des zones. Les producteurs de données ont comme objectif de contrôler cette relation, de construire des données dites représentatives.

Dans le contexte du changement de paradigme des politiques de transports les utilisateurs des données ont besoin de plusieurs supports d'analyse. En effet, l'utilisation d'un support géographique agrège les localisations des observations. Et, l'agrégation des observations doit répondre aux objectifs des analyses. Par exemple, les analyses des aires de rabattement aux transports en commun demandent des supports de petites tailles (de quelques centaines de mètres), et une description des individus assez riche pour comprendre les processus de choix des modes de transport. En revanche une étude sur les formes de villes facilitant l'utilisation des modes doux va nécessiter l'utilisation d'un support différent que celui des aires de

rabattement. Selon les types d'analyse les supports d'enquêtes attendus par les utilisateurs sont différents.

Les supports sont effectivement de plusieurs natures, nous pouvons distinguer les supports zonaux et les supports linéaires. Les premiers peuvent être classés en trois types : les supports morphologiques regroupant les informations selon un ou des critères particuliers, des supports administratifs et des supports réguliers. Les seconds se distinguent également selon leur finesse de description. Dans le premier cas, les supports d'analyse zonaux se construisent par agrégation en partant du support de la base de donnée. Ils servent à mesurer des stocks. Dans le second, les supports servent à mesurer des flux obtenus par comptage ou modélisation. Le transfert de données d'un support zonal vers un support linéaire demande des traitements post-enquêtes sophistiqués. La modélisation à 4 étapes nous paraît bien illustrer ce cas. Nous examinons ici les conditions que doivent remplir ces différents supports spatiaux et les possibilités qu'ils offrent.

Dans un premier temps, nous analyserons les trois fonctions que doivent remplir les supports spatiaux des enquêtes (2.1) :

- une fonction d'**analyse territoriale**, appuyée sur un zonage et/ou un réseau, en vue de l'application des politiques, qui doit contenir un échantillon suffisant pour croiser plusieurs données attributaires.
- une **fonction de production ou de diffusion** des données qui est généralement le support le plus fin disponible et qu'on appellera support de base des données.
- une **fonction de référentiel pour la localisation** des individus ou de leurs déplacements, essentielle pour la mesure des distances des déplacements ou des trajets à l'intérieur de ceux-ci.

Ces trois fonctions entretiennent des liens d'interdépendances. Nous verrons qu'un support régulier ouvre des perspectives pour concevoir un zonage d'analyse adapté à chaque problème posé, mais en tenant compte des contraintes du MAUP (Modifiable Areal Unit Problem).

Ensuite, nous verrons comment ces fonctions sont remplies dans le cas d'une enquête à taux d'échantillon relativement faible, en prenant l'exemple des Enquêtes Ménages-Déplacements (2.2), puis nous montrons les méthodes de construction des données du recensement et les choix de support et d'échantillonnage (2.3).

2.1 Les trois fonctions des supports spatiaux des données de mobilité

2.1.1 Les supports d'analyse

Les données de mobilité sont conçues pour fournir une représentation fidèle des comportements de mobilité sur un territoire. Cette représentation doit être suffisamment précise pour que l'analyse des données aide à mettre des politiques en œuvre. Or ces politiques sont localisées dans l'espace, il faut alors des données sur des supports d'analyse permettant de cibler les lieux de ces politiques. C'est pourquoi nous commençons par examiner les supports spatiaux de la représentation de la mobilité sur un territoire. Les supports peuvent être de deux natures : un zonage basé sur la structure morphologique ou administrative du territoire, ou un réseau basé sur la structure de l'offre de transport. Nous donnons quelques exemples de supports spatiaux construits pour analyser les données de l'Enquête Globale Transport d'Île-de-France. Nous commençons par les supports zonaux puis linéaires.

2.1.1.1 Un support zonal obtenu par agrégation de zones de la base de données pour la description et la communication

Dans cette section nous prenons le rôle d'un utilisateur de données. Nous souhaitons faire des analyses descriptives de la mobilité. Nous remarquons que l'Observatoire de la Mobilité en Île-de-France (Omnil) a effectué des analyses disponibles pour le grand public. Ces analyses sont réalisées dans le cadre du Plan de Déplacement Urbain de la région.

- « *Les analyses de l'EGT 2010 seule ou en comparaison avec l'EGT 2001 s'appuient sur un découpage de l'Île-de-France basé sur la morphologie des territoires, retenu dans le projet de Plan de déplacements urbains d'Île-de-France, arrêté en février 2012, qui distingue :*

- *Paris,*

- *le cœur d'agglomération, qui comprend les communes – en continuité de bâti avec Paris – urbanisées (au moins 80 % d'espaces urbanisés au mode d'occupation du sol de 1999) et denses (au moins 80 habitants et emplois à l'hectare urbain construit),*

- l'agglomération centrale, qui correspond à l'agglomération parisienne définie par l'Insee (1999),
- les autres agglomérations (hors « agglomération centrale ») définies par l'Insee,
- l'espace rural ». (Omnil, Enquête Globale Transport⁴¹, p.3)

Cette typologie de l'espace est centrée sur l'urbain dense. Elle fait l'objet d'une cartographie dans la plaquette de l'Omnil (illustration 13).

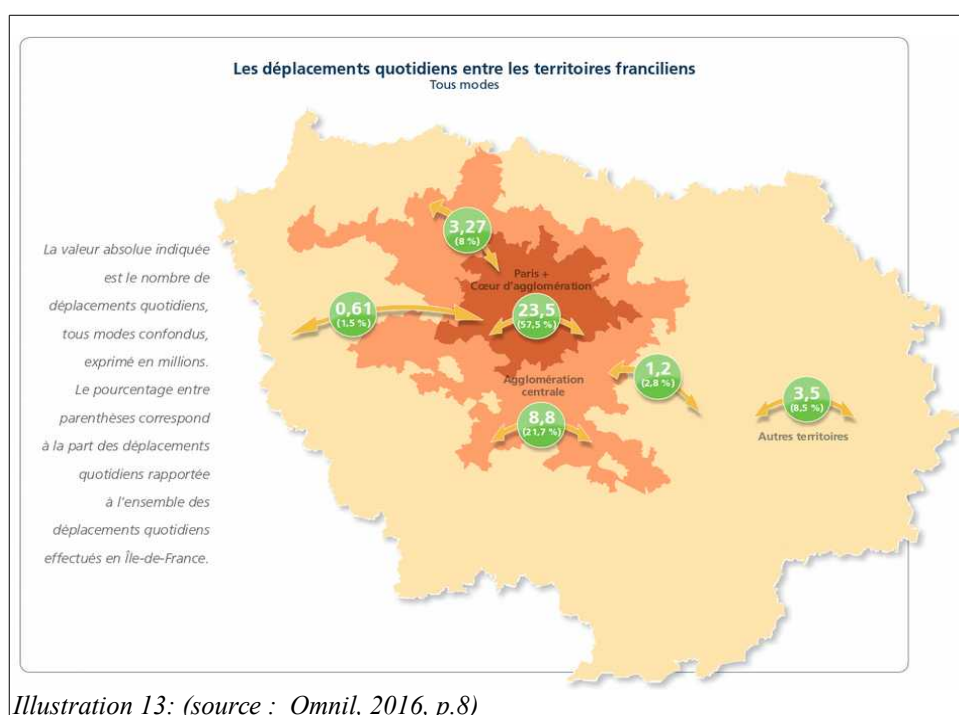
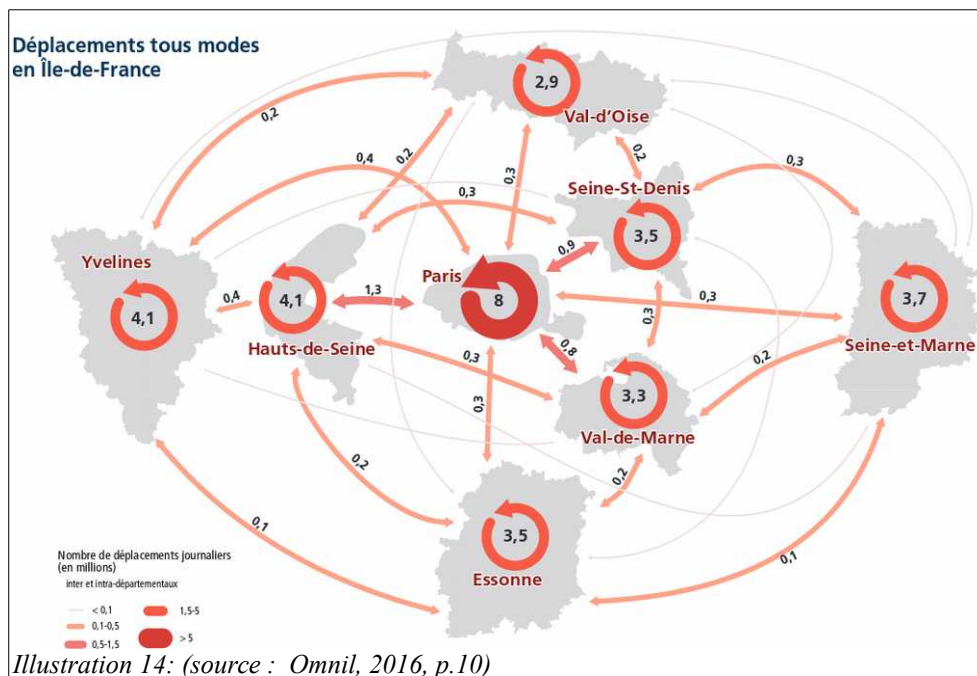


Illustration 13: (source : Omnil, 2016, p.8)

41 Omnil - Enquête Globale Transport - [en ligne] (consulté le 31/06/2026)
<http://www.omnil.fr/IMG/pdf/egt2010_enquete_globale_transports_-_2010.pdf>

L'Omnil utilise aussi le support des départements de la région pour représenter les flux entre les départements (illustration 14).



À l'aide de ces supports nous observons bien la polarisation parisienne dans l'espace régional francilien. Plus généralement, un support d'analyse doit permettre d'identifier des zones d'intérêt et/ou de compétence pour des organisations ou des collectivités territoriales.

Nous comprenons bien que pour obtenir ces découpages nous devons disposer de données à l'intérieur des zones de ces supports, pour la deuxième carte au moins à une échelle infra-départementale. Les départements doivent contenir des échantillons robustes et sont souvent obtenues par agrégations de zones plus petites. L'Omnil a donc agrégé les données à l'échelle des départements puis en a fait des analyses descriptives. Dans ce cas, l'échelle départementale permet bien de communiquer/cartographier la centralité parisienne.

En tant qu'utilisateurs de données nous comprenons que si nous souhaitons des effectifs d'observations importants pour que les analyses soient robustes, il est possible d'augmenter la taille des échantillons par agrégation de zone. Un nombre suffisant d'observations permet de prétendre à une représentativité socio-économique à l'intérieur des zones. L'agrégation permet de faire gonfler les effectifs mais nous perdons en précision géographique. Le besoin de disposer d'un gros échantillon posera donc le problème de l'agrégation (perte de précision)

dans les cas où les données sont obtenues par enquête avec un faible taux d'échantillonnage.

De plus, l'agrégation pose des problèmes de lissage. L'opération d'agrégation pour construire des supports d'analyses lisse l'information (*Smoothing effect*⁴² (Openshaw, 1977) et augmente artificiellement les différences entre zones dans la répartition spatiale des variables à étudier. Du point de vue intra-zone, la répartition spatiale est lissée à une valeur moyenne ou modale. Pour les variables de localisation l'opération de lissage réduit une distribution des points d'observation à une position centrale. Cette opération renforce les différenciations spatiales des indicateurs statistiques car les variables de localisation ne sont plus continues mais discrètes. Elles créent alors des discontinuités et des seuils. En fonction des besoins il est aussi nécessaire d'agréger les variables non spatiales, notamment socio-économiques. Cela peut conduire à la création d'un individu-moyen (Quetelet, 1835), (Desrosières, 2002), (King, 1997b). Pour éviter de faire des hypothèses trop fortes, les utilisateurs de données attendent un support de donnée le plus fin possible pour pouvoir disposer du plus grand éventail de possibilité pour agréger à façon. Cela dépend du support de la base de donnée diffusé par les producteurs.

Avant de présenter la fonctionnalité du support de la base de données, nous montrons un autre type d'adaptation du support de base de données que celui de l'agrégation : le passage d'un support zonal à un support d'analyse linéaire. Cela permet d'introduire le concept de données modélisées pour répondre aux faibles tailles d'échantillon et donc de proposer une première alternative à l'agrégation pour disposer d'échantillon plus grand. Ensuite nous reprendrons le fil annoncé par l'introduction du chapitre : la fonctionnalité de support de la base de données diffusée.

42 Aujourd'hui nous connaissons bien les sources de la variabilité de ces analyses. : -« *The decades of research on the MAUP and the ecological inference problem have clearly identified the source of these problems The smoothing effect that results from averaging is the underlying cause of both the scale problem in MAUP and aggeration biais in ecological studies. As heterogeneity among units decreases through aggregation, the uniqueness of each unit and the dissimilarity among units are reduced. As Openshaw (194) noted, « wheter the ecological fallacy problem exists or not depends on the nature of the aggregation. A completely homogeneous grouping system would be free of this problem ».* (King,2004,p.234)

L'auto-corrélation spatiale est intimement liée à cette problématique. « *There is, however, another mitigating factor : spatial autocorrelation. The decrease in variability is moderated by negative autocorrelation among the original observations and exacerbated by negative autocorrelation (Arbia, 1986 , and Cressi 1993, both illustrate the effcet of autocorrelation on the variance of the sample mean). When areal units are similar to begin with, the aggregaion process results in much les information loss than it does with highly dissimilar unirs.* » (King,2004,p.234)

2.1.1.2 Un support linéaire pour l'analyse du trafic : le passage par des données modélisées

Outre l'Omnil, d'autres acteurs analysent la mobilité en Île-de-France : le Syndicat des Transports d'Ile-de-France (STIF) et la Direction Régionale et Interdépartementales de l'Équipement d'Ile-de-France (DRIEA). Le premier organisme s'occupe exclusivement des transports en commun, le deuxième a parmi ses fonctions l'exploitation et l'aménagement du réseau routier national en Île-de-France. Chacun d'eux développe un modèle de prévision de trafic dédié à leur réseau de transport respectif. Le modèle ANTONIN pour les transports publics et le modèle MODUS pour le routier. Les supports de l'information géographique utilisés sont les zones d'émission et de réception des trafics dans une matrice de demande de déplacement. Cette matrice est construite par les experts à l'aide de l'EGT et d'autres sources de données (données de comptage de leur réseau, données d'enquêtes origine-destination, enquête à bord). D'après la littérature scientifique en modélisation du trafic, le support contenant les observations de déplacements (les flux) est construit avant de procéder à l'enquête. Puis, l'étape n°1 et n°2 de la célèbre chaîne de modélisation à quatre étapes (génération, distribution, choix du mode, affectation) extrapolent les données d'enquêtes. Des flux de déplacements vont être modélisés à partir d'un échantillon d'origines-destinations issues des enquêtes. Selon la finesse du modèle, les motifs et les profils socio-économiques sont pris en compte dans l'extrapolation. La précision géographique dépend du support d'enquête choisi. Mais pour palier à la faiblesse d'échantillonnage, il est donc procédé aux étapes de génération et de distribution. La méthode la plus utilisée pour générer les flux est la régression linéaire catégorielle (Bonnell, 2001). Les variables explicatives des régressions sont les données du recensement de la population. Le support géographique de l'enquête doit donc être aligné à celui du recensement. Par exemple, si un certain nombre d'origine de déplacement est observé dans une zone de tirage de l'enquête un coefficient pondérateur est appliqué pour estimer le nombre total d'origine de déplacement de la zone de tirage à partir de la population totale approché par le recensement. Ensuite, s'il est souhaité disposer d'une matrice plus fine que les zones de tirage, les résultats de l'étape de génération et de distribution de trafic peuvent être désagrégés après coup.

- « Les données de recensement constituant le plus souvent une base indispensable, il est généralement recommandé d'opter pour un zonage compatible avec les îlots de l'INSEE⁴³.

43 L'îlot était l'unité géographique de base pour la statistique et la diffusion des recensements de la

Enfin, il est généralement recommandé de respecter le découpage communal. Cette contradiction entre les objectifs d'étude et les données disponibles conduit parfois à considérer différents niveaux de zonage lors d'un exercice de modélisation. Le zonage peut être ainsi « relativement grossier » pour l'étape de génération pour laquelle on ne dispose pas forcément des données socio-économiques à un niveau suffisamment fin, voire également pour celles de distribution et de répartition modale. En revanche, il sera beaucoup plus fin pour l'étape d'affectation pour laquelle une plus grande précision du zonage est souvent nécessaire. Mais dans ce cas, l'éclatement de la matrice de demande estimée sur un zonage « grossier » en une matrice de demande correspondant au zonage le plus fin est évidemment problématique» (Bonnell, 2001, p.311). Le support de la matrice de déplacements construite dépend donc des choix d'enquêtes et des limites des supports du recensement.

Les supports d'analyses en lien avec les réseaux de transport routier:

La modélisation de transport ne concerne pas seulement les étapes de création de matrice de déplacement. Elle consiste à appliquer sur un réseau de transport, les flux dénombrés par les matrices de déplacements.

Nous faisons une aparté pour présenter les éléments de la littérature de la modélisation de trafic sur la question du choix du support en lien avec les réseaux de transports. Cette parenthèse met en avant le rôle des supports d'analyse dans la schématisation des réseaux et le fait qu'ils peuvent donc influencer les résultats de modélisation de trafic. Les supports ne doivent être ni trop fins (car le modèle n'est pas assez fidèle) ni trop grossier car il ne permet pas de juger de l'effet des aménagements de taille courante. Il y a un équilibre à trouver par un travail de spécialistes des modèles et du territoire. Les supports d'analyse ne sont donc pas déterminés a priori mais construits au fur et à mesure à partir des retours d'expérience et des évolutions des territoires. Nous refermerons la parenthèse à la section 2.2.

Pour étudier les effets de choix des supports zonaux et linéaires (les réseaux de transports) et donc produire des résultats en connaissance de cause il faut faire varier la précision de chacun des supports et examiner les différences dans les résultats. En modélisation de transport routier beaucoup de littérature existent, citons par exemple : (Bovy et Jansen,1993) (Crevo,1991) (Ding,1994) (Khatib et al, 2001) (Chang et al., 2002) (Choi 2003) (Zhang et

population jusqu'à celui de 1999. Désormais, la brique de base pour la diffusion des résultats du recensement rénové, qui débute avec le millésime 2006, est le quartier IRIS et l'îlot est abandonné.

Kukudia, 2005) (Martinez,2007) (Jeon et al., 2010) (Jeon, 2012) et pour une revue de littérature (Nix, 2009).

Ces travaux agrègent les supports zonaux et les graphes représentant les réseaux viaires et comparent les résultats avec divers indicateurs.

Tableau 2 : Exemples de littératures traitant de la résolution des supports en modélisation du transport routier			
Auteurs	Supports	Réseaux	Indicateurs
Bovy and Jansen (1983)	Agrégé	Fin Moyen grossier	-RMSE -PRMSE
Crevo (1991)	Agrégé Détaillé	Fixe	-Coefficient de corrélations -Coefficient de détermination
Ding (1994)	Agrégé	Fixe	-Ratio de flux intra-zone -Vitesse moyennes -Véhicule par heure -Véhicuel par km
Khatib (2001)	Comté census tract census block	Moins détaillé Détaillé	-Coefficient de corrélation et détermination -Percent Root Mean Square Error -Interzonal et intrazonal trips -ANOVA -Annual average daily traffic (AADT)
Chang et al. (2002)	Agrégé	Moins détaillé Détaillé	-PRMSE
Choi (2003)	Détaillé	Fixe	-Coefficient de corrélation -Khi deux-roues -RMSE
Zhang and Kukadia (2005)	Agrégé	Fixe	-significant test of mean
Martinez et al. (2007)	Frequencias Grid TAZ TAZ BBRI	Fixe	-Pourcentage de flux intra- zones-Pourcentage de flux non significatifs dans la matrices -75° percentile of the zone equivalent radius
Source : Jeon et al., 2010			

Sur l'agglomération de New Castle, Crevo (1991) teste un support de 9 zones et un autre de 23 zones et ne remarque pas de différences. Toutefois, la subdivision en 9 et 23 zones est, dans les deux cas grossière, ce qui expliquerait l'absence de différences significatives.

Ding (1994) travaillant sur les régions sud-coréennes découpe l'espace en 8, 15, 25, 35, 60, 75, 100 et 130 zone. A partir de 50 zones, les effets sur l'estimation de la demande est bien plus petit. Nous en déduisons qu'il faut au moins disposer de 50 zones.

Choi (2003) teste 3 supports sur la ville de Eujeongbu city (Corée) ayant respectivement 115, 54 et 13 zones. Il en ressort que le support moyen est plus efficace selon un rapport de coût de production et de valeur ajouté de la finesse des supports.

Les travaux de Martinez et al. (2007) sur l'agglomération de Lisbonne démontrent également qu'un support fin ne produit pas systématiquement de meilleurs résultats. Ces deux derniers résultats pointent du doigt le fait que les supports doivent être en adéquation avec la représentation plus ou moins schématique des réseaux linéaires de transport. Un grand nombre de zones est pertinent si l'on dispose d'une représentation fine des infrastructures de transports.

Bovy et Jansen ont quant à eux testé 3 réseaux de transport plus ou moins détaillés (fine-level network, arterial and collector, only arterial roads) et 3 supports (1 285, 183 et 47). Il en ressort que le niveau de finesse du réseau et du support de la matrice a bien une influence. Mais qu'au delà d'un niveau de finesse les améliorations sont marginales.

Enfin, les résultats de Khatib et al. (2001) en Idaho (USA) ajoutent un effet supplémentaire. Contrairement aux travaux précédents ils concluent que l'utilisation d'un réseau fin apporte toujours de meilleurs résultats (marginaux ou pas) peu importe le niveau de précision des supports. Les auteurs concluent également que l'utilisation d'un graphe grossier apporte de meilleurs résultats lorsqu'il est couplé avec un support grossier. Cette combinaison semble plus efficace qu'un graphe grossier avec un support fin.

Nous pouvons encore citer le travail de Sean Nix (2009) qui répertorie un certain nombre de travaux sur la question de la quantification de l'impact du problème de la modification des zones dans les modèles de la demande de transport. Les résultats concordants de ces travaux démontrent qu'il faut disposer d'un support de déplacement en adéquation avec le niveau de résolution du graphe formalisant le réseau routier. Avec une matrice de déplacement grossière, il est possible de modéliser l'effectif moyen des tronçons des plus grosses infrastructures. En revanche, il est nécessaire de disposer d'une matrice de déplacement fine si l'on souhaite modéliser le trafic sur le chevelu fin des infrastructures. Or, une matrice fine doit s'appuyer sur des données également fine, qu'elles soient obtenues directement par des zones de tirages fines ou par désagrégation à condition de bien contrôler l'opération.

Les supports d'analyses en lien avec les modes de transport alternatifs à la voiture personnelle

Dans le changement de paradigme politique instauré par le Grenelle de l'Environnement nous devons nous interroger sur les choix de supports visant les systèmes de transports alternatifs à la voiture personnelle à carburant fossiles. En effet, un repérage spatial des ménages potentiellement concernés par ces mesures doit être effectué et des zonages d'analyse particuliers peuvent être construits. Pour examiner les comportements actuels et analyser les changements d'offre à effectuer pour faire changer les comportements de mobilité, il faut des données sur des supports adaptés.

Les transports en commun : Le choix du support pour la modélisation du trafic en transports en commun est plus complexe que pour la voiture. Idéalement, chaque zone doit contenir une seule station d'entrée et de sortie. Toutefois comme nous savons que les aires de rabattement autour des grandes stations de transports en commun (gares) sont plus étendues que les bus, le praticien est en droit d'hésiter. Faut-il regrouper les stations de bus desservant les grandes gares dans une même zone ou faut-il différencier chaque station de transport en commun ? Mais dans le deuxième cas, Bonnel nous renseigne *« Il est clair que cette méthode peut difficilement être généralisée à l'ensemble d'une agglomération, car elle conduirait à un nombre de zones extrêmement élevé et donc à un coût de production de données et d'analyse très lourd »* (Bonnel, 2001, p.319).

Les véhicules partagés: Pour les systèmes de transport partagés de types vélib' ou autolib' il faut distinguer les zones résidentielles et d'activités des lots de stationnement pour pouvoir mesurer les distances d'accès aux stations d'entreposage. A l'instar des aires de rabattement, il est supposé que les distances d'accès sont courtes (quelques centaines de mètres maximum) alors le support doit être suffisamment fin pour ne pas produire de déplacements intra-zones.

La mise en place de systèmes de véhicules partagés apporte un défi supplémentaire. Les objectifs sont de déterminer les localisations des parcs de véhicules et de dimensionner le nombre de places de véhicule afin de tendre vers une auto-organisation (Leurent, 2012). Il est entendu ici par auto-organisation un système où la rotation des véhicules se ferait seulement par les utilisateurs et il n'y aurait pas de parc de véhicules vides ou constamment pleins.

Le co-voiturage : Pour inciter les citoyens à effectuer du co-voiturage, il faut pouvoir distinguer les zones résidentielles des points de rendez-vous accessibles par les utilisateurs du co-voiturage. A l'instar des transports en commun et de l'auto-partage des véhicules, il doit être construit un support fin pour analyser les distances de rabattement. De plus, l'étude du co-voiturage nécessite de connaître des paires d'origines-destinations communes à des citoyens pour les inciter à mutualiser leur moyen de transport.

Les modes doux et l'activité physique : Puisque les distances en marche à pied ou en vélo sont relativement de petites longueurs il faut disposer d'un support avec une résolution fine. Celui-ci doit d'avantage respecter les discontinuités naturelles qui sont bien plus contraignantes que dans l'utilisation des autres mode de transport.

Les reports modaux : Enfin, l'étude des reports modaux apporte sans doute le plus de contraintes de production. Pour étudier les potentiels de reports modaux, il est utile de simuler le déplacement réellement effectué et de simuler le déplacement effectué avec un autre mode. Il faut disposer de supports adéquats pour les deux systèmes de transport étudié. Dans le cas de la comparaison entre un déplacement en voiture particulière et un déplacement en transport collectif il faut que le support soit centré à la fois sur les carrefours routiers et à la fois sur les stations d'entrée du système de transport collectif. Cela rappelle les difficultés énoncées par Bonnel lors de la codification de plusieurs réseaux de transport (multi-graphe). Ceci est plus aisé si nous disposons au départ de supports fins pour les systèmes de transport à comparer. Il est possible de recombinaison à façon les supports.

En conclusion de ce panorama nous observons que chaque système de transport nécessite des supports spatiaux différents. Le dernier point énoncé pour étudier les reports modaux est à la base de la proposition méthodologique pour produire ces différents supports. Il faut disposer d'un jeu de données localisant finement les observations et les équipements spatiaux de transport pour pouvoir agréger selon les besoins. L'idée est de construire un jeu de données localisé finement dont la propriété est de permettre la ré-agrégation à façon. Pour autant les supports des bases de données ne remplissent généralement pas cette condition. Ils ont une autre fonction à respecter.

Nous arrêtons de prendre le point de vue d'un utilisateur. Prenons maintenant le rôle d'un producteur de données et décrivons la fonction d'un support de production de données.

2.1.2 Les supports des bases de données de mobilité diffusées par les producteurs

2.1.2.1 La production de données suffisamment représentatives pour être diffusées

Nous prenons maintenant le rôle d'un producteur de données. Son travail est de collecter des observations sur le terrain et de les transmettre aux utilisateurs en stockant l'information dans un support que nous pouvons appeler de base, et qui satisfait aux contraintes de la production et de la diffusion⁴⁴.

Dans une enquête par tirage au sort, le tirage est organisé par zone afin que l'échantillon soit grossièrement représentatif de la population totale à l'intérieur des zones. C'est la fonction principale du support pour un producteur. De plus, si l'enquête suit un plan de sondage, la non-réponse peut avoir un effet négatif qui peut être redressé à l'aide des données du recensement. Ainsi plus l'effectif tiré au sort est faible plus le support de tirage doit être aligné à celui du recensement pour redresser l'échantillon. Par construction, le support de base est le zonage disponible le plus fin pour effectuer des analyses descriptives.

Une base de données de déplacements a pour objectif de contenir un nombre suffisant d'enquêtés pour donner une indication sur la capacité de chaque zone à générer des déplacements. Plus les échantillons par zone sont grands, plus l'échantillon est représentatif des populations totales à l'intérieure des zones. Dans le jargon statistique nous parlons également d'hypothèse d'homogénéité (2 échantillons sont issus de la même population sont dit homogènes). Mais ces hypothèses que nous sommes obligés de prendre dans le cas d'enquêtes statistiques ne sont pas les seules qui déterminent le support de la base de donnée à diffuser. La littérature présente 10 règles pour construire un zonage dédié à l'étude des transports:

1-Achieve a maximum of homogeneity inside the newly created zones, which is important for the trip generation and the modal split phase of the model sequence.

2-Retain a maximum of interaction between newly established zones or a minimum of intrazonal trips, which is an important requirement for the trip distribution and trip assignment models.

44 Il peut y avoir deux supports différents, le support de la production n'est pas toujours conservé pour cause de confidentialité.

3-Limit the number of trip ends for the newly created spatial entity in order to avoid overloading of the adjacent street network in the assignment phase.

4-Respect physical, political, ad historical boundaries as far as they are of importance from planning point of view.

5-Avoid undesirable shapes of newly created zones

6-Group only adjacent basic spatial units

7-Generated only connected zones

8-Avoid the formation of islands, which means zones that are completely contained in another zone

9-Obtain a zonal system in which the number of households, population, area, or trips generated and attracted are nearly equal in each zone (the variation with respect to one of these variables should be kept as small as possible)

10-Base the delineation of the zonal boundaries on the census boundaries. (Baas, 1980, p.1)

La contrainte numéro une à respecter est l'homogénéité des observations à l'intérieur des zones (point n°1). Mais il faut respecter les limites naturelles (fleuves, relief), politiques et historiques (point n°4) qui ne définissent pas forcément des zones homogènes. Il faut prendre en compte le réseau routier (point n°3) tout en minimisant le nombre de déplacements intra-zones et en maximisant les déplacements entre les zones (point n°2) . Il faut respecter toutes ces contraintes en faisant le choix des zones de manière à agréger progressivement des sous-zones adjacentes (point n°6) et en évitant des zones contenues dans d'autres (point n°8), tout en construisant le moins possible de couples de zones sans déplacement (point n°7). Les zones doivent être délimitées de manière à respecter les zones de diffusion des recensements de population pour permettre le redressement. Enfin, les zones doivent contenir approximativement le même nombre d'observations (point n°9).

Les règles à respecter pour un producteur sont nombreuses, peut être un peu trop. La première règle est de diffuser des données représentatives donc s'appuyant le plus possible sur les hypothèses d'homogénéité. A l'intérieur d'une zone les observations doivent se ressembler. Et recouvrir si possible tout l'espace de la zone.

Cependant, un support est « homogène » seulement par rapport à une ou plusieurs variables d'intérêt, quand la valeur de ces variables varie peu et de manière aléatoire à l'intérieur des zones. Ainsi avec un support homogène, peu importe où nous nous trouvons dans la zone, la proportion étudiée ne change pas (illustration 16). Mais cette définition n'est pas

complète. Un support homogène doit être construit en ajustant le périmètre aux valeurs des variables. La construction d'un support de l'information doit tendre à minimiser la variance intra-zone (intra-groupe) et maximiser la variance inter-zone (inter-groupe) des variables d'intérêt. Le choix du support revient donc à discrétiser les coordonnées géographiques des observations statistiques en classes spatiales des variables d'intérêt. La première variable qui peut être utilisée est souvent la densité de population. En considérant d'autres variables le problème est dit contraint puisque nous souhaitons regrouper spatialement des observations qui se ressemblent alors que rien ne dit qu'elles le soient lorsqu'elles sont à proximité. Un support qui essaye d'homogénéiser une variable d'intérêt donne lieu à un support irrégulier.

La construction de tel support est un processus pouvant respecter de multiples contraintes qui sont parfois contradictoires. Parfois même il se peut qu'il n'existe aucune solution satisfaisante. Et, il est difficile de construire des zones devant respecter des "fonctions-objectifs" maximisant l'homogénéité de plusieurs critères tout en préservant la connexité des zones. Le nombre de critères possibles à considérer doit être faible. Or dans le cas de données bi-localisées nous avons déjà deux types de variables à homogénéiser : celles qui déterminent les origines des déplacements et celles qui déterminent les destinations. Nous avons vu au chapitre 1 que les lieux d'emploi sont plus concentrés dans l'espace que les lieux de résidence. Cette méthode peut conduire à produire des zones très petites, biscornues ou avec peu de population ou d'activité. A l'inverse, il n'est pas forcément vrai qu'une zone petite ou de faible densité soit homogène mais cela semble être une règle communément admise.

L'application des règles de Baas pour construire un zonage de base de données de déplacement suppose un examen précis du territoire et une connaissance des zones à analyser. En outre, l'application d'une méthode standard qui assure la comparabilité des enquêtes peut avoir pour inconvénient de se baser trop sur certaines règles, en particulier la 4^{ème} (le respect des zonages administratifs), comme on le verra plus loin dans le cas des enquêtes à faible taux d'échantillonnage. Afin de répondre aux besoins divers des analyses, les nouvelles orientations en matière de diffusion de statistiques publiques sont en évolutions et se dirigent vers la diffusion de supports réguliers de petites tailles ayant vocation à être agrégées à façon par les utilisateurs et non plus par les producteurs.

2.1.2.2 La solution vers laquelle les producteurs se tournent est un support régulier constitué de petites mailles

La directive européenne Inspire (Inspire, 2007) vise à établir une infrastructure d'information géographique dans la communauté européenne pour favoriser entre autres la protection de l'environnement. L'ordonnance du 21 octobre 2010 a transposé dans le droit français la directive européenne Inspire, qui, pour favoriser la protection de l'environnement, impose aux autorités publiques, d'une part de publier sur Internet leurs données environnementales géographiques, d'autre part de les partager entre elles. Ces dispositions ont pour objectif de faciliter la réalisation d'études et donc la conception, le suivi et l'évaluation des politiques environnementales, en favorisant la prise de décision dans un cadre démocratique, avec un bon niveau d'information de tous les acteurs et du grand public.

Sans qu'il y ait un caractère obligatoire l'information géographique diffusée doit permettre l'interopérabilité. Pour ce faire il est proposé de diffuser l'information sur des supports répondant à des recommandations techniques. Les supports préconisés par la directive INSPIRE sont des maillages utilisant la projection géodésique ETRS98. Plusieurs systèmes de projection peuvent être utilisés selon les échelles (IGN. Projection INSPIRE⁴⁵). Un support répondant à la directive est donc un support utilisant la projection géodésique ETRS98 et dont les carreaux sont alignés aux supports européens. Le centre de la grille européenne a comme coordonnées 4321000,0 m ouest et 3210000,0 m sud (Anonin, 2005, p.17). La résolution des supports doit être en base 10 (1km, 10km, 100km) mais il peut être utilisé des résolutions intermédiaires (200m, 500m) du moment que les supports ont les mêmes centre de grille et donc entretiennent des relations hiérarchiques.

La directive propose d'utiliser une grille régulière sur un référentiel géographique commun à tous les pays membres de l'Union européenne (les zones régulières d'un carroyage peuvent être appelées des mailles). L'idée est donc de produire un support neutre, sans a priori c'est-à-dire sans processus de découpage s'appuyant sur des données. Il est souhaité construire un support homogène spatialement (chaque maille ayant la même superficie). « *La fonction des nouvelles mailles n'est plus tant de pouvoir restituer de l'information maille par maille, comme au niveau géographique supérieur, que de permettre de décrire le fonctionnement*

45 IGN. Projection INSPIRE. [en ligne] (Consultée le 06/06/2016)
<http://geodesie.ign.fr/contenu/fichiers/documentation/rgf93/Projections_INSPIRE_ETRS89.pdf>

global de l'ensemble des mailles qui constitue ce niveau géographique supérieur. » (Insee. Données carroyées- Pourquoi des carroyages, 2016)⁴⁶

Pour faciliter l'interopérabilité⁴⁷ le rapport du CNIS du groupe de travail de Francis Cuillier va plus loin, il propose une « géographie du carreau » (Cuillier, 2009, p.37) c'est-à-dire diffuser les données publiques sur un même support carroyé. « *Il est par ailleurs indispensable que l'utilisateur puisse accéder via un point unique à l'ensemble des chiffres clés disponibles « au carreau ». Le groupe préconise à cette fin que l'Insee soit désigné comme le coordonnateur de la diffusion des chiffres clés « au carreau » et qu'à ce titre il soit le pilote d'un projet « diffusion de données au carreau » (Cuillier, 2009, p.37).*

« Les informations souhaitées correspondent à celles qui sont aujourd'hui disponibles à l'IRIS et, pour certaines, sur les quartiers « politique de la ville », à savoir :

- Les chiffres clés du recensement de la population (qui seront définis fin 2009).*
- Les chiffres clés présents dans la rubrique «Données urbaines infracommunales par quartier » du site de l'Insee : revenus/niveau de vie (allocataires CAF, revenus fiscaux), bénéficiaires de la CMU, Entreprises (Clap), Emploi salarié (DADS ou Clap), activité (chômage, qualification des actifs).*
- Les chiffres de présence d'équipements (BPE) actuellement disponibles au niveau de l'IRIS sur le site de l'Insee. Ces informations pourraient être enrichies de données actuellement non disponibles au niveau infra-communal [...]*

Cette mise à disposition suppose que les données soient diffusables, eu égard aux règles de secret statistique et donc :

- que ces règles soient éventuellement revisitées en ce sens,*
- que des méthodes de « brouillage » soient appliquées de façon à respecter ces règles sans pour autant perdre du « sens » statistique à ce niveau géographique. Cela suppose aussi que*

46 Insee, Données carroyées- Pourquoi des carroyages [en ligne] consulté le 05/04/2016 < http://www.insee.fr/fr/themes/detail.asp?reg_id=0&ref_id=donnees-carroyees&page=donnees-detaillees/donnees-carroyees/donnees_carroyees_carroyage.htm>

47 INSPIRE définit l'interopérabilité comme étant la : « *Possibilité d'une combinaison de séries de données géographiques et d'une interaction des services, sans intervention manuelle répétitive de telle façon que le résultat soit cohérent et la valeur ajoutée des séries et des services de données renforcée.* » [Directive INSPIRE: DIRECTIVE 2007/2/EC DU PARLEMENT EUROPEEN ET DU CONSEIL DU 14 mars 2007, Article 3, page L108/5]

les différents producteurs s'accordent pour une mise à disposition à ce niveau « carreau ». Cette mise à disposition suppose enfin que le « carreau » soit un objet géographique bien défini. Dans le cadre d'INSPIRE, une définition du carroyage normalisé européen existe . Il serait utile que la définition du carreau de base français permette de reconstituer le carreau européen sans perte de précision statistique » (Cuillier, 2009, p.19).

En 2010, l'Insee a diffusé un premier jeu de donnée nommé carroyage de population. Les données issues des fichiers administratifs, en l'occurrence les fichiers de la taxe d'habitation ont été exploités pour proposer une base de donnée sur le dénombrement des populations au lieu de résidence sur un carroyage de 1km répondant aux prérogatives INSPIRE (Lipatz, 2011). Les données ont été ensuite diffusées sur d'autres supports : un carroyage intermédiaire de 200m, et plus récemment le carroyage INSPIRE de 1km a été remplacé par un support rectangulaire de maille irrégulière ne répondant plus directement aux recommandations techniques INSPIRE (encadré 1). Les données carroyées donnent des informations sur les populations localisées aux lieux de résidence. Les données sont produites avec comme source d'information les revenus fiscaux localisés, elles ne sont pas des sous-produits des recensements de population.

Encadré 1 : Historique de la diffusion des données carroyées de l'Insee.

Fin 2010 : Diffusion de la population résidente de 2007 sur des grilles à mailles régulières de 1km de côté. Source RFL, cadastre + information sur l'occupation des sols pour les communes dont le cadastre n'était pas numérisé.

Septembre 2011 : Diffusion de la population résidente de 2009 sur des grilles à mailles de 200 mètres de côté, et actualisation de celles à 1km. Sources : RFL, cadastre maintenant disponible sur la France entière.

Janvier 2013 : Diffusion de 23 nouvelles variables socio-économique issues du RFL 2010 sur les grilles à mailles de 200m de côté. Elles concernent :

- la population résidente (totale, par classe d'âge) : 9 var.
- Les ménages (total, taille, présents depuis + ans, propriétaires/ locataire, en logement individuel / collectif, déclarant une résidence principale / secondaire, surface des résidences principales) : 10 var
- les revenus (UC, médiane du revenu par UC, ménages en dessus / en dessous du seuil de haut / bas revenu) : 4 var.

La confidentialité de certaines variables sensibles est assurée par des méthodes de permutation et d'imputation appliquées aux mailles isolées contenant un faible nombre de ménages.

Février 2013 : Suspension de la diffusion de toutes les données carroyées à 200m suite à la publication d'un article du *Canard Enchaîné* qui mettait en doute la confidentialité des données diffusées, mais sans prendre en considération les méthodes de floutage. L'Insee décide de modifier la chaîne de production et d'abandonner la diffusion de données floutées sur la maille de 200m car des dysfonctionnement techniques ont entravé l'anonymisation d'une faible proportion de carreaux.

Novembre 2013 : Diffusion de 18 nouvelles variables socio-économiques issues du RFL 2010 :

- la population résidente (total par carreau/rectangle, classe d'âge) : 10 var.
- Les ménages (total, taille, présents depuis + de 5 ans, propriétaire, en logement collectif, surface des résidences principales) : 6 var.
- les revenus (somme des revenus par UC, ménages en dessous du seuil de bas revenu) : 2 var.

La méthode de diffusion et de floutage des données a changé après une validation de la nouvelle méthodologie par le CNIS. A l'exception de la population résidente qui est diffusée sans floutage sur les carreaux, toutes les variables sont diffusées en zones sensibles sur des rectangles de taille variable, agrégeant suffisamment de ménages pour permettre de la diffusion des données. Des traitements statistiques de type « winsorisation » ont aussi été appliqués sur certaines variables sensibles.

Source : J. baro, 2015, p. 61

Le rapport du groupe Cuillier donne d'autres indicateurs à produire sur des carroyages INSPIRE. Pour les déplacements domicile-travail il est exprimé l'idée de disposer de

distances moyennes, médianes parcourues par mode de transport (Cuillier, 2009, p.63). Et également des temps moyens et médians de déplacements domicile-travail. Ces informations, sur un support de petite taille, permettent de s'affranchir des biais d'agrégation des localisations géographiques (impactant la mesure des distances).

L'Insee dispose de plusieurs sources de données pour effectuer ce travail. Les fichiers des Déclarations Annuelles des Données Sociales (DADS) est une formalité administrative obligatoire que doit accomplir toute entreprise employant des salariés. Ces données renseignent sur les adresses postales des salariés et de leur lieu d'emploi. Les services de l'Insee pourrait diffuser les origines-destinations domicile-travail en agrégeant les couples de localisations sur les mêmes carroyages que ceux des données carroyées de population. Dans ce cas, il faudrait mettre au point des méthodes de brouillage de l'information des localisations des salariés pour respecter les règles de l'anonymat. Au demeurant, les fichiers des DADS apportent des localisations domicile-travail des salariés mais les variables du recensement sont plus nombreuses et décrivent mieux les caractéristiques des ménages. La diffusion des données de mobilité du recensement sur un support carroyé peut aussi être envisagée en se servant des données confidentielles internes à l'Insee (nous allons voir que certaines informations sont accessibles pour la recherche à l'aide d'un centre d'accès sécurisé aux données (CASD)). Pour autant la question de l'anonymat demeure et des méthodes de brouillages devront donc être mises au point.

Les experts du CNIS proposent également de disposer d'autres indicateurs sur l'accessibilité, celles liées aux équipements. Les indicateurs proposés par le groupe de travail sont les distances et les temps moyens d'accès aux 4 gammes d'équipements définies par l'Insee (de base, de proximité, intermédiaire et supérieure). Comme nous le verrons, les chiffres de présence d'équipements disponibles sur le site de l'Insee peuvent être complétés par le répertoire des entreprises (Sirene). Le répertoire Sirene enregistre l'état civil de toutes les entreprises et leurs établissements.

D'autres indicateurs vont également pouvoir être incorporés sur ce carroyage. Nous en avons énuméré un certain nombre en introduction générale en présentant quelques-uns des 32 thèmes de la directive INSPIRE. Le travail de thèse de Johanna Baro consistant à renseigner une information sur la structure urbaine pour chaque carreau est un exemple supplémentaire (Baro, Bonin, Hubert, 2016).

Nous rejoignons ces recherches en proposant de désagréger les données de mobilité du recensement sur le carroyage de 1km initialement diffusé par l'Insee. Nous appelons ce support : carroyage INSPIRE de 1km. La diffusion de données à l'aide de ce support permet d'effectuer des découpages à façon de l'espace géographique, et ainsi de répondre aux besoins divers des analyses. Néanmoins, les supports ont encore une autre fonction à remplir, ils décrivent les localisations.

2.1.3 Les référentiels de localisation

Nous nous sommes mis dans la peau d'un utilisateur de données et nous avons mis en lumière le besoin de disposer de données de mobilité sur des supports répondant aux besoins d'analyse. Un utilisateur peut seulement agréger les données sur des supports alignés à celui de diffusion et donc agrandir les espaces d'analyse. Ceci ne permet pas de répondre aux objectifs du Grenelle de l'Environnement. Les utilisateurs attendent une réponse de la part des producteurs de données. La construction de données zonales représentatives de leur espace nécessitant de connaître à l'avance les critères d'analyse des utilisateurs. Auparavant les besoins exprimés par les collectivités territoriales pouvaient être satisfaits avec des supports alignés à leurs limites administratives. Pour des besoins différents, la solution proposée est un support régulier avec un certain nombre d'indicateurs pouvant être agrégés. Pour autant, le problème de l'agrégation demeure. Nous présentons ici le cas de l'agrégation des variables de localisations. Un support assurant exclusivement la fonction de localisation et de mesure des distances est utilisé. Ce référentiel de localisation est plus fin que le support de base de tirage. Puis ces mesures sont diffusées par l'intermédiaire d'un support plus grossier, la base de tirage diffusable.

Le référentiel de localisation est le dispositif géographique qui permet de localiser les observations et en particulier les origines et destinations des déplacements. Les référentiels peuvent être de plusieurs ordres. La localisation la plus fine est apportée par un support spatial ponctuel repéré dans l'espace par une adresse postale, des coordonnées GPS ou issues d'un positionnement par antennes GSM. Elle peut être obtenue également par un zonage fin comme le code postal au Royaume Uni ou en Amérique du Nord, ou le carroyage utilisé par l'Enquête Globale Transport d'Île-de-France (EGT). En effet, contrairement aux enquêtes Ménages-Déplacements, l'EGT utilise un support de recueil des origines-destinations très fin. Les origines et destinations des déplacements sont récoltées sur un support carroyé de 100m

recouvrant toute l'Île-de-France⁴⁸. Les observations sont localisées sur ce support et la représentativité statistique est assumée à l'échelle du support de tirage. Sur la carte 15 nous avons pu représenter les distances d'accès à vol d'oiseau aux stations de transport en commun des déplacements domicile-travail. L'EGT a fourni les points de passage des trajets composant les déplacements. A l'aide de ces informations, le praticien peut, par exemple, effectuer des analyses sur les aires de rabattement à une échelle fine.

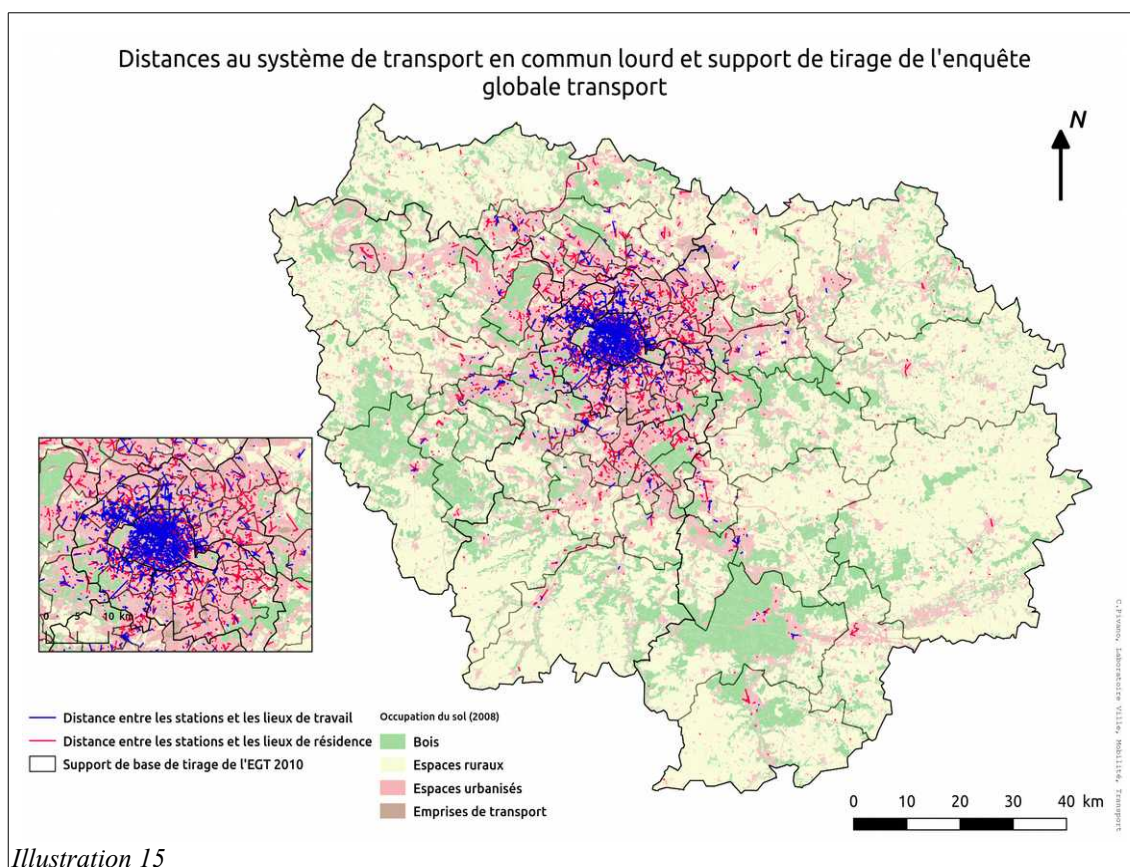


Illustration 15

Mais de manière générale l'information sur les distances est difficile à collecter, l'EGT est un cas particulier. De plus, la précision des réponses des enquêtés varie selon les modes de transport. Les voitures ont un compteur mais pour les transports en commun et les autres modes les distances ne peuvent pas être demandées aux enquêtés. Lorsque nous ne disposons pas des distances parcourues ou si le référentiel de localisation est grossier nous devons effectuer des mesures de distances après coup. Il faut alors avoir une méthode pour

48 Les enquêtes Globale Transport antérieures à 2010 ont un référentiel de localisation moins fin, de l'ordre de 300m.

calculer ces distances a posteriori (Gascon, 2009). Le Certu propose de calculer les distances à posteriori à partir des centres des zones du référentiel de localisation⁴⁹ mais sans connaître les barycentres des origines ou des destinations. Ce qui pose des biais d'agrégation. « *En pratique, la longueur d'un trajet sera supposée égale à la distance entre les zones fines origine et destination du trajet en considérant que les extrémités de ce déplacement sont situées en un point unique par zone : leur centroïde* » (Gascon, 2009,p.12)

La documentation ajoute que l'estimation des distances réellement parcourues peut être réalisée selon plusieurs modes opératoires et donnent des résultats légèrement différents :

- « *si les distances sont estimées à partir des longueurs à vol d'oiseau [...] le passage aux distances parcourues se fait à l'aide de coefficients correcteurs forfaitaires voisins de 1,3* »

- « *si les distances sont estimées à partir de l'utilisation d'un Système d'Informations Géographiques (comprenant des recherches d'itinéraires) et d'une base de données routières, les résultats obtenus sont plus proches des distances réellement parcourues car la méthode intègre directement les réseaux de voirie* »⁵⁰

-« *dans le cas où un modèle de prévision de trafic est utilisé, les résultats sont les plus proches de la réalité car les réseaux de voirie et le comportement des usagers sont pris en compte.* » En effet, le modèle prend en compte la congestion contrairement à l'utilisation des distances dites à vides par les SIG⁵¹.

Les opérations de mesure des distances sont cruciales pour répondre aux enjeux climatiques. A l'aide de la longueur des déplacements parcourus nous pouvons estimer la consommation d'énergie et les émissions de polluants. « *Les premiers travaux de calcul des émissions de CO₂ à partir d'une EMD ont été réalisés par l'INRETS dans les années 1990 à travers l'élaboration de BEED (Budgets Énergie Environnement Déplacements) sur plusieurs agglomérations françaises (Gallez et Hivert, 1998)* », (Bouzouina, Quetlard et Toilier, 2013, p.10). Pour le trafic automobile, il est appliqué un coefficient d'émissions de GES en fonction de certains critères : vitesses, cylindrée des véhicules, la norme et le type de carburant, démarrage à froid, etc. Les coefficients d'émissions sont issus de programmes de

49 Dans la méthodologie standard « Certu » le référentiel de localisation est appelé zones fines.

50 Les indicateurs en lien avec la distance proposée par le rapport du groupe Francis Cuillier est le distancier INRA Odomatrix. C'est un distancier de chef-lieu de commune à chef-lieux de commune.

51 Notons qu'une mesure de distance à l'aide d'un SIG en ayant au préalable calibré des vitesses sur le graph routier revient à faire de la modélisation de trafic.

recherche européens nommé COPERT (Ntziachristos et Samaras, 2000). Les coefficients sont appliqués à la longueur des déplacements. La mesure des distances peut être effectuée de plusieurs manières. Gallez et Hivert ont d'abord mesurés les distances rectilinéaires aux centroïdes des zones du référentiel de localisations des EMD. Plus récemment, Hivert, Nguyen, Merle et al (2014) mesurent la distance sur le réseau de transport à l'aide d'un système d'information géographique. Ces méthodes de mesures de distance a posteriori posent cependant un problème de précision car la localisation fine des origines-destinations n'est pas connue. Cette localisation est un identifiant de zone, qui est approximée par un point central pour le calcul des distances (le point central est au choix : le centre géométrique, le chef-lieux (mairie), etc). Or ce point peut introduire un biais s'il n'est pas confondu avec le barycentre des observations (illustration 16).

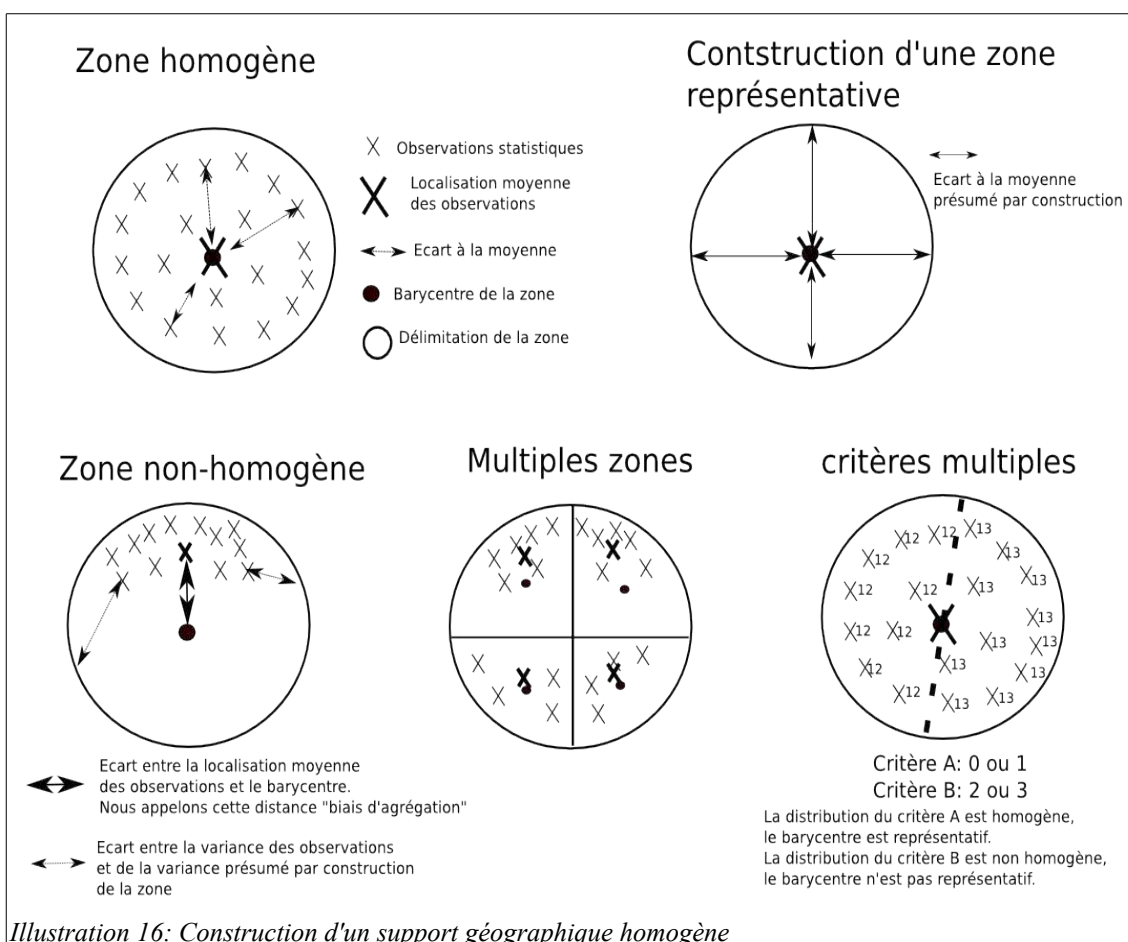


Illustration 16: Construction d'un support géographique homogène

Les mesures des distances sont dans tous les cas dépendantes de la finesse du support de localisation. Et plus le référentiel de localisation est fin moins le biais d'agrégation est grand. Si l'enquête EGT a un support de 100m, c'est rarement le cas des autres EMD.

En dépit de la finesse du référentiel de localisation il faut se méfier de la ré-agrégation des indicateurs. En effet, il faut par la suite agréger sur des supports homogènes selon le critère d'accessibilité pour offrir une donnée représentative. Il faut diffuser les mesures de distances sur des territoires ayant les mêmes accessibilités (homogène en termes de distances et/ou de temps de déplacements). En reprenant l'illustration 15 des aires de rabattement, nous remarquons que si nous construisons un indicateur de distance moyenne d'accès aux stations de transport par secteur de tirage la moyenne des distances peut ne pas être représentative des secteurs de tirage. Les moyennes peuvent ne pas être représentatives car le taux d'échantillonnage est trop faible et l'agrégation rassemble des observations au sein de territoires dont les comportements sont en réalité hétérogènes.

Nous en avons terminé avec la présentation des différentes fonctions que doivent assurer les supports de l'information géographique. Nous remarquons que les trois fonctions à remplir ne convergent pas toujours vers le choix d'un même support.

-les supports d'analyses doivent répondre à des questions diverses. Dans une démarche descriptive, ils peuvent seulement être construits par agrégation. Pour faire des analyses fines il faut recourir à de la modélisation, selon l'exemple des modèles de trafic.

-le support de base (production-diffusion) doit permettre de produire des échantillons représentatifs au sein des zones (le premier critère est généralement la localisation au lieu de résidence).

-le référentiel de localisations doit être le plus précis possible pour mesurer avec précision les distances.

Nous allons maintenant montrer comment ces fonctions sont remplies dans le cas d'enquêtes à faibles effectifs et dans celui du recensement de la population à fort taux d'échantillonnage.

2.2 Les supports avec des données faiblement échantillonnées : le cas des Enquêtes Ménages-Déplacements

Cette section analyse sous l'angle des supports spatiaux la méthode de construction d'enquête de déplacement préconisée par le Certu. Ces enquêtes constituent l'ossature de l'observation de la mobilité en France. Leur solidité statistique permet de calibrer les modèles de trafic et de mesurer l'évolution des parts modale à l'échelle d'un périmètre de transport urbain. La question que nous posons ici est de savoir si les choix des supports spatiaux utilisées ne contraignent pas le potentiel des analyses pouvant être effectuées à l'aide de ces données. Nous nous interrogeons aussi sur le moyen de se libérer de certaines de ces contraintes.

2.2.1 L'alignement des supports utilisés sur la géographie des IRIS de l'Insee

Nous partons de l'existant et partons de la méthodologie officielle de construction des données des enquêtes Ménages-Déplacements pour analyser le rôle des supports spatiaux utilisés dans les enquêtes. Nous reprenons les trois fonctions des supports dans un ordre différent : la fonction de production, de localisation et d'analyse.

Les commanditaires des Enquêtes Ménages-Déplacements (EMD) sont les collectivités territoriales en tant qu'autorités organisatrices des transports. Ces données servent à faire soit des portraits statistiques, soit à modéliser l'écoulement du trafic sous différentes hypothèses d'aménagement ou d'offre de transport collectif. La collecte s'appuie sur une méthodologie rigoureuse mise en place par le Certu. Reprenons une à une les étapes du guide méthodologique du Certu (Certu, 2008).

Les supports de base de tirage :

Le périmètre d'étude des enquêtes Ménages-Déplacements doit être soit celui du SCOTT ou, pour les plus grosses aires urbaines celui du Plan de Déplacements Urbains. Par construction, les limites des aires d'étude des EMDs s'appuient sur les limites communales.

Les supports de tirage servent à appliquer un plan de sondage stratifié (Loonis, 2009). Il faut

découper le périmètre d'étude en zones géographiques dans lesquelles seront tirés au sort des échantillons suffisamment importants pour refléter la composition de ces secteurs.

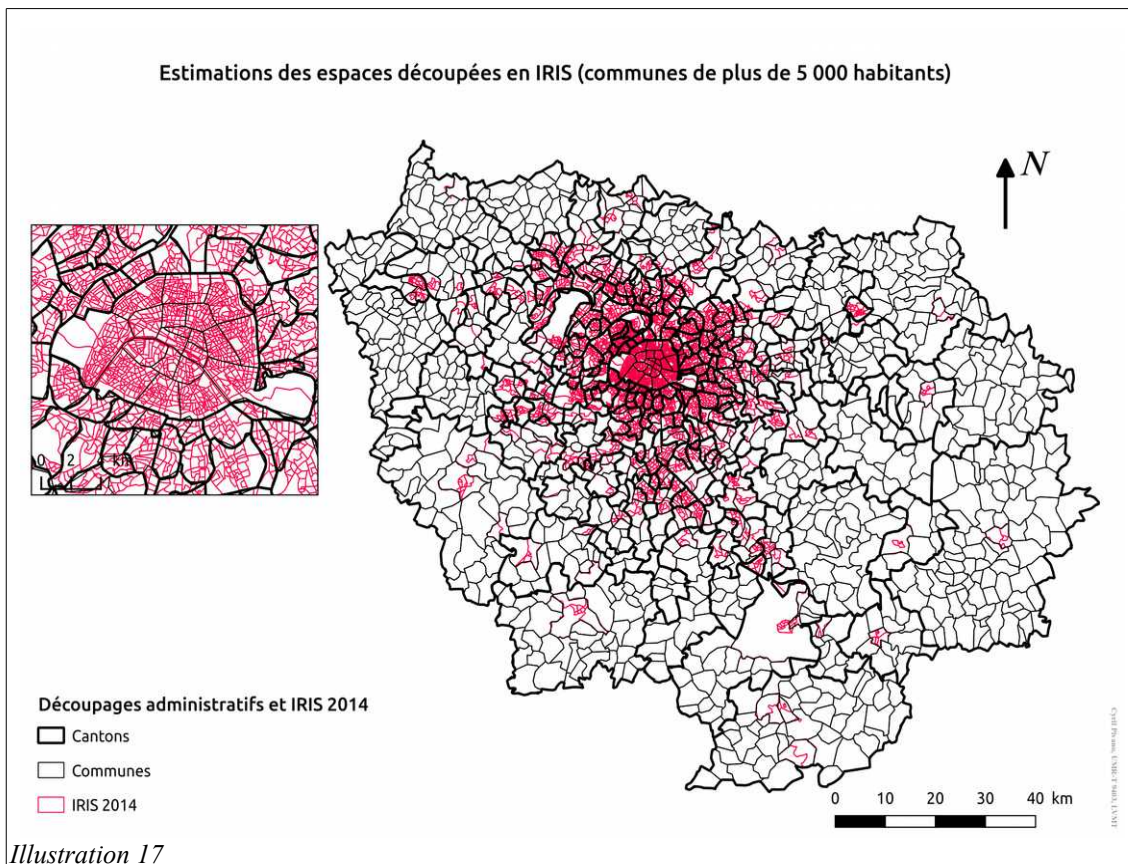
Le Certu préconise de définir une vingtaine de secteurs de tirage avec au moins 70 ménages enquêtés et validés par zone. Les ménages d'un secteur doivent comporter au moins 160 personnes. Au total, une EMD doit rassembler au moins 1 500 ménages enquêtés. La documentation Certu précise que les taux d'échantillonnage étant faibles, c'est la taille de l'échantillon qui détermine la précision des indicateurs. Une taille de 160 personnes par zone constitue selon le Certu un bon compromis entre la représentativité sociale par zone et la précision requise. Cependant, nous savons que les études de mobilité ont aussi besoin de représentativité spatiale (les observations doivent recouvrir l'ensemble de l'espace des zones), l'effectif de répondants doit être mis en relation avec les superficies des zones. Mais les règles pour effectuer le découpage du périmètre d'étude en secteurs de tirage n'abordent pas directement le critère de superficie. Les critères sont :

- le respect des contours IRIS
- la préservation des zonages pré-existants
- l'homogénéité des secteurs (« Chaque secteur de tirage d'échantillon doit être aussi homogène que possible en termes de structure urbaine et sociale » (Certu,2008,p.23.)

La combinaison de ces trois règles donne dans la pratique des supports de tirage construits à partir d'IRIS ou d'agrégats d'IRIS. Rappelons que le zonage en IRIS est construit par l'Insee afin d'étudier la population dans les zones urbaines denses avec un zonage fin mais suffisamment peuplé pour garantir l'anonymat des répondants au recensement (au moins 2 000 individus par IRIS). Les IRIS remplacent le découpage en îlots déficient pour la garantie de l'anonymat. En se servant des IRIS pour construire le support de tirage, on est assuré de disposer des indicateurs socio-démographiques au lieu de résidence nécessaires pour le redressement de la non-réponse des EMD. Mais le support de tirage devient aussi la base du support d'analyse, sans pour autant que ce soit obligatoire et en faisant attention aux évolutions temporelles du découpage en IRIS (chapitre 3.1.2)

En 2008, le Certu ne spécifie aucune règle pour les communes non découpées en IRIS (illustration 17), car moins peuplées, ce qui signifie que la diversité géographique des communes non découpées en IRIS ne peut pas être prise en compte dans le plan de sondage.

Dans ce cas le choix du support de tirage doit s'appuyer sur les limites communales (illustration 17).



Le référentiel de localisation :

Les secteurs de tirage sont donc souvent trop grossiers pour reconstituer de manière fine la structure du territoire et l'offre de transport actuelle ou projetée. C'est pour disposer de localisations plus précises sur les origines et les destinations des déplacements que les secteurs de tirage peuvent être sous-découpés en zones fines. Le Certu appelle ce support «Découpage fin», qui correspond aussi à ce que nous avons appelé «référentiel de localisation».

Là encore, il n'existe pas de règles clairement définies pour définir le référentiel de localisation. Le guide du Certu aborde tout de même plusieurs points. Il doit tenir compte de :

- l'identification des zones génératrices de déplacements

- l'identification des principaux pôles d'échanges
- l'identification des plans des politiques de la ville,
- et être adapté à la modélisation du trafic.

Le référentiel de localisation doit permettre d'identifier les «principaux lieux de génération de déplacements ». Ces lieux sont parfois concentrés dans l'espace (Chapitre 1.1.1.2) et donc situés dans un même secteur de tirage. Ils sont plus rarement résidentiels. « *Ces générateurs peuvent être des centres commerciaux, des hôpitaux, des domaines universitaires, des zones industrielles, des équipements administratifs ou culturels, etc.* » (Certu, 2008, p.23) Certains générateurs sont donc plutôt des points, d'autres des zones.

Ce découpage de l'espace peut aussi être construit pour répondre à une question spécifique d'aménagement local. Par exemple, dans le cas de la création d'une ligne de tramway le référentiel de localisation des déplacements peut découper plus finement les zones concernées par le projet d'infrastructure.

Le Certu invite dans ces cas à procéder à un sur-échantillonnage au lieu de résidence pour espérer capter plus de déplacements dans ces zones d'intérêts (Certu,2008,p.25)

Enfin, le référentiel des localisations des déplacements doit être adapté aux travaux de modélisation éventuellement prévus.

Le Certu nous informe qu'en cas de doute, il est préférable d'être plus fin que pas assez. Par contre, compte tenu des effectifs enquêtés souvent très faibles dans une zone de tirage, il est évident que des exploitations sur le référentiel de localisation plus fin ne pourront être réalisées qu'avec précautions (Certu, 2008, p.25).

L'idéal est donc de procéder avec un support de tirage fin (pas forcément confondu avec le référentiel de localisation) et avec un grand nombre d'observations pour observer finement les lieux à mettre en lumière selon les projets urbains des collectivités. Mais cela a un prix que seules les collectivités les plus riches peuvent payer. L'ordre de grandeur de coût d'une EMD est de 160 € à 180 € HT par ménage suivant la taille de l'échantillon (Certu, note de synthèse, p.4⁵²). Sachant qu'un minimum de 1 500 ménages doit être enquêté, le coût minimum d'une EMD est de 240 000 € HT.

52 Certu - note de synthèse 02 du CERTU « Mobilité et transports » - Février 2008 – [en ligne]
(consulté le 06/06/2016)
<http://www.bv.transports.gouv.qc.ca/mono/0968315/02_Note_de_synthese_02.pdf>

Le choix d'un référentiel de localisation peu précis car aligné sur des secteurs de tirage relativement grossiers a des conséquences sur le calcul de la distance, comme il a été vu plus haut.

L'EGT se distingue par son support de recueil des localisations. C'est en effet la seule enquête de la famille EMD à utiliser un carroyage fin et régulier pour repérer les origines-destinations dans l'espace du périmètre d'étude. Ce carroyage précis permet de calculer une distance réseau proche de la réalité, quel que soit le mode. Toutefois les secteurs de tirage et la taille d'échantillon demeurent faibles pour avoir une représentativité spatiale fine en périphérie peu dense.

Les zonages d'analyse

Les commanditaires des enquêtes doivent respecter les conditions de constructions d'échantillonnage pour recevoir une aide financière de l'État (de 20%). Certains types d'analyses doivent également être réalisées. L'exploitation «standard Certu» est un rassemblement des exploitations les plus souvent utilisées et demandées par les collectivités locales. L'exploitation standard doit traiter de l'ensemble du questionnaire standard et ne prend pas en compte les questions locales. Il convient de réaliser au moins des portraits statistiques, à défaut d'utiliser ces données dans une chaîne de modélisation. Les données des EMD doivent être réparties spatialement selon au moins trois supports dénommés : DTIR, D30 et D10. Le découpage DTIR est le découpage de base de tirage de l'échantillon. Les découpages D30 et D10 regroupent ces secteurs. *« Les appellations D30 et D10, qui sont des appellations génériques, n'imposent pas des regroupements en 30 et 10 secteurs. Il appartient à la maîtrise d'ouvrage de définir ces regroupements.*

- Les découpages D30 et D10 doivent être constitués de zones qui regroupent au moins deux secteurs de tirage d'échantillon (minimum 140 ménages et 320 personnes) ;
- Le découpage D10 doit être constitué d'au maximum 12 zones ;
- Le nombre de zones du D30 et du D10 résulte de la taille de l'échantillon et des deux principes ci-dessus. »

Le Certu invite donc à disposer de plusieurs supports d'analyses en agrégeant les données socio-économiques et/ou les données de localisations pour pouvoir analyser les données. Reprenons l'exemple des études menées par l'Observatoire de la Mobilité en Île-de-France

réalisées à partir de l'Enquête Globale Transport d'Île-de-France de 2010, (illustrations 13 et 14). Nous pouvons effectuer des analyses croisant les profils socio-économiques seulement à une certaine échelle d'agrégation. L'Omnil effectue des tris croisés seulement au secteur de tirage. Certes auparavant les analyses ne se faisaient qu'aux départements. « *Concernant l'échantillonnage, l'Île-de-France a été découpée en plus d'une centaine de **secteurs de tirage**. Les résultats de l'EGT 2010 sont donc représentatifs à l'échelle de ces secteurs de tirage alors que les résultats des précédentes enquêtes étaient représentatifs à l'échelle des départements.* » (Omnil, méthodologie EGT⁵³)

Le support de tirage est effectivement assez fin dans certains espaces. A Paris et pour les communes de moins de 10 000 habitants les secteurs de tirage sont construits à l'aide d'agrégats d'IRIS. Pour les espaces plus périphériques les zones de tirages sont bien plus étendues. Selon les règles en vigueur les zones doivent contenir au moins 160 personnes résidentes peu importe leur superficie. Étant certainement la région la plus riche de France, L'EGT Île-de-France de 2010 a enquêté 18 000 ménages (soit près de 43 000 personnes). En moyenne, nous disposons de 430 personnes enquêtes par zone.

Nous pouvons faire des analyses descriptives au niveau individuel, mais nous n'avons pas une représentativité spatiale, ni sociale à l'intérieur des zones de tirage d'autant plus qu'elles sont grandes. Et ce en raison du faible taux d'échantillon. Nous pouvons effectuer des analyses croisant les profils socio-économiques seulement à une certaine échelle d'agrégation. Et cela pose des problèmes quant à la finesse des analyses pouvant être effectuées.

53 Omnil. Méthodologie EGT [en ligne] (Consultée le 15/06/2016) <<http://www.omnil.fr/spip.php?article122>>

2.2.2 L'agrégation des données d'enquêtes à faible taux d'échantillonnage pour l'analyse

Certes, ce n'est pas l'échelle individuelle qui intéresse les politiques d'aménagement. Toutefois il est souhaité observer les territoires, aussi petits soient-ils. Ce qui amène à traiter de l'effectif d'échantillon au sein de petits territoires (aires géographiques restreintes). Plus l'échantillon est petit moins il est possible d'observer une significativité dans les tests statistiques. Il faut alors de grandes zones pour disposer de plus d'observations mais plus sujette à être non homogène et donc à ne pas révéler des relations statistiques. Suivant le niveau d'agrégation et donc de la taille d'échantillon il est plus ou moins possible d'étudier des relations statistiques. Des éléments de la littérature démontrent qu'avec les EMD il faut un niveau d'agrégation parfois assez élevés.

Le rapport d'étape de Raux (1996) est édifiant. Les auteurs analysent les évolutions des temps de déplacements des usagers des transports en commun de Lyon entre l'EMD de 1976 et celle de 1995. Bien qu'en 20 ans de considérable développements urbain ont été faits il est difficile d'en observer les effets sur les différentes zones du territoire lyonnais. Une grande critique est donc faite sur la taille des échantillons des EMD's de Lyon qui présente pourtant un effectif important comparativement à la plupart des agglomérations françaises. Bonnel présente cette critique.

« Les auteurs montrent que la précision sur les temps de déplacements déclarés par les enquêtés en transports collectifs entre zones pour un découpage en 5 zones de l'agglomération (sur le périmètre de l'enquête 1976) ne permet de mettre en évidence une évolution que pour 6 flux sur les 25 alors que la différence globale est assez importante (37 minutes en 1976 avec une précision de 2 %, et 32 minutes en moyenne en 1996 avec une précision de 2 %). Le gain de temps en transports collectifs est pourtant probablement général sur l'ensemble de l'agglomération, compte tenu de l'amélioration du système des transports collectifs au cours de ces 20 années. On peut ainsi observer une diminution quasi systématique des temps de parcours en transports collectifs pour chacune des origines-destinations, mais ces différences ne sont pas statistiquement significatives. Ce constat résulte d'intervalles de confiance trop larges pour chacune des enquêtes (supérieurs à 10 % pour 18 des 25 flux en 1976 et pour 14 des 25 flux en 1995). Il faut ainsi des écarts d'au

moins 25 % pour la moitié des flux pour que les résultats soient statistiquement significatifs au seuil de confiance de 5 %. Le découpage n'est pourtant pas particulièrement fin. On imagine aisément qu'avec un découpage légèrement plus fin, il n'y a plus aucun flux pour lequel le résultat puisse être significatif.

Pour des calculs de pourcentage, la précision est également bien souvent insuffisante. Ainsi lorsqu'il s'agit de calculer un pourcentage à partir d'un échantillon de 100 observations (Raux et al., 1996), le demi-intervalle de confiance est de 10 % pour un pourcentage de 50 %, soit un intervalle [40 % - 60 %]. Même avec un échantillon de 1.000 observations le demi-intervalle est encore de 3 %, toujours pour une proportion de 50 %. L'un des objectifs de l'enquête ménages déplacements est de pouvoir mesurer l'incidence d'une modification importante de l'offre de transports collectifs. Prenons l'exemple de la mesure de l'incidence de la création d'un axe lourd de transports collectifs en site propre. La part de marché des transports collectifs (sur l'ensemble voiture + transports collectifs) est fréquemment de l'ordre de 20 %. Si le nombre de déplacements enquêtés est de l'ordre de 1.000 unités, l'intervalle de confiance (avec une précision de 5 %) est [17,5 % - 22,5 %]. Même avec 10.000 déplacements (soit de l'ordre de 2.000 ménages enquêtés), l'intervalle de confiance est encore de [19,2 % - 20,8 %]. Les intervalles des deux enquêtes s'additionnant, il faut donc que la part de marché des transports collectifs augmente de 5 % (soit 25 %) si l'on a 1.000 déplacements, et de 1,5 % (soit 21,5 %) si l'on a 10.000 déplacements. Lorsque l'on connaît la modération de la croissance de la part de marché des transports collectifs (par exemple à Lyon, sur le périmètre 1976, la part de marché des transports collectifs (sur l'ensemble voiture + transports collectifs) est de 22 % en 1976, 24 % en 1985 et de 21 % en 1995), la probabilité est forte pour que les écarts ne puissent être au mieux que mesurés sur l'ensemble de l'agglomération. En revanche, toute segmentation qu'elle porte sur des découpages géographiques ou des caractéristiques socio-économiques conduira probablement à des résultats statistiquement non significatifs.

Un accroissement de la taille de l'échantillon permettrait de mieux répondre à certains objectifs assignés à l'enquête ménages déplacements. » (Bonnell, 2001, p.129)

Pour l'Île-de-France, rappelons que l'EGT de 2010 regroupe en moyenne près de 430 observations par secteur de tirage. Nous sommes loin des 2 000 ménages nécessaires pour étudier statistiquement une zone. L'utilisation des données d'enquêtes probabilistes avec de petits échantillons en tant que données descriptives doit se faire en agrégeant les observations

sur des grands territoires afin de disposer d'échantillon suffisant et d'une répartition homogène. Les EMD ne permettent alors d'étudier que les comportements à l'échelle de l'ensemble des zones de tirage. Il ne peut pas être fait d'analyses statistiques à l'intérieur ou même entre plusieurs zones. A l'échelle de la zone d'étude, les moyennes régionales sont représentatives et les relations statistiques peuvent éventuellement être significatives. Nous comprenons désormais pourquoi le Certu invite à agréger les observations.

Les EMD apportent néanmoins des informations très riches sur la description des comportements de mobilité, elles décrivent l'ensemble des motifs de déplacements avec des informations sur la temporalité des déplacements (heures de départs et d'arrivées). La chaîne de modélisation à quatre étapes permet d'enrichir les EMD (Certu, 2003). Les EMD ayant un faible taux d'échantillonnage les méthodes de modélisation de trafic extrapolent les échantillons à l'aide des données du recensement et des données de comptage des gestionnaires des infrastructures de transport. L'avantage de ces méthodes est de pouvoir traiter de plusieurs motifs de mobilité et d'en donner une photographie d'un jour moyen d'enquête. Nous nous inspirons de cette méthodologie pour répondre au besoin en données exprimé par le changement de paradigme des politiques de transport. Comme la thèse se concentre sur les données de mobilité domicile-travail nous pouvons alors ne pas utiliser les données EMD. Les données du recensement sont beaucoup moins détaillées mais ont un plus grand nombre d'observations qui leur donne une potentialité pour contrôler l'enrichissement d'autres jeux de données, ce qui n'empêche pas d'être enrichi par ailleurs. Au premier chapitre nous avons posé l'idée de disposer d'une base de couplage représentative statistiquement de la population sur un support permettant le couplage de données issues de différentes sources. La diffusion de données sur un support régulier avec un référentiel géographique commun est la voie indiquée par la directive INSPIRE. Nous suivons cette voie en proposant de produire et diffuser les données de mobilité du recensement sur le support INSPIRE de 1km. Ce chapitre termine donc par présenter les méthodes de collecte du recensement rénové de population. Nous orientons la description du recensement en mettant en avant le concept de poids de sondage des enquêtes. Ce sont des outils renseignant sur le niveau de représentativité statistique : la relation entre tailles des échantillons par zone de tirage et résolutions des supports irréguliers.

2.3 Les supports spatiaux et l'échantillonnage du recensement rénové de la population

Le recensement rénové est une enquête utilisant plusieurs méthodes de sondage. Les méthodes de collecte sont différentes pour les communes de moins de 10 000 habitants et pour les communes de 10 000 habitants et plus. Aussi les résultats du recensement sont diffusés à l'aide de plusieurs fichiers renseignant sur les localisations avec des supports différents. Nous détaillons dans cette section comment sont construits les données du recensement rénové et accordons à nouveau une place privilégiée aux supports spatiaux intervenant dans la production. Nous introduisons ainsi un outil permettant d'évaluer du niveau de représentativité des supports d'enquêtes : les poids de sondage. Et, ils vont nous être d'une grande aide dans notre proposition méthodologique de désagrégation.

Le recensement de la population n'est plus une enquête exhaustive. Les résultats sont produits à partir d'un sondage représentatif de la population entière, c'est-à-dire qui permettent d'analyser les caractéristiques de l'ensemble des individus à partir d'une fraction conséquente de la population.

Dans le cas d'un sondage parfaitement équilibré dans une population homogène, l'échantillon présente un rapport d'homothétie avec la population entière : on enquête x % des individus, et on extrapole les analyses faites à partir de l'échantillon en multipliant les résultats obtenus par x . Dans le cas d'une enquête par sondage quelconque, on reconstitue la représentativité statistique de l'échantillon en munissant les individus statistiques de poids. Le poids de sondage compense la probabilité de tirer un individu dans sa strate de sondage. Des opérations de redressement de la non-réponse interviennent souvent pour modifier les poids de sondage en calant certaines variables d'intérêt à des effectifs indiqués par une autre source de données (Caron, 2005). Le recensement est d'ailleurs souvent utilisé dans les étapes de redressement. Le cas du recensement est un peu particulier puisque les poids de sondage sont modifiés afin d'harmoniser les différents types d'exploitation et les différentes méthodes du recensement collectant les observations à différentes dates.

Les données du recensement rénové sont diffusées à l'aide de plusieurs fichiers détail. Certains sont issus de ce que l'Insee appelle l'exploitation principale et d'autres de l'exploitation complémentaire (tableau 3). Nous nous concentrons dans cette thèse sur les

fichiers MOBPRO et INDCANTVILLE de l'exploitation complémentaire. Précisons aussi que ces fichiers sont issus d'un seul et même fichier que nous appelons fichier confidentiel.

Nom des fichiers	Localisation(s)	Type d'exploitation
Fichier Logement ordinaires	IRIS-Communes de résidence	Exploitation principale du recensement ⁵⁴ : 40 % dans les communes de 10.000 hab et plus, exhaustif ailleurs
Fichier Individus localisé au canton-ville (INDCANTVILLE)	IRIS-Cantons de résidence	Exploitation principale du recensement
Fichier Individus localisés à la région	Région ou département et EPCI pour les départements de plus de 700 000 habitants	Exploitation principale du recensement
Fichier Mobilité scolaire des individus	Commune de résidence et commune de scolarisation	Exploitation complémentaire du recensement : 40 % dans les communes de 10,000 hab et plus, 25 % ailleurs
Fichier des Mobilités professionnelles (MOBPRO)	Commune de résidence et communes d'emploi	Exploitation complémentaire du recensement
Fichier des activités professionnelle des individus	Zone d'emploi de travail	Exploitation complémentaire du recensement
Fichier des migrations résidentielle (MIGCOMM)	Plus disponible	Exploitation complémentaire du recensement
Source : Insee		

La différence entre l'exploitation principale et l'exploitation complémentaire est que cette dernière contient des variables socio-économiques plus fines permettant de construire des profils socio-économiques plus précis. Ainsi, les professions ou catégories sociales (PCS) ainsi que les secteurs d'activité sont mieux décrits grâce à un travail supplémentaire sur les données collectées.

Ces exploitations ne sont pas réalisées sur l'ensemble des bulletins du recensement collectés auprès des ménages. Sont enlevés 75 % des bulletins du recensement des communes de moins de 10 000 habitant pour réduire les coûts et diffuser le plus rapidement les résultats⁵⁵. Pour combler ce vide, les poids de sondage sont augmentés. Quant aux communes de plus de 10 000 habitants, le nombre de bulletins collectés est identique à celui de l'exploitation principale, donc les poids de sondage ne changent pas. Nous expliquons dans un premier temps la construction des poids de sondage des petites communes puis celle des grandes communes.

54 Insee, Les exploitations principales et complémentaires [en ligne] (Consultée le 16/06/2016) <http://www.insee.fr/fr/methodes/sources/pdf/RP_exploitations_principales_et_complementaires.pdf>

55 ce qui est pas surprenant comme argument lorsque l'on sait qu'il s'agit de millésimes d'enquêtes menées à des années de collectes différentes.

Ensuite, toujours en se rapportant aux supports spatiaux, nous présentons les méthodes d'estimation de la précision du recensement. Le tirage au sort des échantillons totaux des ménages s'appuyant sur les adresses alors les poids de sondage sont relatifs aux effectifs totaux de ménages habitant les adresses enquêtées et non pas aux modalités socio-économiques.

2.3.1 Échantillonnage des communes de moins de 10 000 habitants

Un cinquième des communes de moins de 10 000 habitants sont enquêtées exhaustivement chaque année (Insee. Fiche pondération⁵⁶). Et comme le fichier MOBPRO rassemble cinq enquêtes annuelles nous disposons à terme de l'ensemble des ménages des petites communes enquêtées⁵⁷. Nous pouvons dire que le recensement rénové de la population est exhaustif sur les petites communes. Il n'y a pas de poids de sondage ou autrement dit les poids valent 1.

Le taux d'échantillonnage de l'exploitation complémentaire est plus faible. Cette dernière retient seulement 25 % des observations enquêtées des petites communes. Les poids de sondage des observations décrites dans les fichiers de l'exploitation complémentaires sont égaux à 4. Ainsi, les résultats obtenus en totalisant les poids des unités statistiques reflètent l'ensemble des logements, ménages, familles ou individus concernés.

Aussi, pour harmoniser temporellement les observations, les poids de sondage sont utilisés pour ramener les effectifs du recensement au 1^{er} janvier de l'année médiane du millésime. Cette étape de redressement permet d'obtenir des poids finaux variant autour de 4 dans l'exploitation complémentaire⁵⁸ et de 1 dans l'exploitation principale.

56 Insee. Fiche pondération [en ligne] (Consultée le 05/05/2016) < <http://www.insee.fr/fr/ppp/bases-de-donnees/recensement/resultats/doc/pdf/fiche-ponderation.pdf> >

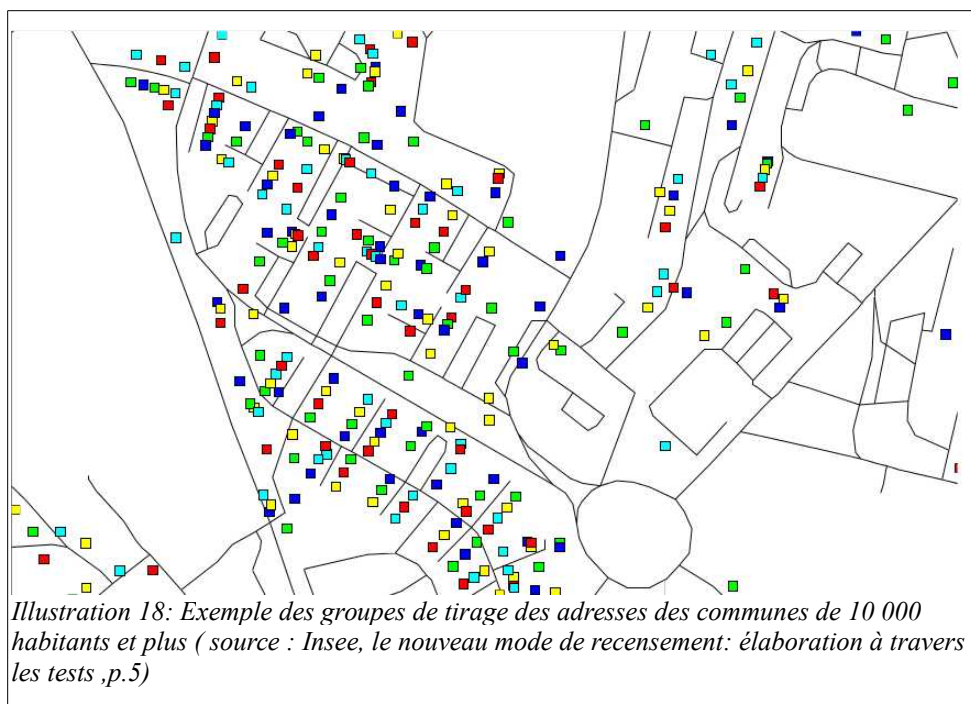
57 Avec le principe de rotation, certains ménages peuvent par le jeu des déménagements être recensés deux fois (ou plus) ou pas du tout. Il est supposé que ces effets se compensent.

58 Les poids ne sont pas des entiers.

2.3.2 Échantillonnage des communes de plus de 10 000 habitants

Pour les communes de plus de 10 000 habitants la méthode de sondage est plus sophistiquée. Il est espéré enquêter 40 % de la population des ménages en cinq ans, mais ce chiffre est un peu incertain du fait que la base de sondage utilisée est un répertoire d'adresses des logements.

L'Insee se sert d'une base de sondage sur les logements (le RIL – répertoire d'immeubles localisés) pour tirer au sort 8 % d'échantillons d'adresses chaque année. Pour chaque commune, la base de sondage est divisée en strates géographiques, et en strates non-géographiques à l'intérieur de chaque strate spatiale. Les strates géographiques correspondent à des secteurs de tirage, et les strates non géographiques à des typologies de nombres de logement par adresse (illustration 18⁵⁹). Nous reprenons longuement la documentation de l'Insee afin d'expliquer l'utilisation des différents supports géographiques et comprendre les méthodes de construction des poids de sondage des observations du recensement rénové.



59 Insee, le nouveau mode de recensement: élaboration à travers les tests [en ligne] (Consultée le 14/06/2016) <<http://www.insee.fr/fr/ppp/sommaire/imeths01e.pdf>>

« Ce schéma représente les voies d'un quartier quelconque d'une commune de grande taille. Les petits carrés représentent les adresses d'habitation le long des voies. Chaque adresse a été affectée à un groupe identifié par sa couleur. Il y a cinq groupes au total dans la commune et tous ont le même nombre de logements. Ces groupes sont stables dans le temps et sont mis à jour uniquement en fonction des créations et des destructions d'adresses. Une adresse qui a été placée dans un groupe y demeure, du moins jusqu'à sa disparition. On décide, par exemple, que l'échantillon des adresses recensées en 2005 sera tiré dans le groupe « rouge » et celui des adresses recensées en 2006 dans le groupe « jaune ». Dans ces conditions, le groupe d'adresses "rouge" fera l'objet d'un nouveau tirage d'échantillon en 2010, puis 2015, puis 2020, etc., tandis que le groupe "jaune" sera concerné en 2011, puis 2016, puis 2021, etc. » (Insee, le nouveau mode de recensement: élaboration à travers les tests, p.5)

Les groupes de rotations sont construits afin qu'ils contiennent des adresses réparties de manière homogène dans l'espace. Aussi ils sont construits pour regrouper des adresses de tailles différentes afin d'éviter le phénomène de grappe. Nous présentons successivement les deux strates de tirage.

-la strate des adresses :

Les unités d'échantillonnage sont des adresses d'habitation et non des logements. Dans le recensement, l'Insee considère qu'un ménage est l'ensemble des individus résidant dans un même logement. Échantillonner à partir des adresses pose problème, puisqu'une adresse est associée dans bien des cas à plusieurs logements. Le nombre de logements peut même être considérable pour les grands ensembles collectifs. Les résultats du tirage au sort vont être biaisés par ces grandes adresses qui auront une influence plus importante qu'une adresse avec peu de logements associés. L'Insee remarque que les profils socio-démographiques des logements collectifs sont souvent similaires ; une adresse composée de beaucoup de logements biaisera alors fortement les résultats. Ce problème est appelé « *effet de grappes* ». Le choix de considérer les adresses comme unités d'échantillonnage est donc non optimal. Il provient de la réponse élaborée pour répondre à un problème amont. Il a été choisi de demander aux enquêteurs d'enquêter toutes les habitations d'une même adresse, car il était plus difficile de déterminer précisément les logements au sein des adresses. L'Insee ne souhaite pas tirer au sort des logements dans des adresses qui seraient difficilement accessibles par les enquêteurs. Cela serait source d'erreur. Il n'existe pas de base de données

déterminant aux sein des immeubles les numéros des logements. Seul l'enquêteur une fois sur place peut dénombrer les logements. C'est donc pour répondre à ce problème de manque de données qu'il a été imposé de choisir les adresses comme unité d'échantillonnage.

- « *Les groupes de rotation sont constitués en trois strates caractérisées par le nombre de logements des adresses.*

– *les adresses de grande taille d'une commune sont celles qui, à la fois, sont les plus grandes en nombre de logements, ont au moins 60 logements chacune et qui, réunies, cumulent au maximum 10% des logements de la commune. Elles constituent la première catégorie. Elles sont réparties en cinq groupes pour être enquêtées exhaustivement en cinq ans ;*

– *les adresses nouvelles sont les adresses qui apparaissent année après année. Celles qui sont de grande taille rejoignent la catégorie précédente. Les autres constituent la deuxième catégorie ;*

- *les autres adresses d'habitation regroupent toutes les autres adresses où il existe au moins un logement. Elles constituent la troisième catégorie. Les adresses nouvelles qui forment la deuxième catégorie sont réparties entre les groupes de la troisième catégorie de manière à être toutes recensées en cinq ans. Pour les mêmes raisons qui ont conduit à équilibrer les cinq groupes de communes de moins de 10 000 habitants, mais cette fois en raisonnant au niveau de la commune et non plus au niveau de la région, chaque groupe d'adresses doit être aussi représentatif que possible de l'ensemble de sa catégorie dans la commune et chaque échantillon d'adresses doit être aussi représentatif que possible de l'ensemble des adresses de la commune. Les critères retenus sont les mêmes que pour l'équilibrage des groupes de communes de moins de 10 000 habitants : nombre de logements individuels, nombre de logements collectifs, population par sexe et âge (la référence étant le recensement de 1999 à l'initialisation du répertoire). » (Insee, le nouveau mode de recensement: élaboration à travers les tests, p.10)*

Afin d'éviter les effets de grappe concernant les grandes adresses il est choisi de toutes les enquêter en cinq ans. Ces adresses sont réparties en cinq groupes équilibrés (en nombre de logements) en cherchant à optimiser la répartition spatiale dans la commune ; chacun de ces groupes sera traité à part et recensé exhaustivement, son année venue. Dès lors sur une période de cinq ans toutes ces adresses auront été enquêtées et l'effet de grappe aura disparu.

La notion de grande adresse est relative : une adresse est grande si elle contient un nombre de logement dans le dernier décile de la distribution de la taille des adresses de la commune⁶⁰. Seuls les grandes adresses du dernier décile sont enquêtées exhaustivement. Ces observations ont alors avant la prise en compte des différentes années de collecte des poids de sondage de 1.

En ce qui concerne les autres types d'adresses, elles ne sont pas enquêtées de manière exhaustives. « *Chaque année, un échantillon aléatoire d'adresses est tiré dans le groupe de l'année. Pour ce tirage de deuxième phase, on introduit comme critères d'équilibrage le nombre de logements de la strate, le nombre de logements collectifs et le poids des IRIS en nombre de logements* » (Insee. Pour comprendre le recensement Annexe B2 : Le plan de sondage dans les communes de 10 000 habitants ou plus, p.3⁶¹). Nous comprenons que le tirage au sort vise à retenir des logements représentatifs spatialement de l'ensemble de la commune. Pour cela il est utilisé une strate géographique présentée par la suite. La documentation nous informe que les poids de sondage des observations autres que les grandes adresses sont en général compris entre 2,5 et 5 et le plus souvent près de 3.

Les poids sont ajustés par la suite pour estimer un effectif moyen sur les cinq années de collecte. Ils demeurent en moyenne proche de 3.

Enfin, contrairement aux communes de moins de 10 000 habitants l'exploitation complémentaire utilise tous les bulletins de recensement des grandes communes. Donc les poids de sondage demeurent identiques à celui de l'exploitation principale.

-La strate géographique : les IRIS et IRIS-2000

Les échantillons des adresses sont tirés au sort chaque année de manière à avoir à peu près le même nombre de logements par IRIS. Le support spatial IRIS est utilisé pour stratifier le tirage au sort et assurer sa représentativité dans l'espace des communes de plus de 10 000 habitants.

Les IRIS sont construits à l'aide d'un autre découpage, les IRIS-2000. Jusqu'en 1999, le recensement général de la population (RGP) utilisait déjà un découpage infra-communal de

60 Une commune avec beaucoup de grandes adresses aura toujours un effet de grappe. Sauf si bien sûr il n'y a que des grandes adresses.

61 Insee. Pour comprendre le recensement Annexe B2 : le plan de sondage dans les communes de 10 000 habitants ou plus [en ligne] (consulté le 06/05/2016)
<<http://www.insee.fr/fr/ppp/sommaire/imeths01g.pdf>>

collecte : l'îlot. L'ancienne méthode du recensement était un comptage exhaustif. Selon la loi sur le secret statistique, l'information du RGP ne pouvait être diffusée sur ce support géographique tellement fin que l'on peut y reconnaître les individus. Alors les résultats ont été agrégés sur un support nommé IRIS-2000 signifiant « Îlots Regroupés pour l'Information Statistique » et qui fait référence à la taille visée de 2 000 habitants par maille élémentaire.

Avec le recensement rénové, les IRIS-2000 ont été mobilisés indirectement comme strate de la base de sondage et non plus comme support de diffusion. En effet, les critères d'équilibrage des groupes de rotation des adresses sont issus du recensement de 1999 (nombre de logement individuel et collectif, population par sexe et âge).

Mais comme le territoire change, le recensement rénové diffuse les résultats des grandes communes sur un support IRIS actualisé régulièrement. Ce découpage est le support sur lequel les localisations résidentielles sont diffusées dans certains fichiers détail du recensement pour préserver l'anonymat. Indirectement, les IRIS sont issus des IRIS-2000 et de leur méthodologie, mais ils diffèrent. En effet, ils sont construits de manière à contenir 2 000 habitants et les effectifs dénombrés dépendent du tirage au sort des adresses au sein des groupes, qui sont construits à l'aide des IRIS-2000.

Afin de garantir que les IRIS soient homogènes en termes de population le découpage évolue dans le temps. Les IRIS sont mis à jour après chaque année de collecte. Lorsque un IRIS devient trop peu habité il est agrégé avec un autre IRIS. Lorsque celui-ci redevient de taille suffisante il est recréé. Comme le support IRIS évolue dans le temps pour être homogène en population il ne peut pas être utilisé comme strate de sondage. Des adresses pourraient changer de groupe de rotation d'année en année et être enquêtées plusieurs fois. De plus, nous remarquons que les évolutions du support IRIS peuvent rendre délicate les analyses temporelles des résultats du recensement. C'est pourquoi ce sont les IRIS-2000 qui sont utilisés pour définir les groupes de rotation des adresses. Et, dans un second temps, les adresses à enquêter une année donnée sont tirées au sort en assurant une répartition homogène en introduisant comme critère d'équilibrage le poids des IRIS en nombre de logement. Ces deux supports sont donc utilisés comme strate de sondage. Il n'y pas de problème de non-alignement des supports puisque l'Insee dispose des adresses des logements comme support le plus fin.

Maintenant que nous connaissons les méthodes de construction de poids de sondage nous

devons nous poser la question sur les hypothèses de construction qu'ils posent. Les poids de sondage sont construits pour estimer l'effectif de population à l'année médiane des cinq enquêtes annuelles. Le dénombrement des populations par type de variable socio-économique peut être alors source de biais. Nous nous intéressons donc brièvement à la précision du recensement.

2.3.3 La précision du recensement

Comme le recensement rénové de population est une enquête par sondage, sa précision est dépendante du processus d'échantillonnage que nous venons de décrire.

Le recensement général de la population permettait de connaître l'effectif total de population. Pour mesurer la précision du recensement général, on tire au sort des sous-échantillons afin de mesurer l'écart à l'effectif de la population totale. Cette opération de tirage au sort de sous-échantillons puis de mesure de l'écart à l'effectif total est répétée de très nombreuses fois. L'écart moyen à cet effectif total est ce que l'Insee appelle un coefficient de variation.

Pour le recensement de la population rénové, le même principe est adopté, à la différence que nous ne connaissons pas avec certitude l'effectif global. Les écarts sont donc comparés à la moyenne des autres écarts⁶² (dans le processus la moyenne est mise à jour à chaque tirage).

Enfin à l'aide des coefficients de variation un intervalle de confiance à 95 % peut être calculé, c'est-à-dire un intervalle dans lequel la vraie valeur (inconnue) aura 95 % de chance de se trouver. Cet intervalle de confiance caractérise ce que l'Insee nomme la précision.

$$[\text{valeur du RP } x (1 - 2CV) ; \text{valeur du RP } x (1 + 2CV)]^{63 \ 64}$$

Toutefois cette mesure est effectuée pour l'effectif global de la population. La précision doit être calculée pour les autres variables socio-économiques si nous voulons être assurés de la validité d'une exploitation socio-économique du recensement, comme par exemple le nombre d'actifs ou d'agriculteurs parmi la population totale. Il faut alors tenir compte de l'imprécision sur l'effectif de la modalité à étudier (le numérateur) et de celle sur l'effectif total de

62 Nous remarquons que le recensement général étant quasi-exhaustif l'effectif total était en fait un effectif moyen à l'instar des résultats du recensement rénové.

63 Insee. La précision des résultats du recensement [en ligne] (Consultée le 14/06/2016)

<<http://www.insee.fr/fr/ppp/bases-de-donnees/recensement/resultats/doc/pdf/fiche-precision.pdf>>

64 Remarquons que l'intervalle de confiance à 95 % utilise le quantile 0,975 de la loi normale, soit 1,96, dont 2 est une valeur approchée.

population estimée (le dénominateur). Chaque variable socio-économique de l'exploitation complémentaire possède ainsi sa propre précision. En effet, le tirage au sort des échantillons s'appuyant sur les adresses alors les poids de sondage sont relatifs aux effectifs totaux de ménages habitant les adresses enquêtées et non pas à une modalité socio-économique particulière. Ainsi la somme d'une variable socio-économique quelconque n'est pas connue, elle est supposée être égale à la somme des poids de sondage. Comme l'indique l'Insee, cette supposition est cohérente en cas d'homogénéité de la sous-population dans la population totale (Insee. La précision des résultats du recensement⁶⁵). Par exemple, si nous souhaitons connaître la proportion de cadres dans la population totale il faut tenir compte de l'imprécision de la population totale et du nombre de cadres. La méthode de calcul de l'imprécision consiste à sommer les deux coefficients de variation (Insee, la précision des résultats du recensement,p.3).

$$CV_{\text{taux de cadre}} = \sqrt{(CV_{\text{cadre}})^2 + (CV_{\text{population totale}})^2}$$

Il est supposé que les deux coefficients sont indépendants. La documentation précise : « *ce calcul ne tient pas compte de la corrélation entre le numérateur et le dénominateur ; si on en tenait compte, la précision serait en réalité meilleure.* » Plus exactement cette hypothèse d'indépendance est moins gênante lorsque nous effectuons les calculs sur la population totale. Mais dans notre projet de descente d'échelle nous devons nous interroger sur les spécificités locales. Localement la situation d'homogénéité est bien plus rare. Il est bien plus exceptionnel de trouver des petites zones dont la structure de la population est similaire à celle de la population totale. Dès lors à partir de quel effectif de population pouvons-nous mesurer la précision ? La documentation précise qu'il n'est pas judicieux d'opérer avec des effectifs de moins de 2 000 habitants, c'est-à-dire à l'intérieur des IRIS et des très petites communes. Toutefois la documentation est assez lacunaire sur ce point. Pour évaluer de la précision des petites communes, la documentation de l'Insee indique que, pour un effectif de 10 000 habitants, l'écart-type vaut 200, la précision est de 2 % et l'intervalle de confiance à 95 % de plus ou moins 400. (Insee. la précision des résultats du recensement. p.4). La documentation précise à la suite de cette exemple que ce calcul permet d'expliquer pourquoi les résultats tirés de l'exploitation complémentaire ne sont pas utilisés pour les zones de moins de 2 000 habitants. Les bases téléchargeables contiennent les résultats pour toutes les communes mais

65 Insee. La précision des résultats du recensement [en ligne] (Consultée le 10/04/2016
<<http://www.insee.fr/fr/ppp/bases-de-donnees/recensement/resultats/doc/pdf/fiche-precision.pdf>>

à seule fin de permettre des agrégations sur des zones d'au moins 2 000 habitants. Les informations pourront être utilisées avec un niveau de détail d'autant plus grand qu'elles concernent une zone plus peuplée.

En conclusion de cette présentation des supports utilisées par les méthodes de collecte des données du recensement de la population, le support de base de tirage est le support IRIS-2000 servant à équilibrer les groupes de rotations des adresses à enquêter chaque année. Le support de localisation est un support ponctuel composé des adresses postales. De cette manière les alignements aux autres supports est possible (alignement au support IRIS). Enfin, les supports d'analyses diffusées par l'Insee est le support IRIS pour les communes de 10 000 habitants et plus et le support communal pour les communes de moins de 10 000 habitants. Par commodité nous parlons de support IRIS-Communes.

La grande taille de l'échantillon du recensement est conçu pour apporter une bonne précision pour des unités de 2 000 habitants et plus. Il nous faut alors faire attention aux communes de moins de 10 000 habitants surtout lorsque on utilise les données de l'exploitation complémentaire renseignant seulement sur un quart de la population de ces communes. L'utilisation de MOBPRO est donc sous-optimal mais ce problème de sous-échantillonnage peut être contourné par l'utilisation des données confidentielles. Le processus que nous allons appliquer à l'Île-de-France (chapitre 4) pourrait donc être utilisé dans un cadre plus précis. Pour les grandes communes le sondage peut créer des problèmes dans les carreaux peu peuplés, il faut alors espérer une homogénéité sociale suffisante.

Dans notre processus de désagrégation nous avons fait le choix de produire un jeu de donnée sur un carroyage de 1 km, qui parfois contiendra moins de 2 000 habitants ou actifs occupés au km². Ainsi selon la documentation, ces zones de moins de 2 000 habitants seront potentiellement moins précises. C'est pour améliorer la précision du tirage du recensement que nous allons utiliser des données de contrôle permettant de répartir spatialement les observations. Par exemple, la sous populations des actifs appartenant à la catégorie des cadres, devra être localisée en s'appuyant sur des données de contrôle prenant en compte les spécificités de cette catégorie.

Conclusion du chapitre 2

En analysant les diverses fonctions des supports spatiaux utilisées dans la méthodologie de construction d'enquête nous nous sommes aperçu que le besoin en données provient en partie du fait que les supports de production des données ne sont pas initialement construits pour répondre à tous les besoins des analyses. Ils doivent d'abord respecter des contraintes de représentativité. Et si le taux d'échantillonnage est faible nous sommes contraints de préserver le support de tirage pour effectuer des analyses. Nous ne pouvons que ré-agréger sur un autre support d'analyse qu'à condition qu'il soit aligné sur le support de tirage et plus grossier que le support de diffusion. Nous perdons donc en précision géographique.

L'analyse de la chaîne de production et d'analyse des données a ainsi révélé un antagonisme entre les fonctions de production et d'analyses des données. Les utilisateurs souhaitent des données représentatives selon les critères qu'ils souhaitent étudier. Les producteurs de données doivent construire à l'avance des jeux de données répondant aux questions des utilisateurs. Ils doivent construire des données représentatives en fonction des demandes. L'antagonisme entre les deux fonctions de ces supports provient du fait qu'un support construit pour étudier un objet d'étude ne peut pas être utilisé pour étudier un autre phénomène.

Pour concilier les divers besoins d'analyse et les étapes de production de données sans passer par la création d'autres enquêtes, les limites administratives jouaient un rôle d'intermédiaires (supports des IRIS-Communes). Avec les objectifs du Grenelle nous ne pouvons plus nous satisfaire de ces bornes à respecter. Nous devons disposer de supports d'analyse plus fins ou permettant d'étudier des territoires à cheval sur les limites administratives.

L'objectif est aujourd'hui de produire des données les plus neutres possibles vis-à-vis des compétences politiques pour laisser le soin aux utilisateurs de créer eux-mêmes leurs supports d'analyse et faciliter les comparaisons internationales. Pour produire ces données neutres, les observations sont regroupées dans des mailles régulières (carreaux). Cette maille doit répondre à des impératifs de construction pour être aux normes INSPIRE et doit être suffisamment fine pour permettre d'analyser des localisations infra-communales. L'Insee a commencé la diffusion de données de population au carreau en regroupant des informations

issues des fichiers fiscaux localisés à l'adresse et en procédant à des méthodes de brouillage de l'information pour respecter la règle de l'anonymat. Plusieurs maillages sont utilisées, nous retenons un de 200m et un second de 1km de résolution. Nous proposons de continuer dans cette voie, en produisant des indicateurs de demande de mobilité sur le support carroyé de 1km. Plusieurs sources de données et méthodes permettent potentiellement d'effectuer ce travail. Les EMD et la chaîne de modélisation à quatre étapes ont été présentées. Les EMD ayant un faible taux d'échantillonnage les méthodes de modélisation de trafic extrapolent les échantillons à l'aide des données du recensement et des données de comptage. L'avantage de ces méthodes est de traiter de plusieurs motifs de mobilité et d'en donner une photographie d'un jour moyen d'enquête, mais leur désagrégation spatiale ferait selon nous trop d'hypothèse mal contrôlées.

Les sources de données internes à l'Insee peuvent également être utilisées. Nous choisissons de poursuivre une autre voie, en reprenant certains souhaits exprimés par le CNIS (Cuillier, 2009, p.37) la désagrégation des données du recensement au carreau INSPIRE de 1km va permettre de disposer de l'ensemble des variables du recensement sur le carroyage INSPIRE de 1km. De plus, la méthodologie de désagrégation proposée peut être vue comme un moyen d'anonymiser les données sur des carreaux contenant de faibles effectifs.

Pour effectuer ce travail de descente d'échelle du support communal au support carroyé de 1km nous devons disposer d'effectifs d'enquêtés importants et d'une bonne description des profils socio-économiques pour ventiler les observations dans tout l'espace géographique en s'appuyant sur des correspondances avec des données de contrôle. Le chapitre suivant présente la méthodologie permettant de ventiler sous contrainte les données du recensement.

Chapitre 3 : Clonage et ventilations spatiales sous contraintes d'observations statistiques

Chapitre 3 : Clonage et ventilations spatiales sous contraintes d'observations statistiques

Ce chapitre présente la méthodologie permettant de reconstruire une description socio-spatiale fine de la demande de mobilité domicile-travail sous forme de données origines-destinations recouvrant l'espace géographique à relativement grande échelle. Pour effectuer ce travail de descente d'échelle du support communal nous devons disposer d'effectifs d'enquêtés importants et d'une richesse de description des profils socio-économiques. De cette manière nous pouvons ventiler les observations dans tout l'espace géographique des données de contrôle, en s'appuyant sur des correspondances entre les données de contrôle et les profils socio-économiques des observations. La désagrégation spatiale de données de mobilité sur un support plus fin que la commune passe par trois étapes que nous nommons « clonage des observations », « ventilation spatiale sous-contraintes » et « ré-agrégation ».

La ventilation sous contraintes de données de contrôle

Le travail de ventilation sous contraintes consiste à localiser les observations des données de mobilité du recensement à l'intérieur du support communal de diffusion. L'étape de ventilation s'appuie sur des éléments de connaissance externes appelés données de contrôle. L'objectif est d'intégrer la structure spatiale des données de contrôle dans les données à désagréger. Par exemple, nous pouvons utiliser une information fine sur les bâtiments et les effectifs de population qu'ils contiennent pour localiser finement des données d'enquêtes de mobilité. Et, cette localisation reproduira la distribution spatiale des ménages par types de bâtiments à une échelle plus fine que celle des données de mobilité du recensement. Si l'on peut montrer, par ailleurs, que cette information sur les bâtiments structure la mobilité des individus à l'échelle des zones à désagréger, alors nous considérerons que le fichier désagrégé contient une information enrichie sur le phénomène qui nous intéresse.

Par exemple, si nous enquêtons à domicile un ménage sur ses déplacements, les individus ont bien des origines-destinations spécifiques. Si nous revenons l'année d'après et que nous nous apercevons que le ménage a déménagé et a été remplacé par un autre alors les origines-destinations des individus du nouveau ménage peuvent avoir changées. Entre l'ancien ménage

et le nouveau ménage il y aura des variants et des invariants. Et ce sont ces invariants que nous tentons de préserver. D'une part l'origine des déplacements sera toujours le domicile. Ensuite, nous pouvons imaginer que le nombre d'individu du ménage aura une probabilité assez forte d'être sensiblement le même ou proche, car il est corrélé à la superficie du logement. Aussi les revenus par unité de consommation seront similaires en considérant que la dépense pour se loger (loyer ou achat) seront proches. Enfin, les lieux de destinations seront bien différents mais dépendront toujours de l'accessibilité du logement aux emplois pour la mobilité domicile-travail et aux aménités urbaines pour les autres formes de mobilité. De la même manière que les localisations résidentielles, les localisations aux lieux d'emploi des individus peuvent donc être approchées à l'aide de données structurantes. En dépit du turn-over des salariés les individus remplaçant les anciens salariés ont des profils sociologiques similaires. Ces exemples démontrent que nous pouvons nous servir d'une description des bâtiments pour ventiler les observations statistiques. Peu importe qui occupe réellement les bâtiments, ceux-ci pré-déterminent en partie les profils socio-économiques de leurs utilisateurs. L'opération de désagrégation risque de réduire néanmoins la part des profils atypiques (par rapport à la zone de désagrégation) au profit des profils les plus courants.

Du moment que nous connaissons les caractéristiques des bâtiments et/ou de zones plus fines que les communes, nous pouvons affecter des individus à l'intérieur des communes en fonction des caractéristiques des données, dites alors « de contrôle », renseignant sur les bâtiments. L'hypothèse sous-entendue est que ce sont les infrastructures qui déterminent un éventail de possibilités des profils socio-économiques que les infrastructures peuvent accueillir. La question est alors de quelles données disposons-nous sur les bâtiments ou au moins à une échelle infra-communale ? Les meilleures données sont celles qui associent la localisation fine avec un maximum de données attributaires. Toutefois, nous allons être confrontés à un problème d'adéquation entre les effectifs d'observations des données du recensement à ventiler et les effectifs d'observations des données de contrôle. Nous proposons alors d'augmenter quelque peu les effectifs d'observations du recensement.

Le clonage d'observations statistiques :

La ventilation des observations en couvrant le maximum possible l'espace géographique pose le problème de l'adéquation entre les effectifs d'observations et les effectifs, nécessairement plus complètes, des données de contrôle que nous utiliserons. Si

l'effectif de l'enquête est inférieur aux données de contrôle, nous ne pouvons pas distribuer correctement les observations d'enquêtes dans tous les bâtiments. Nous devons alors augmenter artificiellement la taille de l'échantillon des enquêtes. Pour ce faire nous présentons le concept de clonage d'observations statistiques qui détourne quelque peu l'utilisation des poids de sondage statistique. Nous considérons qu'une observation du recensement en représente autant que son poids de sondage l'indique. L'utilisation des poids de sondage comme coefficient multiplicateur pour cloner les observations est à considérer avec prudence. Surtout lorsque les observations ont des poids de sondage élevés⁶⁶. Une observation en représente dans ce cas beaucoup d'autres. C'est pour minimiser les hypothèses de construction que nous allons laisser groupées le plus possible les observations clonées. La première section du chapitre définit et précise l'utilisation des clones statistiques.

La ré-agrégation :

L'opération de ventilation spatiale sur un support plus fin que le support de tirage apporte des incertitudes. Nous nous appuyons sur des données de contrôle mais il y a également de la stochasticité dans la ventilation. Nous allons utiliser des tirages de Monte-Carlo. Pour ré-introduire de la robustesse statistique nous allons ré-agréger les observations sur un support plus grossier que celui sur lequel la ventilation a été effectuée mais plus fin que le support communal de diffusion. Ainsi nous espérons gagner en robustesse tout en affinant les localisations des observations. Par exemple, nous allons ventiler les localisations résidentielles des observations du recensement sur le carroyage de 200m de l'Insee et les ré-agréger sur le carroyage INSPIRE de 1km. Nous espérons ainsi obtenir une meilleure représentativité statistique qu'en désagrégeant directement sur le support carroyé de 1km. Nous ventilons les observations au lieu d'emploi en utilisant des données sur les lieux d'emploi plus fines que la commune.

L'opération de ré-agrégation des localisations résidentielles ne pose aucun problème technique puisque les supports carroyés de 200m et de 1km sont alignés. Toutefois dans certains cas le carroyage de 1km peut ne pas être pertinent dans les analyses. Par exemple, si nous souhaitons étudier les aires de rabattement en transport en commun, il se peut que les carreaux de 1km ne soient pas centrés sur les stations de transport. Pour pouvoir s'en assurer nous laissons dans les fichiers de résultats les identifiants et les coordonnées des carreaux de 200m attribués par la ventilation. De cette manière, les résultats peuvent être ré-agrégés sur

⁶⁶ Le poids de sondage moyen de l'Enquête Globale Transport de 2010 est d'environ 320.

un autre support que le carroyage de 1km que nous allons utiliser. Nous insistons sur le fait que les résultats de la ventilation doivent être ré-agrégés pour consolider la représentativité statistique.

Ce chapitre expose donc premièrement les choix méthodologiques pour cloner les observations du recensement rénové et obtenir ainsi des individus statistiques plus nombreux dont nous localiserons le domicile et l'emploi sur le carroyage de 1 km. Deuxièmement, la méthodologie générale de ventilation spatiale sous contrainte est présentée. Enfin comme nous visons la désagrégation sur un carroyage de 1km, nous faisons le constat qu'il est nécessaire d'utiliser des données de résolution inférieure à 1 km, avant de pouvoir ré-agrégérer les couples synthétiques à cette échelle. Il est donc utile de disposer de données de contrôle sur un support plus fin que celui du carroyage cible de 1 km. La dernière section du chapitre présente les données que nous allons utiliser dans le processus appliqué à l'Île-de-France.

3.1 Le clonage des observations du recensement

La méthode de désagrégation spatiale des données du recensement utilise une étape de clonage. Cette opération consiste à augmenter artificiellement le nombre d'observations d'une enquête pour avoir plus d'individus statistiques à ventiler. Le clonage a pour but d'améliorer l'adéquation entre les profils disponibles dans l'échantillon et ceux des données de contrôle. L'espace occupé par les données de contrôle sera mieux rempli qu'avec un petit nombre d'observations.

Bien que l'opération de clonage ne crée pas de nouveaux profils, l'augmentation artificielle d'observations par clonage peut s'apparenter aux méthodes de génération de populations synthétiques. Cette construction vise à respecter le lien entre les localisations spatiales, les profils socio-économiques et les modes de transport utilisés pour aller travailler, de manière à obtenir des données cohérentes et statistiquement représentatives. De plus, nous devons disposer de descriptions socio-économiques riches pour effectuer des correspondances avec les données de contrôle présentées par la suite. La construction d'une population synthétique, par clonage des observations d'enquêtes selon leur poids de sondage, permet de recouvrir l'espace informé par les données de contrôle.

Prenons le cas d'un couple de zones dans le support source dans laquelle nous voulons localiser les observations au lieu de résidence mais dont l'effectif d'observations est inférieur au nombre de bâtiments cibles découpant plus finement les zones résidentielles sources (cela peut être des bâtiments mais aussi des zones). Par exemple, nous avons trois observations de déplacements dans une zone source et quatre bâtiments cibles à l'intérieur de la zone source. Nous ne pouvons pas distribuer ces trois observations dans les quatre bâtiments cibles. Afin de disposer d'unités dans chaque bâtiment la méthode de désagrégation doit produire des unités dites alors artificielles. La création de ces unités passe par l'utilisation des poids statistiques. Dans notre exemple, disons que chacune des trois observations ont au début un poids de 4. Il faut créer une quatrième observation, nous pouvons diviser la somme des poids de sondage des observations du couple de zone source par l'effectif d'observation. Mais dans ce cas il faut construire les données attributaires pour la nouvelle observation. Nous proposons plus simplement de cloner une observation du couple de zone pris au hasard. Cette étape est conceptuelle. Nous supposons qu'une observation représente autant d'individus que son poids de sondage l'indique. Une

observation est alors constituée de plusieurs clones. Et pour couvrir l'ensemble de l'espace des données de contrôle nous séparons dans certains cas les clones des observations. Dans notre exemple, deux observations auront un poids de 4 et deux autres auront un poids de 2 et auront les mêmes données attributaires. Cette opération consiste simplement à reconstituer des individus statistiques identiques (c'est-à-dire possédant exactement le même profil socio-économique) à partir d'une même observation et de son poids de sondage. Nous pouvons bien entendu créer autant d'unités d'observations artificielles que souhaité.

La critique que l'on peut porter à cette approche est que l'on ne construit pas de nouvelles destinations (communes de travail) pour les observations clonées ni même de nouveaux profils socio-économiques et modes de transport. Ce fait peut être gênant avec un échantillon trop petit qui ne contient pas la totalité des profils des données de contrôle.

Il existe pourtant des méthodes permettant de construire de nouvelles localisations et profils socio-économiques à partir d'un échantillon. Nous avons déjà présenté les étapes de modélisation de trafic pour ce qui concerne les variables de localisations. Pour la construction de profils socio-économiques plusieurs méthodes peuvent être utilisées : les méthodes d'Iterative Proportional Fitting (IPF) sont utilisées dans la littérature de création de populations synthétiques (Deming et Stephan, 1940) (Beckman, 1996). D'autres méthodes peuvent être utilisées, par exemple nous pouvons construire des matrices de similarité (et/ou de dissimilarité) (Duncan et Duncan, 1955) afin de comparer les partitions (Hubert et Arabie 1985), puis nous pouvons appliquer des méthodes de ré-échantillonnage (Efron et Tibshirani, 1994) pour produire de nouvelles partitions les plus similaires possibles mais pas de trop.

Mais ces méthodes s'appuient sur des effectifs totaux à respecter. Ces effectifs totaux sont dans bien des cas les données du recensement. En travaillant directement avec les données du recensement rénové les effectifs totaux des profils des individus doivent être considérés avec prudence (chapitre 2.3.1). C'est pourquoi nous choisissons de ne pas appliquer les méthodes de création de nouveaux profils socio-économiques dans le cadre de cette thèse. Il est supposé que la taille initiale de l'échantillon enquêté dans MOBPRO est suffisamment grand (au quart dans les petites communes et au deux cinquièmes dans les grandes) pour offrir des observations de l'ensemble des situations les moins exceptionnelles. Plus l'effectif d'observation est grand plus nous nous approchons de l'ensemble des modalités des variables de localisations et socio-économiques des zones de tirage. Avec un grand effectif, il est supposé que les modalités sont assez représentatives. Nous prenons le parti qu'il n'est pas

nécessaire ici de créer de nouvelles modalités de variables socio-économiques, ni même de nouvelles communes de travail. De plus, les méthodes de générations synthétiques s'appuient sur des marges totales à respecter. Pour appliquer ces méthodes il faut de toute manière multiplier les nombres d'observations par leur poids de sondage. L'opération de clonage est une étape préalable à la création de nouveaux profils socio-économiques. Nous pourrions éventuellement explorer cette voie ultérieurement.

L'augmentation artificielle des observations d'enquêtes se fait donc simplement en considérant que les observations représentent autant d'individus que leur poids de sondage l'indique. Et, lors de la ventilation des observations statistiques dans l'espace des communes, nous pouvons soit séparer dès le départ les clones des observations statistiques, soit nous pouvons les laisser agrégées. Pour ne pas faire d'hypothèses de construction trop fortes nous suivons la deuxième proposition: nous allons laisser les clones groupés le plus possible. C'est-à-dire nous allons leur affecter, autant que faire se peut, les mêmes localisations résidentielles et d'emplois. De cette manière, au lieu de créer autant de lignes (d'observations) dans les fichiers numériques qu'un poids de sondage l'indique, nous maintenons groupés les observations que nous ne souhaitons pas dissocier. Ajoutons que de cette façon les temps de calculs sont plus courts et la mémoire vive nécessaire moins importante. Dans certains cas nous sommes toutefois amenés à séparer des clones car sans cela nous ne respecterions pas les effectifs de population indiqués par les données de contrôle. L'opération revient donc, dans ces cas, à modifier marginalement les poids de sondage. La séparation de clones intervient dans les étapes de ventilation spatiale sous contraintes (chapitre 4.2).

3.2 La ventilation spatiale sous contraintes

Une fois les d'observations statistiques clonées (mais pas encore séparés, comme nous venons de le voir), nous allons ventiler spatialement les observations du recensement rénové en utilisant des données de contrôle pour choisir les lieux de résidence pour une grande partie des observations du recensement⁶⁷ et des lieux de travail pour les actifs occupés. Les observations du recensement sont diffusées à l'aide de plusieurs fichiers détail, et pour faciliter l'exposé de notre méthode, nous considérons pour l'instant que seul le fichier des mobilités professionnelles est à désagréger. D'autres fichiers détail sont en fait utilisés comme des sources de données de contrôle. En traitant du fichier MOBPRO décrivant les communes

⁶⁷ Nous allons voir que nous ne pouvons pas traiter les ménages avec aucun actif occupé.

de résidence et d'emploi, la question est de savoir ventiler les couples de localisations domicile-travail.

Notre objectif est de produire des localisations domicile-travail plus précises que les références des communes présentes dans MOBPRO, à l'aide de données de contrôle, pour respecter des structures socio-spatiales existantes. La difficulté d'un processus de descente d'échelle est de construire une répartition spatiale des couples de localisations domicile-travail cohérente.

Quand nous passons en revue les différents jeux de données pouvant servir de données de contrôle, nous constatons que, de manière générale, plus un jeu de données possède une grande précision spatiale, moins celui-ci offre une information détaillée sur les caractéristiques attributaires des zones et des individus, pour des raisons de coût et en raison du secret statistique. En revanche plus les zones sont grandes, mieux elles peuvent être caractérisées, ce qui peut sembler un peu paradoxal⁶⁸.

Pour désagréger, nous allons construire un jeu de données riches en variables socio-économiques portant sur un support spatial équivalent ou plus fin que le carroyage de 1 km qui est la référence spatiale cible de la désagrégation, par appariement de différentes sources de données. Ce jeu de données, à condition qu'on puisse le construire concrètement, agirait comme un moyen pour contrôler la répartition spatiale des observations à l'intérieur des zones sources (les communes) vers les zones cibles (les carreaux).

Cette construction simple est intellectuelle, et se heurte au principe de réalité. Elle ne va pas se matérialiser directement par un seul et unique support pivot de l'information géographique. En pratique, nous allons effectuer plusieurs ventilations spatiales sur des supports de plus en plus fins. Les premières descentes d'échelle pourront s'appuyer sur une description riche des sous-espaces, les suivantes introduiront plus de stochasticité. La méthodologie générale de ventilation sous contraintes est présentée dans la suite de ce chapitre. Le chapitre 4 exposera explicitement le cas d'application et les différentes étapes de ventilation sous contraintes utilisées dans cette recherche.

Le processus de ventilation se décompose en plusieurs étapes. Comme nous traitons de

68 Pour décrire les zones il faut a priori compter les éléments individuels et donc procéder à une enquête coûteuse. Nous retrouvons ici l'intérêt de procéder à une enquête par échantillonnage et le concept de représentativité statistique (chapitre 2) et notre besoin en désagrégation.

données de déplacements et donc de données bi-localisées il est important de traiter de l'appariement des couples domicile-travail, ce qui différencie notre travail d'une désagrégation spatiale de données classique.

Le travail de descente d'échelle de données bi-localisées se décompose en trois étapes théoriques :

- le choix des données de contrôle pour construire un éventail de possibilités des lieux de résidence pouvant recevoir des observations à l'intérieur des communes ;
- le choix des données de contrôle pour construire un éventail de possibilités des lieux de travail pouvant recevoir des observations à l'intérieur des communes ;
- la distribution des observations dans des lieux de résidence et de travail dont le couple constitue un déplacement réalisable (contraintes de bi-localisations).

En utilisant le vocabulaire de la modélisation à quatre étapes, nous pouvons dire qu'il s'agit de la première étape de génération (sources et puits de mobilité) et de la deuxième étape de distribution (appariement des origines et des destinations⁶⁹). Cependant, notre approche ici est purement statistique et désagrégée, puisque nous nous appuyons sur les observations du recensement.

Bien que notre objet d'étude soit des données bi-localisées, nous présentons d'abord les éléments généraux de la méthode de changement de support appliqués à des données localisées : l'inférence écologique (Chapitre 3.2.1).

Forts de ces éléments de connaissance, nous montrons ensuite les données que nous allons utiliser pour localiser plus finement les lieux de résidence (partie 3.2.2) et d'emploi (partie 3.2.3) à l'intérieur des communes d'Île-de-France. Nous introduirons ensuite une variable de contrôle un peu particulière : la distance des trajets composant les déplacements (3.2.4) pour faire face aux enjeux spécifiques de la bi-localisation des informations.

69 La modélisation à quatre étapes réalise en fait une descente d'échelle (sur un réseau). Un enrichissement des données collectées par une enquête de terrain (EMD) est effectué avec les données du recensement et/ou des données de comptage des gestionnaires des réseaux (étape 1). Les origines-destinations sont construites à l'aide d'un modèle théorique : le modèle gravitaire (étape 2). Ensuite, les modes de transports sont construits et les déplacements synthétiques sont affectés sur les réseaux (étape 3 et 4). Nous appliquons finalement la même démarche d'extrapolation dans notre travail de désagrégation des données de mobilité du recensement. Cependant, nous nous servons le moins possible de modèle théorique.

3.2.1 L'inférence écologique – extension aux données bi-localisées

La distribution spatiale des points de génération de déplacements (origines : domiciles, et destinations : emplois) doit être approchée à l'aide de données de contrôle. Nous disposons dans le recensement d'observations dont les localisations ont été agrégées dans une zone assez vaste : la commune. En leur choisissant une nouvelle localisation au hasard à l'intérieur de la zone de regroupement nous avons de grandes chances de ne plus respecter les logiques de choix restreint (chapitre 1.1). Nous pouvons placer des individus ayant un mode de transport ou un profil socio-économique de manière non cohérente avec les logements, avec les activités, ou encore avec le système de transport. Pour éviter des hypothèses fallacieuses, et nous tenir à notre approche statistique (c'est-à-dire ne pas utiliser de modèle trop normatif comme par exemple le modèle gravitaire) nous devons utiliser des variables socio-économiques pour localiser les observations du recensement rénové aux bons endroits ou au moins à des endroits possibles ou, mieux encore, probables.

Nous voulons construire la répartition spatiale des couples domicile-travail en considérant la distribution spatiale des données de contrôle qui permettent de traduire les contraintes de localisations qui ont été en œuvre lors du processus de localisation des ménages. Nous allons nous appuyer sur des données désagrégées, ainsi que sur des indicateurs agrégés nous renseignant sur les contraintes de localisations.

Afin de faciliter la compréhension de la méthode, faisons pour l'instant comme si nous disposions d'un jeu de données sur les bâtiments offrant des variables similaires aux variables décrivant les observations d'enquêtes à ventiler. Il sera ainsi possible de réaliser un appariement entre les bâtiments et les observations statistiques. Par exemple, si nos observations comportent une variable décrivant la superficie des logements occupés, faisons comme si nous avions un jeu de données de bâtiments par commune comportant une information de superficie.

Les couples domicile-travail sont des données bi-localisées. Nous devons alors ventiler les origines et les destinations sur les supports géographiques des données de contrôle. La spécificité de la ventilation de données bi-localisées est de devoir affecter à chaque observation deux localisations sur les supports de données de contrôle en veillant à ce

que le couple origine-destination ainsi construit soit cohérent. La question de la ventilation pose alors trois problèmes : le choix du support de contrôle représentant les lieux d'origine, le choix du support de contrôle pour représenter les lieux de destination et le choix de l'appariement entre les origines et les destinations.

Ces choix de construction sont à mettre en interdépendance. Le processus de désagrégation doit construire d'un seul coup les deux localisations. Contrairement à la désagrégation de données uni-localisée la désagrégation de données bi-localisées nous confronte à une double inconnue dont aucune donnée de contrôle unique ne peut approcher ces deux composantes d'un seul coup.

Il nous faut des données de contrôle pour les origines et pour les destinations. Mais cela ne nous dit pas comment ensuite effectuer l'appariement des couples origines-destinations domicile-travail. L'idée est la suivante : en disposant de données attributaires à la fois sur les origines et à la fois sur les destinations pour une même observation nous effectuons deux méthodes de ventilations successives. Dans un premier temps les origines sont ventilées à l'aide de méthodes de ventilation de données ponctuelles. Par construction, cette première étape va en même temps localiser les variables attributaires décrivant les destinations aux lieux d'origines. Ceci permettant de traiter de l'interdépendance du couple origine-destination. Ensuite, nous ventilons les destinations aux lieux de destinations. Cela est possible si nous disposons de variables décrivant les observations à la fois à leur lieu de résidence et à leur lieu d'emploi (le secteur d'activité renseigné par le recensement rénové).

La ventilation spatiale multicritère

Illustrons le problème en commençant par la ventilation de données uni-localisées. Nous disposons d'une population dans une zone dite source. Nous souhaitons distribuer spatialement les individus dans deux zones cibles à l'intérieur de la zone source. Nous cherchons à créer une variable de localisation tout en respectant les logiques de localisation des ménages (chapitre 1.1). A l'échelle des zones, il s'agit d'indicateurs agrégés regroupant plusieurs ménages. Il faut respecter dans chaque zone cible la proportion de plusieurs modalités des variables socio-économiques décrivant les individus. Prenons, le sexe et le mode de transport. Comment respecter les proportions⁷⁰ d'hommes et de femmes utilisant la voiture personnelle ou les transports en commun à ventiler dans les deux zones cibles?

⁷⁰ Nous ne parlons pas d'effectif car il dépend du taux de sondage de l'enquête. De plus dans la littérature le problème est toujours posé avec des proportions.

Pour proposer une réponse opérationnelle nous pouvons adopter un raisonnement probabiliste. Construisons des probabilités conditionnelles à l'aide de données de contrôle pour construire une variable de localisation. Nous avons vu que les variables caractérisant les localisations peuvent être discrétisées avec un support géographique. Ainsi les modalités sont en nombre finis. Nous notons cette variable Z . L'objet de notre problème est de construire les probabilités conditionnelles des modalités de la variable de localisation à ajouter ($z_1, z_2, z_3 \dots z_n$) avec plusieurs variables de contrôle.

$$P(Z = z_1 | X = x_1 \text{ et } Y = y_1 \text{ et } \dots \text{ et } T = t_1)$$

Ce type de problème est récurrent. Pour le rendre générique, nous reprenons la formulation des travaux de Gary King quand il traite de ce qu'il nomme « l'inférence écologique », probablement en référence aux problèmes d'erreurs écologiques introduits par Robinson (Robinson, 1950). Adaptons sa formalisation à notre contexte en ajoutant une troisième variable, celle de la localisation. L'exemple suivant démontre l'utilisation que nous allons faire des données de contrôle. Si nous disposons des couples (n-uplets) de modalités des variables nous sommes en face d'un système plus contraint. Le fait que le système soit plus contraint à résoudre nous permet de réduire les espaces des solutions (Goodman 1953, 1959), (Duncan et Davis, 1953)⁷¹.

Pour reprendre la manière dont est classiquement exposé le problème de l'inférence écologique nous partons de l'exemple suivant. Nous disposons d'une population dont nous connaissons certaines caractéristiques. Si aucune variable ne décrit la localisation des individus nous supposons que la population est contenue dans une seule zone. Effectivement, lorsque on compte un effectif il est toujours compris dans un ensemble spatial. L'inférence écologique consiste à retrouver les croisements des différents modalités de variables non spatiales à l'intérieur d'une zone. Dans le cas de deux variables, nous pouvons représenter le problème à l'aide d'un tableau de contingence. Nous disposons des marges du tableau i.e les effectifs de la zone et nous cherchons à estimer les cellules.

71 Précisions que Goodman et Duncan étaient des amis et collègues du même laboratoire de sociologie à l'université de Chicago. La méthode des *bounds* utilise la régression de Goodman.

	Genre(x)		
Mode de transport(y)	Femmes (x1)	Hommes (x2)	Total
Bus (y1)	β_{x_1,y_1}^z	$1 - \beta_{x_1,y_1}^z$	$Y_1 = 0,73$
Voiture (y2)	β_{x_1,y_2}^z	$1 - \beta_{x_1,y_2}^z$	$1 - Y_1 = 0,27$
Total	$X_1 = 0,40$	$1 - X_1 = 0,60$	
Tableau 4			

Les travaux présentés dans les ouvrages de Gary King posent le problème de savoir comment construire les probabilités jointes des intersections du tableau de contingence sans hypothèses de construction. Il reprend les travaux de Duncan pour montrer que l'on peut estimer les espaces de solutions possibles à l'aide des méthodes de bounds. Ensuite, il montre comment à l'aide de tirage de Monte-Carlo on peut proposer plusieurs distributions possibles (King, 1997a).

Notre cas est un peu différent, puisque nous introduisons une nouvelle variable dans une population pour décrire sa localisation (en fait, deux localisations, celle du domicile et celle du travail). Il faut alors construire les probabilités jointes non pas de tous les croisements de modalités mais construire celles que l'introduction d'une nouvelle variable demande. Dans le cas de l'introduction d'une variable de localisation Z à deux modalités, nous cherchons à construire les probabilités conditionnelles pour les modalités z_1 et z_2 de la variable de localisation. Pour les lister nous devons faire deux tableaux, car on est en présence de trois variables qui ne peuvent être croisées simplement.

	Genre(X)		
Zones cibles (Z)	Femmes	Hommes	Total
1	β_{z_1,x_1}^y	$1 - \beta_{z_1,x_1}^y$	Z_1
2	β_{z_2,x_1}^y	$1 - \beta_{z_2,x_1}^y$	$1 - Z_1$
Total	$X_1 = 0,40$	$1 - X_1 = 0,60$	
Tableau 5			

Zones cibles(Z)	Mode de transport (Y)		Total
	Bus	Voiture	
1	β_{z_1,y_1}^x	$1 - \beta_{z_1,y_1}^x$	Z_1
2	β_{z_2,y_1}^x	$1 - \beta_{z_2,y_1}^x$	$1 - Z_1$
Total	$Y_1 = 0,73$	$1 - Y_1 = 0,27$	

Tableau 6

L'idée proposée est donc de construire les probabilités jointes de chaque croisement de modalités à l'aide de données de contrôle C.

$$\beta_{z_1,x_i} = \frac{c_{z_1,x_i}}{\sum c_{z_1,x_i}}$$

Puis de les appliquer aux effectifs N des observations des zones sources à ventiler. Ces effectifs dépendent du taux d'échantillonnage de l'enquête utilisée. Avec cette méthode nous observons que z_i peut être approché de plusieurs manières selon la finesse de description souhaitée et selon le nombre quelconque de variables de contrôle C dont on dispose pour construire ces coefficients. Toutefois nous ne pouvons pas résoudre ce système de probabilités jointes simplement.

Nous voulons distribuer les observations dans l'espace en respectant les effectifs donnés par chacun de ces produits. Nous abordons cette question comme la réunion de deux problèmes de combinatoire ou de deux problèmes de satisfaction sous contraintes: le problème de multi-sac-dos et le problème de sac-à-dos multi-dimensionnel (Kellerer, 2004) :

- le premier consiste à répartir un ensemble d'objets dans plusieurs sacs-à-dos (plusieurs modalités de variables de localisations) de capacités différentes (d'effectifs de population qui sont ici inconnus) ;
- le deuxième considère que le sac-à-dos à plusieurs dimensions. Le nombre de dimension est égal aux modalités des variables et il ne faut déborder sur aucune des dimensions. La *pycnophylactic constraint* (Tobler, 1979) est multiple. C'est en cherchant à résoudre le problème du sac-à-dos multidimensionnel que nous allons approcher une solution pour le problème du multi-sac-à-dos (un effectif de population par zone cible z_1).

Nous présentons dans le chapitre 4 les différentes méthodes de résolution de ces problèmes de satisfaction sous contraintes. Nous verrons également que, si besoin il y a, nous pouvons séparer nos clones d'observations pour atteindre précisément les effectifs à placer dans les zones (chapitre 4.2) et ainsi respecter la *pycnophylactic constraint*.

Le fait de multiplier les données de contrôle (les dimensions) dans la ventilation permet d'outrepasser l'hypothèse d'indépendance de ces variables. Il se peut qu'une donnée de contrôle que nous utilisons ne permette pas de distribuer les observations de manière contrôlée en un endroit donné alors qu'ailleurs cette même donnée de contrôle pourrait être utile. C'est pourquoi nous devons maximiser l'emploi de variables de contrôle spatialisées.

Toutefois, lorsque nous ventilons dans l'espace géographique des observations avec un grand nombre de variables socio-économiques nous ne pouvons que rarement disposer ou construire ces coefficients pour tous les couples de modalités de variables. Il y a des variables disponibles et d'autres non.

Les modalités des variables prises en compte vont contraindre la répartition spatiale des observations. Les coefficients appliqués aux grandeurs des effectifs à ventiler vont nous donner des effectifs de modalités à respecter.

Les modalités des variables non prises en compte dans la construction des quantités à répartir pour chaque modalité des variables de localisation peuvent aussi être mises à contribution mais de moindre manière. Si nous ne construisons pas les couples de coefficients de certaines modalités celles-ci seront réparties selon les coefficients des autres croisement de modalités. Par exemple, s'il se trouve qu'il y a plus de cadre que d'ouvriers parmi les propriétaires, lors de la ventilation d'un propriétaire, la probabilité que ce propriétaire soit un cadre sera différente de la probabilité que ce soit un ouvrier. On se rapproche d'un problème de calage sur marges, qui ventile la répartition des critères des modalités non prises en compte en fonction des effectifs des variables considérés dans le système de contraintes des probabilités jointes (Deming et Stephan, 1940). Ces probabilités, exprimées comme un rapport de côtes, sont estimables à l'aide des zones sources. Par construction, nous reproduisons les relations entre les variables socio-économiques de l'échelle des zones de tirage aux zones cibles.

Les relations peuvent être positives ou négatives. Premièrement les variables non prises en compte ayant un lien de corrélation positif avec les variables corollaires aux variables de contrôle auront tendance à être réparti de la même manière que la distribution spatiale des données de contrôle. Deuxièmement les variables non prises en compte et n'entretenant pas

de lien de corrélation avec les variables considérées dans la ventilation vont être réparties de manière aléatoire. Les observations n'ayant pas les modalités à placer en priorité par les données de contrôle seront réparties ailleurs et de manière homogène. Ces effets peuvent être bénéfiques pour le problème de ventilation de données bi-localisées.

Dans le cas de données bi-localisées la distribution des variables socio-économiques non prises en compte dans la ventilation sous contraintes nous intéresse tout particulièrement. En effet, parmi les variables qui ne sont pas contrôlées il se trouve forcément les variables décrivant les localisations autres que celles que les données de contrôle permettent d'approcher. Dans notre cas d'application, quand on ventile les observations dans les lieux de résidence, nous localisons dans le même temps des modalités de variables décrivant des lieux de travail (commune d'emploi et secteurs d'activités économiques des actifs occupés). Par construction, les localisations des secteurs d'activités aux lieux d'origines dépendront des probabilités jointes entre les variables prises en compte dans la ventilation au lieu de résidence et la variable secteur d'activité. C'est avec ce lien que nous supposons considérer d'un seul tenant les origines, les destinations et l'appariement des couples origines-destinations domicile-travail. Le fait de localiser au lieu de résidence les variables décrivant les lieux d'emploi puis de procéder dans un deuxième temps à la ventilation spatiale des lieux de travail permet de traiter le cas de l'interdépendance des données bi-localisées (illustration 23).

Afin de renforcer encore le lien entre l'origine et la destination de chaque couple origine-destination domicile-travail, nous pouvons utiliser une donnée de contrôle supplémentaire et un peu particulière : la distance des trajets composant les déplacements. Celle-ci peut être celle entre les lieux de résidence et d'emploi ou une distance d'accès aux systèmes de transport décrite par les observations. Pour rester dans le formalisme reposant sur des variables discrètes, nous construisons donc une variable supplémentaire : la classe de distance. Bien sûr s'il l'on souhaite considérer les distances en les profilant selon des variables socio-économiques il nous faut également construire autant de coefficient que de couples de modalités souhaités. Les contraintes que l'on peut introduire dans les méthodes de ventilation sous contraintes dépendent des données de contrôle disponibles.

L'hypothèse de l'existence d'un jeu de données sur les bâtiments avec des variables similaires aux variables des observations du recensement rénové est bien sûr purement

fictive. Un tel jeu de données n'existant pas dans la réalité nous ne pouvons pas disposer de l'ensemble des variables de contrôle souhaité idéalement à l'échelle des bâtiments. Nous allons effectuer une série de descente d'échelle, ce qui revient à utiliser des sources de données de contrôle hétéroclites pour créer ces coefficients. Les différentes données de contrôle utilisées dans le cas d'application à l'Île-de-France sont maintenant présentées.

3.2.2 Les données de contrôle pour ventiler les lieux de résidence

La méthodologie de désagrégation spatiale des données du recensement s'appuie sur des données de contrôle. La suite de ce chapitre présente les données de contrôle que nous allons utiliser dans notre cas d'application à l'Île-de-France. Nous commençons par présenter la source de données à désagréger. Les données de mobilité du recensement sont diffusées par l'intermédiaire du fichier MOBPRO de l'exploitation complémentaire (chapitre 2.3). Nous avons annoncé qu'il existe un fichier confidentiel non diffusé publiquement regroupant l'ensemble des variables de l'exploitation complémentaire des fichiers diffusés librement. Toutefois nous faisons le choix d'utiliser les données disponibles pour le grand public. De cette manière nous considérons que le fichier MOBPRO est la source de données à désagréger. Et, nous considérons les autres fichiers détail, en particulier le fichier INDCANTVILLE, comme des sources de données de contrôle.

Nous présentons donc le fichier MOBPRO, le fichier INDCANTVILLE puis les données carroyées de population. Cette présentation s'attache à montrer les différences de supports des données et les différences d'effectifs de population dénombrés par ces fichiers.

3.2.2.1 Le support communal du fichier MOBPRO

Puisque nous nous intéressons à la dimension spatiale de la mobilité domicile-travail, nous commençons par un travail de cartographie du support de l'information géographique du fichier MOBPRO. Ce travail permet de présenter la zone d'étude, et de comparer les différents niveaux de finesse des données de contrôle présentées par la suite. Cela permet aussi d'exposer les effectifs d'individus que les différents sources de données et supports comptabilisent.

Dans le cas d'application de cette thèse, le fichier utilisé est le fichier MOBPRO 2012 diffusé en octobre 2015 sur le site de l'Insee. Il estime 5 398 596 actifs occupés au 1er Janvier 2012. Il contient 1 955 156 observations avec un poids moyen de 2,8. Comme la cartographie des couples origines-destinations serait illisible, la répartition spatiale des observations MOBPRO est décomposée en deux cartes : les effectifs aux lieux de résidence et aux lieux d'emploi (illustrations 19 et 20).

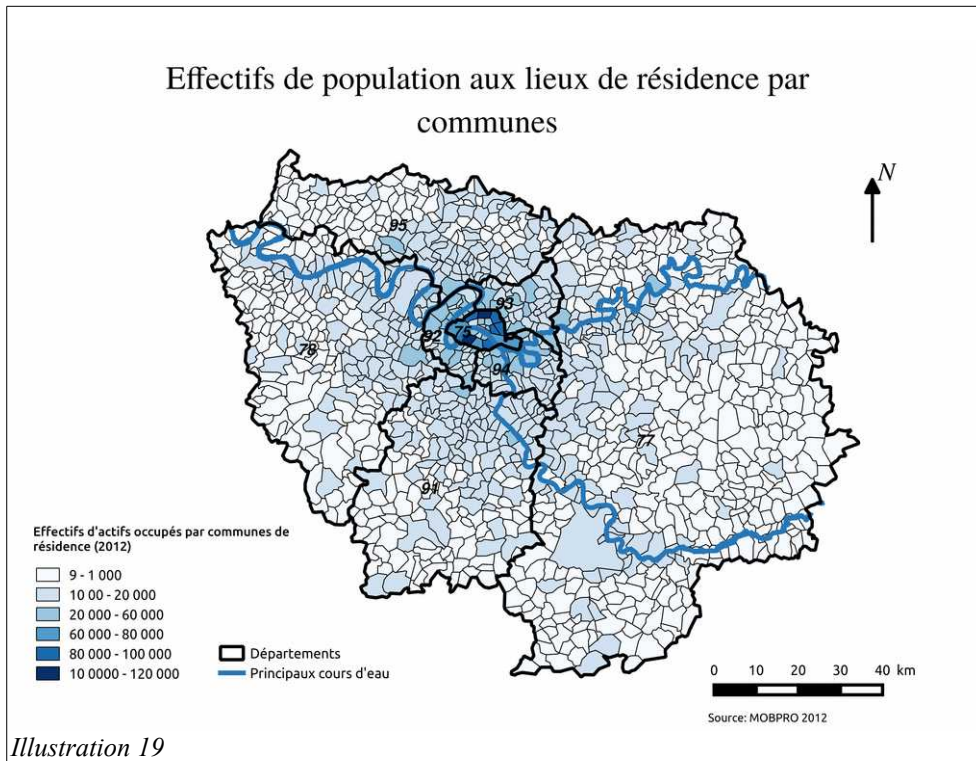


Illustration 19

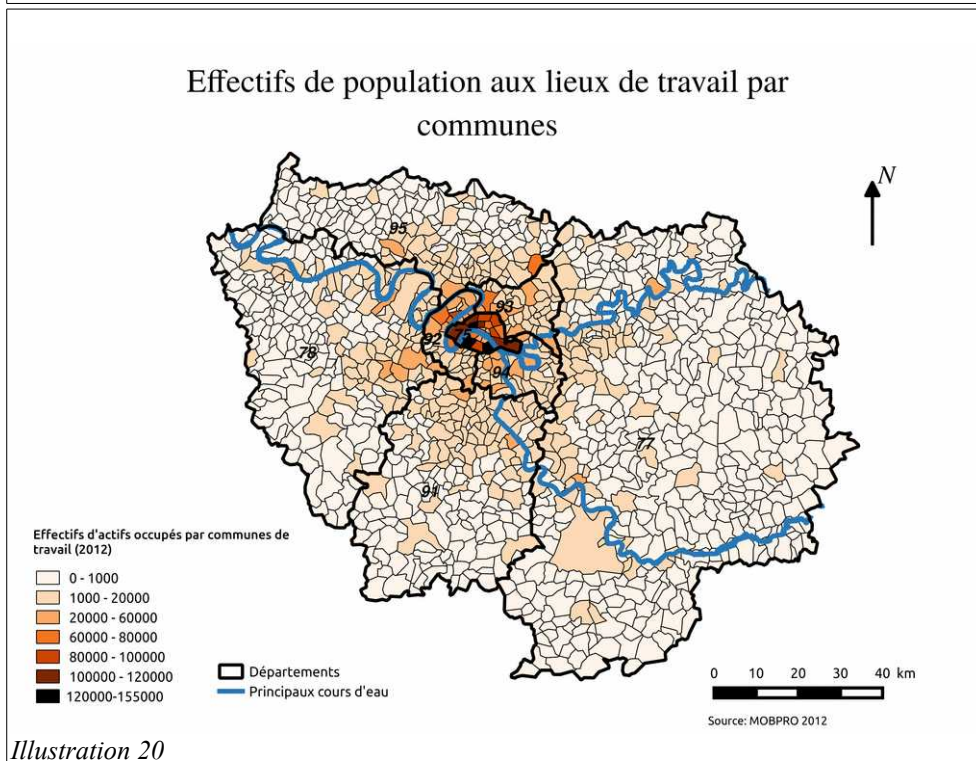


Illustration 20

Le département parisien puis les départements constituant la petite couronne (92,93,94) sont ceux concentrant les populations aux lieux de résidence et d'emploi. Toutefois ce sont les arrondissements périphériques parisiens qui constituent l'anneau de plus forte densité de population et de travail et non les arrondissements plus centraux.

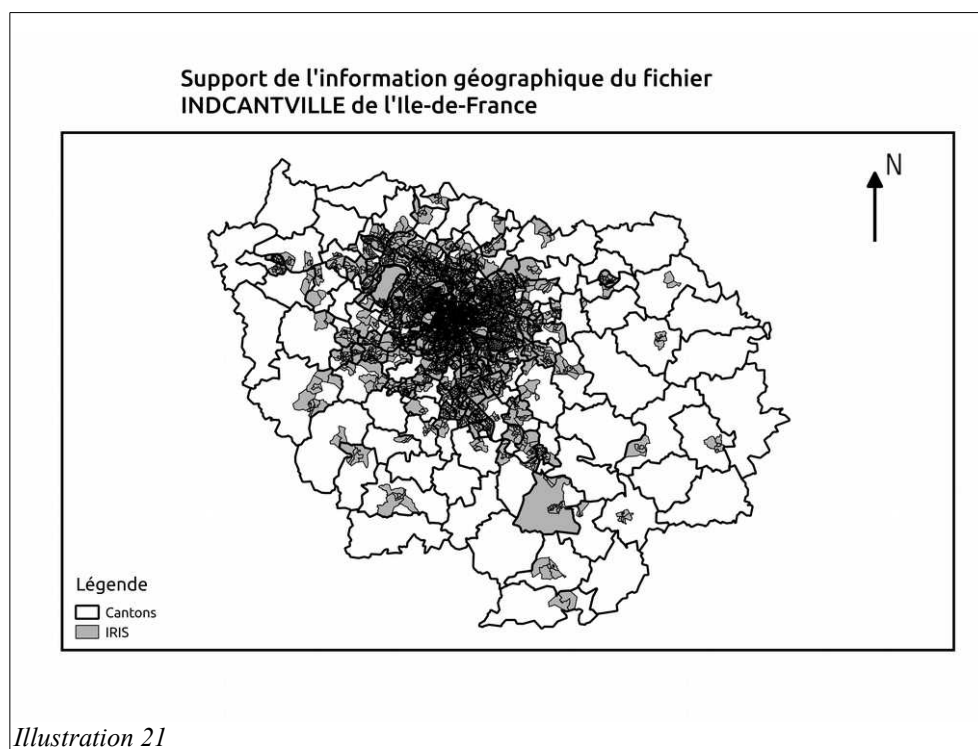
Précisons avant toute chose que ce type de carte présente des biais inhérents à l'approche adoptée pour la représentation des populations, du fait qu'elle viole des règles importantes de sémiologie graphique (Bertin, 1967). Ici, une règle élémentaire n'a pas été respectée : la cartographie d'une variable dénombrant des effectifs doit se faire avec des cercles proportionnels. Mais une telle carte serait illisible. C'est pourquoi les effectifs sont représentés à l'aide d'une carte choroplète. Il faut donc faire attention aux variations de taille des zones du support. Est-ce que les arrondissements parisiens externes sont plus forts en effectif parce qu'ils sont de plus grande taille ou bien sont-ils plus denses ? C'est bien en descendant d'échelle ou en disposant d'un support régulier que nous allons pouvoir le savoir.

3.2.2.2 Le fichier INDCANTVILLE

Parmi les jeux de données utilisés dans la méthode de changement de support le fichier détail INDCANTVILLE est utilisé. Ce fichier décrit l'ensemble de la population enquêtée qu'elle soit active occupé ou non. À l'instar du fichier MOBPRO le fichier INDCANTVILLE fait partie de l'exploitation complémentaire et renseigne sur la composition des ménages, les secteurs d'activité et sur d'autres caractéristiques socio-démographiques de manière plus précise que le fichier MOBPRO. Rappelons que nous cherchons dans ce fichier des données de contrôle c'est-à-dire apportant une connaissance plus fine que la source de données à désagréger, qu'elle soit spatiale ou sociale. Ce fichier, malgré sa localisation à peine plus fine que MOBPRO, voire moins précise dans certains cas, sera une source de données de contrôle. Bien que les deux fichiers contiennent les mêmes actifs occupés, les observations ne sont pas décrites de la même manière et il n'existe pas d'identifiant permettant de retrouver les correspondances entre les deux fichiers. Il s'agit bien de sources de données différentes que nous allons apparier (chapitre 4.1). De plus le fichier INDCANTVILLE contient toutes les autres observations y compris les actifs non occupés et les inactifs. La méthode d'appariement des fichiers MOBPRO et INDCANTVILLE que nous avons conçue est présentée au chapitre 4.1.

Parmi les différences entre les deux fichiers retenons que les supports géographiques

décrivant les localisations ne sont pas les mêmes. Premièrement dans le fichier INDCANTVILLE il n'y a pas de description des lieux de travail. Deuxièmement la description des lieux de résidence est moins précise pour les habitants des communes de moins de 10 000 habitants. Dans ces cas nous ne connaissons plus les communes de résidence mais les cantons. Mais pour les grandes communes nous gagnons en revanche en précision sur les lieux de résidence. Pour ces types d'espaces, nous allons passer de la connaissance de la commune aux Iris de résidences. Comme le montre la carte du support Iris/pseudo-canton (illustration 21) nous gagnons en précisions pour les zones à forts effectifs. Précisions que l'IGN ne diffuse pas ce support géographique en une seule couche d'information géographique. Cette carte est une estimation des limites de ce support. Rappelons que les IRIS et les cantons évoluent dans le temps.



Aussi avec l'utilisation du fichier INDCANTVILLE comme données de contrôle nous allons gagner en description des profils socio-économiques. Nous avons notamment un secteur d'activité décrit en 17 classes et non en 5 comme dans le fichier Mobpro. Cette plus grande finesse de description des activités de travail des individus va nous être grandement utile pour la ventilation des lieux d'emploi. L'utilisation du fichier INDCANTVILLE nécessitera d'apparier les fichiers détail du recensement afin de disposer des descriptions les plus fines de

chacun des deux sources d'information pour le mettre en relation avec Mobpro (chapitre 4.1). Nous disposerons à l'issue de ce processus d'un fichier dit apparié informant sur les lieux de résidence sur un support IRIS-Communes pour les observations appartenant à des ménages avec au moins un actif occupé, et indiquant les lieux d'emploi à la commune pour l'ensemble des actifs occupés.

Insistons sur le fait que parmi les données internes à l'Insee, le fichier confidentiel du recensement regroupe l'ensemble des variables diffusées dans les deux fichiers détail INDCANTVILLE et MOBPRO. Si le travail de désagrégation est réalisé à partir du fichier confidentiel alors l'utilisation d'INDCANTVILLE et l'étape d'appariement n'est plus nécessaire.

3.2.2.3 Les carroyages de population

La désagrégation des lieux de résidence est constituée d'une seconde étape, en continuant notre descente d'échelle. Les lieux de résidence du fichier apparié (que nous nommons INDCANTVILLE-MOBPRO) sont transférés du support IRIS-Communes au support carroyé de 200 m de côté diffusé par l'Insee en janvier 2013 (encadré 1). Ce support a été diffusé par l'Insee conjointement au support carroyé de 1 km répondant aux recommandations INSPIRE. Ce carroyage de 200 m constituera une étape intermédiaire avant la ré-agrégation au carroyage INSPIRE de 1 km.

A l'aide du support Insee de 200m nous disposons des effectifs de ménages et d'individus. Les carreaux ne contenant pas de population ne sont pas présents, le carroyage diffusé est donc discontinu. Les données sont issues des fichiers de la taxe d'habitation et des revenus fiscaux localisés ; il s'agit donc d'une source relativement distincte du recensement rénové.

« Les revenus fiscaux localisés (RFL) sont établis à partir des fichiers exhaustifs des déclarations de revenus des personnes physiques et de la taxe d'habitation fournis à l'INSEE par la Direction Générale des Impôts. L'INSEE procède au rapprochement de ces deux fichiers afin d'estimer le revenu fiscal à des niveaux géographiques finement localisés, tout en préservant la confidentialité des données.

Les données fournies ici proviennent d'une exploitation spécifique des fichiers fiscaux visant à attribuer à chaque ménage une position géographique précise. Les données carroyées de population sont cohérentes avec les données diffusées par la source

« revenus fiscaux localisés », mais elles ne peuvent fournir qu'une vision partielle de la population, rappelons que par rapport au recensement :

- les populations non référencées à la taxe d'habitation ne sont pas présentes : personnes sans domicile, collectivités ;
- les étudiants sont généralement localisés au domicile de leurs parents ;
- La notion de résidence principale peut être sensiblement différente.

Les données des carreaux de 200 m de côté sont dans les mêmes systèmes de projection que les données à 1km. Les carreaux de 200m de côté correspondent à une subdivision interne en 25 éléments des carreaux de 1km de côté. » (Insee, Documentation générale sur le carroyage⁷²)

Le carroyage de 200 m va servir de support de contrôle (illustration 23). C'est donc à l'aide du carroyage de 200 m que nous allons descendre plus bas que la résolution de 1 km, au risque de perdre la représentativité statistique sur les carreaux de 200 m. Cependant, la ré-agrégation sur les carreaux INSPIRE de 1 km permettra de retrouver cette représentativité. L'opération de ré-agrégation sera facilitée puisque ces deux supports entretiennent une relation hiérarchique : ils sont emboîtés.

Les sources d'information pour construire ces carroyages sont bien différentes des enquêtes annuelles du recensement. Les données de population carroyées ont introduit la structure des données des revenus fiscaux déclarés par les ménages. L'utilisation des données de population carroyées en tant que données de contrôle permet d'ajouter cette information externe au recensement. L'utilisation des données confidentielles (CASD) n'utiliserait que cette source d'information supplémentaire.

Le tableau suivant décompte les effectifs des données de contrôle utilisées pour ventiler les lieux de résidences.

72 Insee. Documentation générale sur le carroyage [en ligne] (Consultée le 12/04/2014)
http://www.insee.fr/fr/themes/detail.asp?reg_id=0&ref_id=donnees-carroyees&page=donnees-detaillees/donnees-carroyees/donnees_carroyees_doc.htm

Tableau 7 : Décompte des populations en Île-de-France				
		INDCANTVILLE (1er janv 2012)	MOBPRO (1er janv 2012)	Carroyage (2009)
Population totale	Total	11 898 049 (4 301 151 observations)	N.A	N.A
	En ménages ordinaires	11 661 516 (4 231 656 observations)	N.A	11 368 420
	Hors ménage ordinaires	236 532,6 (69 495 observations)	N.A	N.A
Effectif d'actifs occupés de + 15 ans révolus	Total	5 398 596 (1 955 156 observations)	5 398 596 (1 955 156 observations)	N.A
	En ménages ordinaires	5 348 251 (1 939 160 observations)	5 348 251 (1 939 160 observations)	N.A
	Hors ménage ordinaires	50 344 (15 996 observations)	50 344 (15 996 observations)	N.A
Source : RP 2012 exploitation complémentaire, Revenu Fiscaux localisés 2009				

Les données carroyées à 200m de l'Insee nous permettent de disposer d'une information sur les effectifs de population plus fine que les communes de résidence (illustration 22). Toutefois nous ne connaissons pas les effectifs des actifs occupés. C'est pourquoi le processus de changement de support de cette thèse va affecter de nouvelles localisations résidentielles (sur des carreaux de 200 m) à l'ensemble de la population du fichier INDCANTVILLE et non pas seulement aux actifs occupés. Ainsi, pour les lieux de résidence le travail de désagrégation est effectué pour les observations de l'exploitation complémentaire (ménages avec au moins un actif occupé).

Une fois que nous aurons transféré les localisations des observations de l'exploitation complémentaire au lieu de résidence et au lieu d'emploi pour les actifs occupés nous allons ré-agrégérer les localisations sur le carroyage INSPIRE de 1 km diffusé en janvier 2013 par l'Insee⁷³ (illustration 23). Le choix d'utiliser le support de 1 km comme support final est conforme à la directive européenne INSPIRE (chapitre 2). L'idée est de produire une source

⁷³ Précisons que les supports des données carroyées ne recouvrent pas l'ensemble de l'espace géographique. Seul les carreaux contenant une population résidente sont retenus. Des lieux d'emploi peuvent ne pas être contenu dans les carreaux. Nous avons ajouté quelques carreaux au carroyage originalement diffusé par l'Insee, notamment dans les zones aéroportuaires.

d'information dont les mailles du support de diffusion permettraient aux différents utilisateurs de ré-agréger à façon le support régulier.

De plus, pour revenir sur le choix cartographique d'une carte choroplète afin de représenter des effectifs de population un support régulier permet d'outre-passer la règle de sémiologie graphique nous imposant d'effectuer des cartes par cercles proportionnelles. Les cartes d'effectif de population sur un support régulier sont des cartes de densités de population.

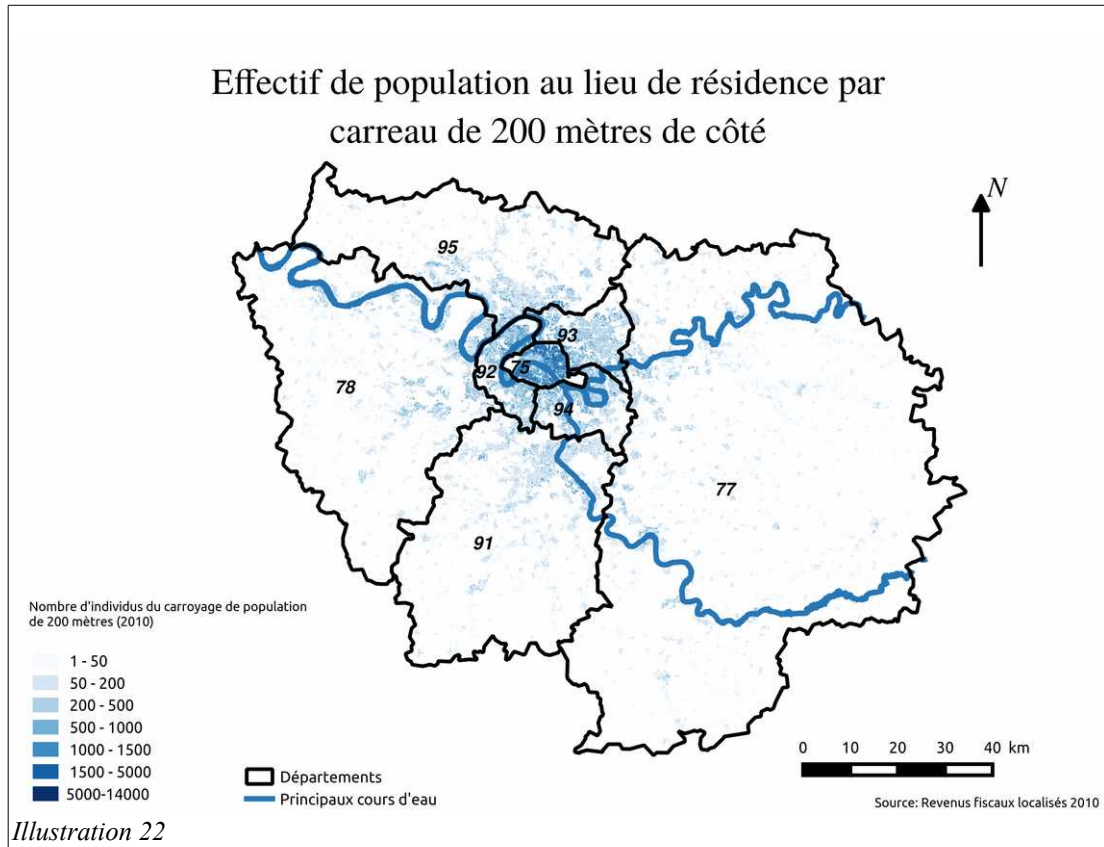


Illustration 22

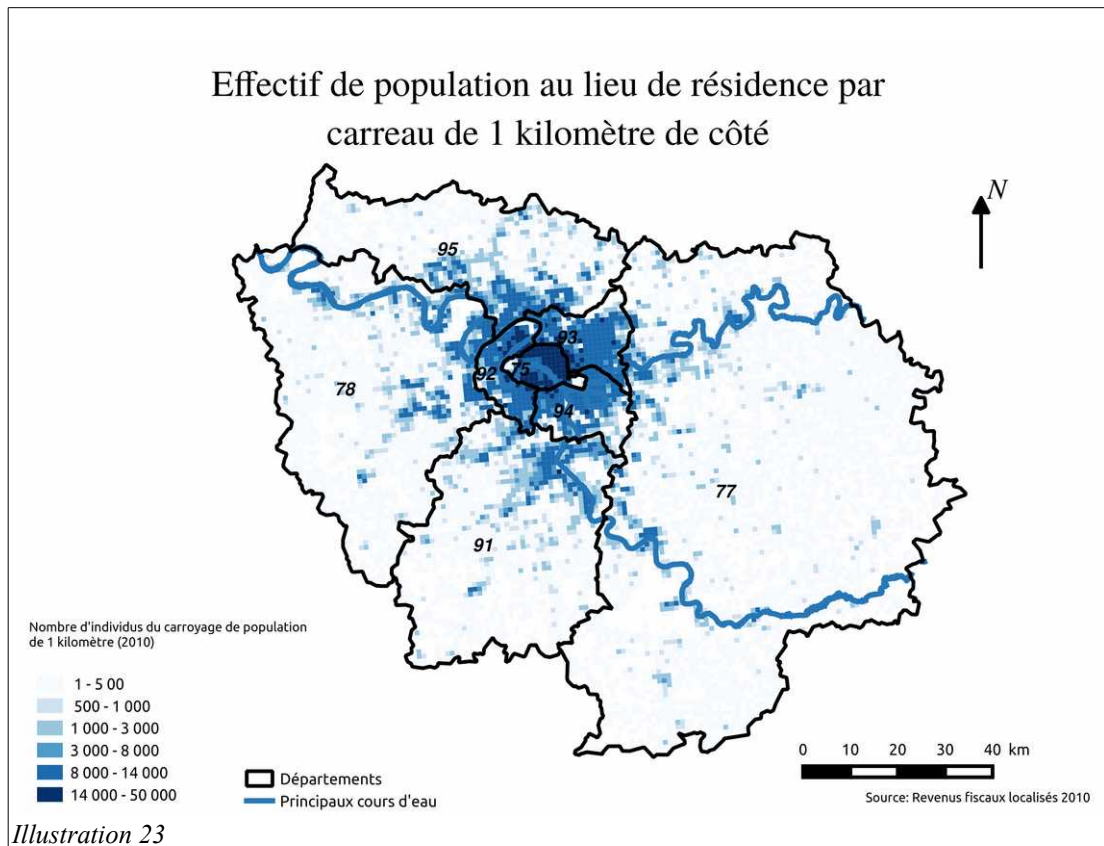


Illustration 23

3.2.3 Les données de contrôle pour ventiler les lieux de travail

Pour ventiler les observations d'actifs occupés au lieu de travail il nous faut un jeu de données renseignant sur la localisation des lieux d'emploi. Mais les bases de données sur les effectifs d'emplois librement diffusés par l'Insee sont le Clap (Connaissance de l'appareil productif) ou encore le fichier détail nommé Activité professionnelle des individus (tableau 1 et 3).

Ces fichiers offrent une information sur les effectifs d'établissements, le nombre de postes de salariés et les activités économiques. Toutefois ces fichiers n'offrent pas de localisation infra-communale. Nous ne pouvons pas utiliser ces données pour notre travail. Nous avons donc utilisé le répertoire national des entreprises et des établissements (Insee. Connaître et comprendre le répertoire SIRENE⁷⁴).

-« Ce système informatisé du répertoire national des entreprises et des établissements (SIRENE) dont la gestion a été confiée à l'Insee enregistre l'état civil de toutes les entreprises et leurs établissements, quelle que soit leur forme juridique et quel que soit leur secteur d'activité, situés en métropole, dans les Dom (Guadeloupe, Guyane, Martinique, La Réunion et Mayotte) et à Saint-Pierre et Miquelon. Les entreprises étrangères qui ont une représentation ou une activité en France y sont également répertoriées. » (Insee. Définition de SIRENE⁷⁵)

A ce jour, ce fichier est payant et coûte environ 70 000 euros pour la France entière. C'était donc une condition qui limitait la reproductibilité et l'actualisation de notre travail de thèse. Soulignons toutefois qu'à l'ère de la libéralisation de la donnée le secrétaire d'État au numérique Axelle Lemaire a annoncé le 12 janvier 2016 l'ouverture gratuite du répertoire SIRENE pour 2017 (Ministère des finances. Open data⁷⁶).

Les variables de ce répertoire utiles à la ventilation spatiale sont les variables décrivant le secteur d'activité des établissements économiques (APET700) et la description des effectifs

74 INSEE, Connaître et comprendre le répertoire SIRENE [en ligne] (Consultée le 04/04/2014) < <http://www.insee.fr/fr/bases-de-donnees/default.asp?page=sirene/institutionnel.htm> >

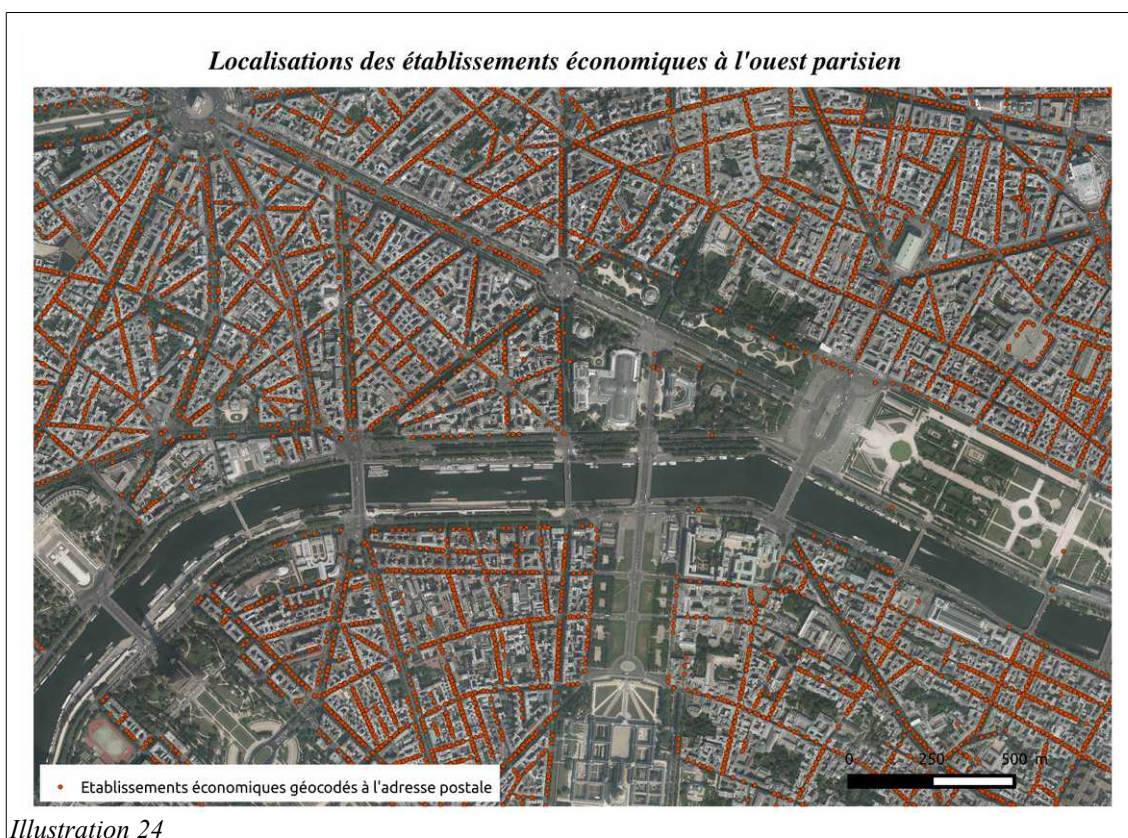
75 Insee- définition de SIRENE [en ligne] (Consultée le 14/05/2016) < <http://www.insee.fr/fr/methodes/default.asp?page=definitions/sys-inf-rep-nat-ent-etab.htm> >

76 Finance.gouv.fr, open data [en ligne] consulté le 11/06/2016 < <http://proxypubminefi.diffusion.finances.gouv.fr/pub/document/18/20339.pdf> >

de salariés des établissements décrite à l'aide de la variable EFFETCENT. Cependant, ce répertoire n'est pas une base de données géographique. Pour pouvoir utiliser ce répertoire nous devons procéder avant à une opération de géo-codage. En effet, le répertoire SIRENE localise les entreprises à l'aide des adresses postales. Il n'y a pas de coordonnées géographiques permettant de géo-localiser les données.

Un algorithme retrouvant les coordonnées géographiques des adresses postales du répertoire SIRENE à l'aide de la couche d'information géographique de l'Institut National Géographique (IGN) a été mis au point pour l'occasion. Ce programme perd 37 205 entreprises (3,6 % de pertes), ce qui semble être une performance pouvant être amélioré.

Une fois le processus de géocodage effectué nous obtenons un jeu de données sur les établissements économiques avec une précision à l'adresse postale et donc infra-communale. La carte suivant montre qu'a l'aide du répertoire SIRENE géocodé à l'adresse nous retrouvons une structure des lieux d'emploi en suivant le linéaire de la voirie (illustration 24).



Pour chaque établissement nous disposons d'une information sur les classes d'effectifs de travailleurs et d'une information sur le secteur d'activité en à peu près 700 classes. Il s'agit de la nomenclature d'emploi la plus fine possible. Cette information va permettre de choisir les lieux d'emploi aux observations du recensement en respectant le secteur d'activité en 17 classes que l'appariement entre le fichier MOBPRO et INDCANTVILLE nous permet d'obtenir (le fichier MOBPRO initial indiquant seulement le secteur d'activité en 5 classes).

La carte ci-dessous (illustration 25) montre pour chaque carreau de 1 km la modalité du secteur d'activité en 17 classes la plus élevée du répertoire SIRENE géo-codé. Nous reviendrons sur ce travail d'agrégation des établissements au carroyage de 1km pour comparer les effectifs de salariées renseigné dans SIRENE avec les résultats de la désagrégation (chapitre 4.3).

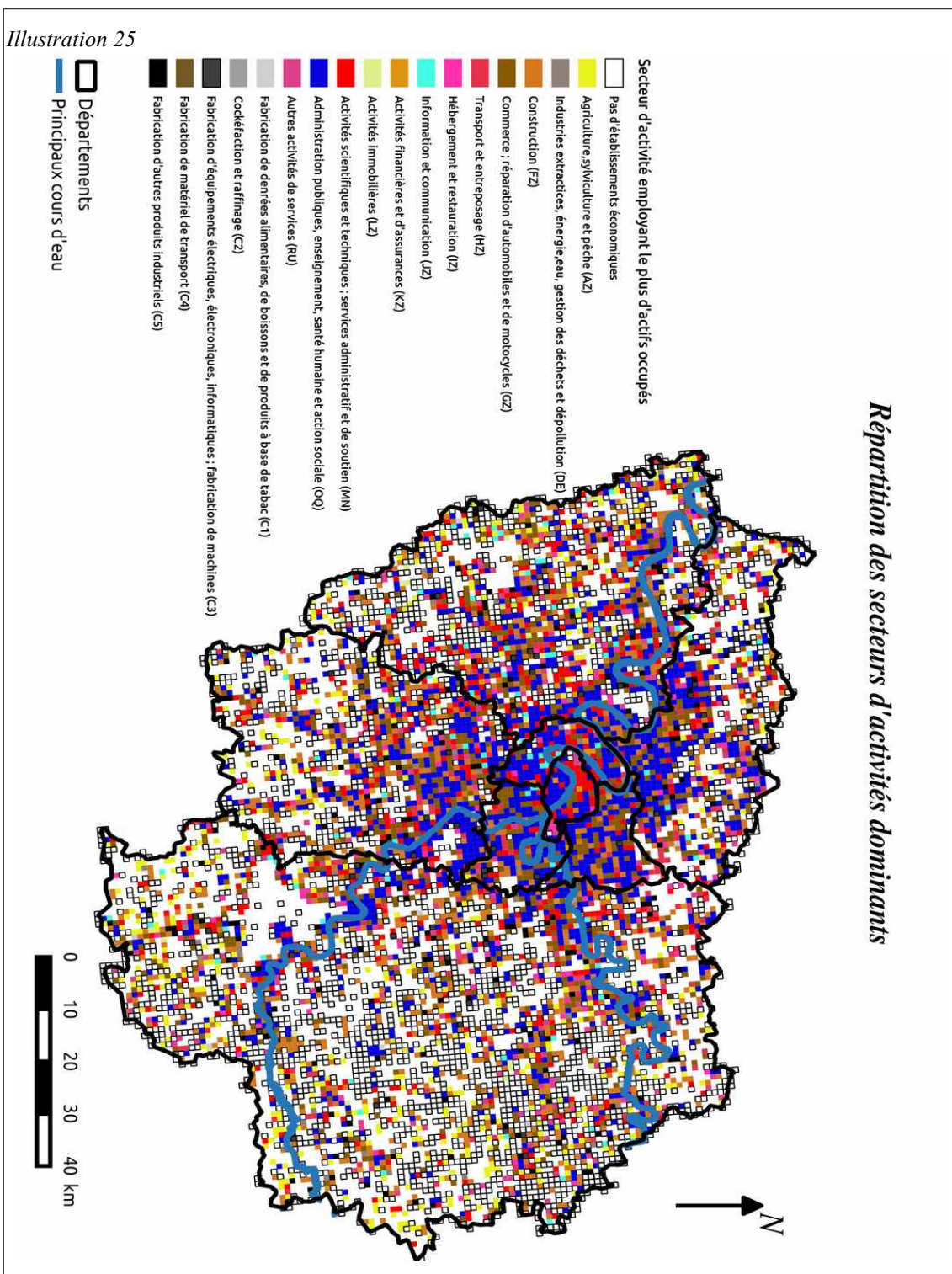


Illustration 25

3.2.4 Les descriptions des trajets des déplacements comme variables de contrôle

Nous venons de voir quelles données nous allons utiliser pour constituer un panel de localisations résidentielles et d'emplois plausibles pour chaque actif. Appariement des couples origines-destinations domicile-travail consiste à sélectionner un couple parmi les croisements possibles de ces localisations ponctuelles. Le problème est que l'on ne connaît pas l'agencement de ces couples origines-destinations à construire : il n'existe aucune donnée nous permettant d'acquérir cette information. La difficulté est donc que l'on ne connaît pas beaucoup de choses sur l'appariement des couples origines-destinations à l'intérieur des communes. Toutefois, selon les modes de transport nous allons pouvoir introduire des contraintes de distances à parcourir.

Nous pouvons utiliser un modèle gravitaire non pas sur la structure globale des déplacements mais en l'utilisant pour localiser les individus selon un gradient de proximité aux points d'entrées et de sorties des systèmes de transport utilisés par les individus. Il s'agit de veiller à ce que les individus se déplaçant en transport en commun ne soient pas trop éloignées des stations de transport par exemple. Il s'agit aussi de veiller à ce que ceux qui déclarent marcher pour aller travailler effectuent des distances raisonnables.

Dans ce but, nous utilisons comme variable de contrôle des distributions sur les longueurs à parcourir selon les différents modes de transport. Prenons les cas de la voiture particulière (VP), des transports en commun (TC), puis des modes doux et actifs.

Les véhicules personnels

L'idée est de localiser les observations utilisant un véhicule personnel à proximité des infrastructures routières. Pour ce faire nous pourrions construire une variable de contrôle surfacique de densité de chevelu fin des infrastructures routières. Cela pourrait nous indiquer une proportion d'observations à placer en fonction de la densité du réseau. Toutefois cela pose la question de la redondance des informations. En effet, dans les modèles de transport on utilise bien plus le taux de motorisation des ménages (Dupuy, 1975, p.25). Cette information est disponible dans les fichiers INDCANTVILLE et MOBPRO. Pour chaque actif occupé nous disposons du mode de transport fréquemment utilisé pour aller travailler. Il est aisé de construire un taux de motorisation des ménages pour les zones (IRIS pour les communes de

plus de 10 000habitants et communes pour les petites communes⁷⁷. Dans le cas d'application de cette thèse nous n'allons pas nous servir de la distances aux infrastructures de transport routier, nous nous servirons directement du taux de motorisation.

Les transports en commun

Comme nous disposons du mode de transport habituel dans le fichier MOBPRO nous pouvons placer les individus les utilisant à proximité des stations de transport en commun. Pour ce faire nous utilisons une règle de probabilité décroissante : plus on s'éloigne des stations de transport, moins un usager aura de chances de se localiser loin des stations. Pour cela, il nous faut disposer d'une information externe sur les aires de rabattement autour des stations.

Pour étudier les aires de rabattement autour des stations de transport en commun nous utilisons l'Enquête Globale Transport d'Île-de-France 2010 pour créer une variable de distance contrôlant les distances d'accès.

Dans le cas d'application proposé dans la suite de cette thèse, la distance sera utilisée comme variable de contrôle pour le choix des localisations des lieux de travail et non des lieux de résidence, dont le processus est encore plus complexe, et résulte d'arbitrages entre les membres d'un même ménage. Nous considérons la distance aux stations de transport seulement pour la ventilation des lieux d'emploi.

La distribution des distances entre les stations de sorties des réseaux de transport en commun et les lieux de travail. La distribution issue de l'enquête globale transport va servir de loi de probabilité lors des affectations des lieux de travail . Nous utiliserons la distribution suivante comme loi de probabilité (illustration 26).

⁷⁷ C'est bien à l'aide de l'appariement des fichiers de détails que nous allons obtenir les communes de résidence des habitants des communes de moins de 10 000 habitants.

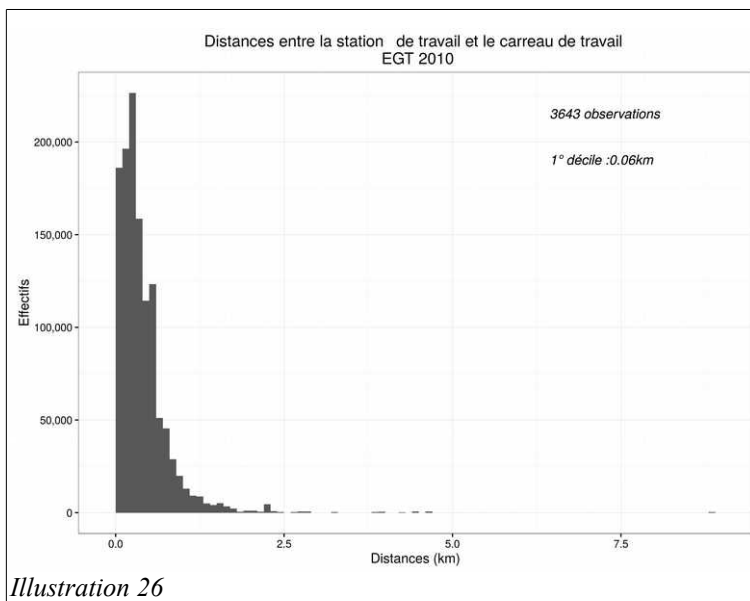


Illustration 26

La marche à pied

Pour contrôler les distances effectuées en marche à pied, nous utiliserons la distribution des distances de l'enquête globale transport (illustration 27).

Nous pourrions utiliser cette méthode pour d'autres modes de transport, notamment pour les déplacements effectués en vélo. Toutefois le fichier MOBPRO ne renseigne actuellement pas sur ce mode.

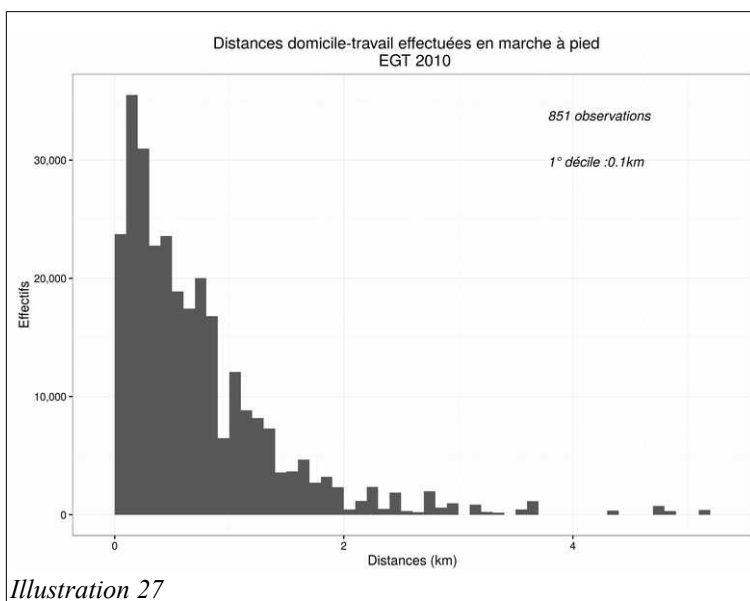


Illustration 27

les sans déplacements

A la question demandant le mode de transport le plus souvent utilisé pour aller travailler le questionnaire du recensement comporte une modalité de réponse nommé «Pas de déplacement» (Insee, bulletin du recensement 2016⁷⁸). Cette modalité permet en principe de considérer les actifs travaillant à domicile. Mais en examinant les actifs de MOBPRO déclarant ne pas utiliser de mode de transport, nous remarquons que les communes de résidence et d'emploi ne sont pas toujours identiques et parfois même elles sont très éloignées. Nous sommes là devant un problème de qualité des données que nous allons devoir laisser de côté, puisque nous ne disposons d'aucune information auxiliaire permettant de traiter la question. Nous ne pouvons pas savoir s'il y a bien un déplacement réalisé ou si ces individus ne se rendant pas à l'adresse de leur employeur pour travailler. La question posée dans le questionnaire du recensement est inadaptée pour capter des mobilités irrégulières, des lieux d'emploi changeant, etc. Ces actifs « problématiques » sont 72 738 dans le fichier MOBPRO de 2012, ils représentent 205 850 actifs occupés soit 3,7% de la population active occupée. Dans le processus de changement de support nous considérerons ces observations comme utilisant un véhicule personnel. Ceci pourra éventuellement amener à une surestimation des distances parcourues par l'ensemble de la population.

78 Insee, bulletin du recensement 2016 [en ligne] (Consultée le 15/06/2016) <<http://www.le-recensement-et-moi.fr/rpetmoi/pdf/Bulletin-individuel.pdf>>

Conclusion du chapitre 3

Les hypothèses de construction de la désagrégation spatiale de donnée de mobilité sont de deux ordres. Premièrement, il faut disposer de suffisamment d'observations pour être en mesure de les répartir dans l'ensemble de l'espace géographique. Deuxièmement, nous devons choisir les données de contrôle qui nous apportent une information sur la répartition spatiale des origines et des destinations des localisations à désagréger.

L'étape de clonage des observations du recensement est conceptuelle et utilisée avec parcimonie. Elle est réalisée au plus simple et peut être raffinée. Nous nous servons des poids de sondage statistique comme coefficient multiplicateur des observations. Et, lors de la ventilation sous contrainte nous tentons de laisser groupés au maximum les clones statistiques, c'est-à-dire que nous leur affectons les mêmes localisations domicile-travail. Nous procéderons à la séparation des agrégats d'observations seulement pour respecter la répartition des effectifs qu'indiquent les données de contrôle. Dans ce cas les clones auront des localisations différentes.

La répartition spatiale des observations ventilées est donnée par les variables de contrôle. Les modalités des variables prises en compte vont contraindre la répartition spatiale des observations. Les modalités des variables non prises en compte dans la construction des quantités à répartir pour chaque modalité des variables de localisation sont mises à contribution mais de moindre manière. Si nous ne construisons pas les couples de coefficients de certaines modalités, celles-ci seront réparties selon les coefficients des autres croisement de modalités des variables de contrôle.

Dans le cas de données bi-localisées, la distribution des variables socio-économiques non prises en compte dans la ventilation sous contraintes nous intéresse tout particulièrement. Parmi les variables qui ne sont pas contrôlées il se trouve forcément les variables décrivant les localisations autres que celles que les données de contrôle permettent d'approcher. Dans notre cas d'application, lorsque on ventile les lieux de résidence, nous localisons dans le même temps des variables décrivant les lieux de travail (commune d'emploi et secteurs d'activité des actifs occupés). C'est avec ce lien que nous supposons considérer d'un seul tenant les origines, les destinations et l'appariement des couples origines-destinations domicile-travail. Le fait de localiser au lieu de résidence les variables décrivant les emplois

puis de procéder dans un deuxième temps à la ventilation spatiale des lieux d'emploi permet de traiter quelque peu le cas de l'interdépendance des données bi-localisées.

Nous allons donc effectuer des ventilations sous contraintes des lieux de résidence. Puis des ventilations sous contraintes des lieux d'emploi. Le tableau suivant (Tableau 9) récapitule les données de contrôle utilisées dans notre cas d'application à l'Île-de-France.

Tableau 9 : Les données de contrôle utilisées dans le processus de désagrégation appliqué à l'Île-de-France	
Données	Date
Fichier de mobilité domicile-travail	Millésime 2012
Fichier Individus localisé aux canton-ville	Millésime 2012
Couche SIG IRIS-Communes	Millésime 2011
Carroyage 200m	2010
Carroyage 1km	2010
SIRENE	2010
PT Adresse	2010
Distance aux stations de travail (EGT)	2010
Couche SIG des stations de transport en commun (APUR)	2015
Distance domicile-travail en marche à pied	2010

D'autres données de contrôle peuvent être utilisées, pour ventiler plus finement les localisations résidentielles que sur le carroyage de 200m les couches d'information géographique du bâti de l'IGN peuvent être utilisées. Nous choisissons de ne pas les utiliser, nous supposons que la localisation résidentielle au carroyage de 200m est suffisante puisque nous ré-agrégerons les résultats des ventilations sous contraintes aux carreaux de 1km.

Nous avons désormais tous les éléments théoriques pour présenter le processus de désagrégation appliquée à la région Île-de-France.

Chapitre 4 : Le processus de désagrégation spatiale des localisations résidentielles et d'emplois appliqué à la région Île-de-France

Chapitre 4 : Le processus de désagrégation spatiale des localisations résidentielles et d'emplois appliqué à la région Île-de-France

Au chapitre précédent nous avons présenté les données de contrôle que nous allons utiliser. Nous présentons maintenant l'application concrète de la méthodologie de désagrégation spatiale. La méthodologie de désagrégation est constituée de deux étapes : le clonage des observations et les ventilations spatiales sous contraintes. L'étape de clonage est conceptuel et elle est imbriquée dans les étapes de ventilations sous contraintes.

Ce chapitre montre comment utiliser efficacement les différentes données de contrôle pour ventiler les données du recensement le mieux possible, en conservant une représentativité statistique. Nous allons opérer une succession de ventilations sous contrainte pour descendre progressivement en résolution spatiale. Le processus de désagrégation est constitué de trois étapes (appariement des fichiers de détails, ventilation sous contraintes de lieux de résidence et ventilation sous contraintes des lieux d'emploi), (illustration 28).

Une première section expose la méthode d'appariement des fichiers détail du recensement. L'appariement permet de descendre en échelle différemment selon que les individus représentés par les observations habitent dans des communes de moins de 10 000 habitants ou de plus de 10 000 habitants. Cette étape permet également de localiser les variables décrivant les emplois exercés au lieu de résidence (4.1).

La deuxième étape du processus de désagrégation consiste à affiner encore plus la description des localisations résidentielles. L'appariement de la première étape produit des localisations sur le support IRIS-Communes, cette seconde étape ventile les observations des ménages sur le carroyage de 200m de l'Insee (4.2).

La dernière étape du processus consiste à ventiler dans leur commune de travail les observations représentant les actifs occupés (4.3).

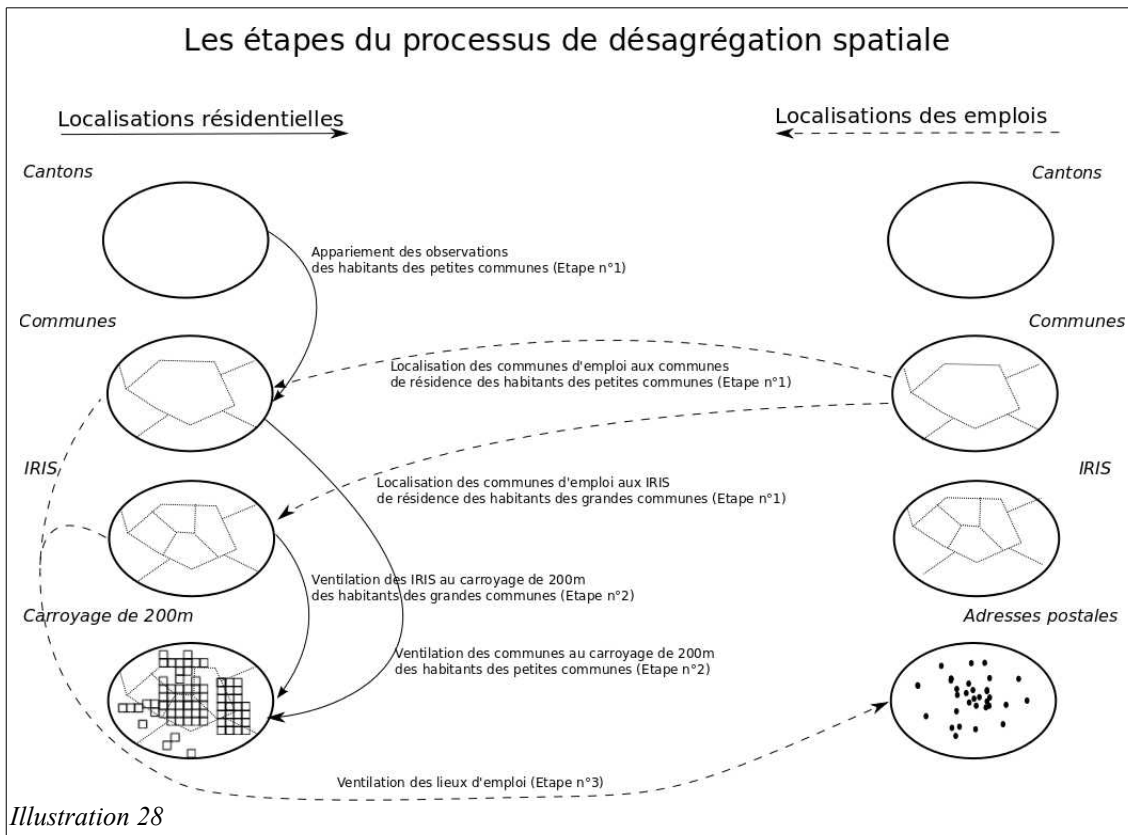


Illustration 28

Pour chacune de ces étapes nous vérifions si les observations ont été ventilées et nous nous assurons que la ventilation a bien réparti les observations dans tout l'espace de la région Île-de-France.

4.1 Appariement des fichiers détails Individus-CANTVILLE et MOBPRO

Nous avons vu dans le chapitre 3 que la désagrégation spatiale du fichier MOBPRO bénéficie de l'utilisation du plus grand niveau de détail socio-économique apporté par le fichier INDCANTVILLE du recensement. Nous allons présenter le processus que nous avons mis au point dans ce but, et expliciter les gains que l'on obtient en appariant ces deux fichiers. Rappelons que le fichier confidentiel contient l'ensemble des variables des fichiers INDCANTVILLE et MOBPRO et qu'en partant de ce fichier l'étape d'appariement présentée dans cette section serait inutile.

Ces fichiers anonymisés décrivent les mêmes individus mais avec des variables différentes. Chaque bulletin du recensement donne lieu à une synthèse différente dans chacun des fichiers de détail : un sur les individus, un sur les logements, un sur la mobilité professionnelle, un sur la mobilité scolaire, un sur les migrations résidentielles (ce dernier étant indisponible temporairement à la suite d'un changement de question dans le questionnaire). Les fichiers de détail sont bien entendu anonymisés. Ils sont conçus pour que l'identification des répondants soit impossible.

Les différents fichiers détail anonymisés comportent des variables communes, et des variables spécifiques. L'objet du travail d'appariement est de se servir des variables identiques entre les fichiers pour enrichir chaque fichier des variables disponibles uniquement dans les autres fichiers. Nous obtenons ainsi une description plus complète des individus, et, nous le verrons, des ménages, qu'en utilisant les fichiers de manière individuelle.

Ici, nous allons traiter de deux fichiers construits à l'aide de l'exploitation complémentaire: le fichier INDCANTVILLE et le fichier MOBPRO. L'objet est d'obtenir l'information la plus fine disponible dans les deux fichiers.

Nous avons dans le fichier INDCANTVILLE tous les bulletins des grandes communes et 25 % des bulletins des petites communes. De plus, nous connaissons la localisation résidentielle des individus aux IRIS pour les grandes communes. En revanche, pour des raisons de secret statistique, l'Insee regroupe spatialement les individus des petites communes aux cantons. Ce fichier nous intéresse particulièrement, car il nous apporte trois sources d'information qui vont nous être utiles :

- la variable NUMMI qui nous permet d'identifier les individus appartenant au même ménage : on pourra ainsi reconstituer des ménages dans le fichier des mobilités professionnelles, qui n'en comporte pas ;
- la variable NA17 qui informe sur le secteur d'activité en 17 classes ;
- la variable CANTVILLE qui localise les individus aux IRIS de résidence pour les grandes communes et aux cantons pour les petites communes.

Quant au fichier MOBPRO, il décrit la mobilité domicile-travail des actifs occupés de plus de 15 ans. La spécificité de ce fichier est d'indiquer à la fois la commune de résidence et d'emploi de tous les actifs quelle que soit la taille de ces communes. En revanche, du fait de l'anonymisation certaines variables socio-économiques sont décrites de manières moins fines. Par exemple, le secteur d'activité des observations est décrit en 5 classes alors que le fichier INDCANTVILLE donne 17 classes.

Le travail d'appariement a pour but de disposer d'un fichier dit apparié offrant les variables décrites de la manière la plus fine qui soit dans ces deux fichiers détails du recensement. Pour chaque actif occupé de MOBPRO nous allons récupérer les variables « *commune de résidence* » et « *commune d'emploi* » et les écrire dans le fichier INDCANTVILLE. Puis, nous transmettons les variables de localisation résidentielle à l'ensemble des individus des ménages des actifs occupés. Remarquons que nous allons pouvoir récupérer et transmettre des variables de localisation plus fines seulement pour les ménages dont un individu est un actif occupé. Nous y reviendrons plus tard.

Cette opération d'appariement revient à transférer les données de MOBPRO du support de l'information géographique communal au support IRIS/pseudo-cantons. En récupérant les variables de localisations les plus fines de MOBPRO nous allons produire une information à une résolution IRIS-Communes. Ce travail nécessite de dissocier le cas des petites et grandes communes. Pour les habitants des petites communes du fichier INDCANTVILLE, l'appariement va permettre d'obtenir la commune de résidence au lieu du canton de résidence, et la commune de travail des actifs occupés de plus de 15 ans. Pour les grandes communes, l'appariement va nous permettre de renseigner la commune d'emploi dans le fichier INDCANTVILLE.

Le travail d'appariement consiste à construire des identifiants permettant de retrouver les

enregistrements similaires dans les deux fichiers (profils socio-économiques identiques). Le premier identifiant que nous allons nommer (IDENTIFIANT) permet de relier les observations des deux fichiers entre elles. Le deuxième identifiant que nous allons créer (ID_MEN) est l'identifiant unique pour chaque ménage d'Île-de-France et il provient de la variable NUMMI du fichier INDCANTVILLE⁷⁹. Il permettra de recomposer les ménages de MOBPRO, qui ne contient aucune information de cette nature. C'est à l'aide de ces identifiants que nous allons pouvoir identifier les observations d'un fichier à l'autre.

La construction de ces identifiants passe par la création d'une clef d'appariement pour chaque actif occupé des deux fichiers. Une fois que nous avons apparié les actifs occupés, nous transmettons les variables gagnées par ce travail d'appariement (commune de résidence) aux individus composant les ménages des actifs occupés de plus de 15 ans.

La première étape à réaliser est bien entendu de sélectionner les observations du fichier INDCANTVILLE correspondant à des personnes actives occupées de plus de 15 ans en âge révolu (variable AGER20). Par construction, nous obtenons exactement le même nombre d'observations d'actifs occupés dans les deux fichiers.

La deuxième étape d'appariement est de construire la clef d'appariement en concaténant les modalités de 24 variables (tableau 10).

⁷⁹ La variable NUMMI est une variable identifiant les ménages par Iris-cantons. Des observations de plusieurs Iris-Cantons ont les mêmes identifiants.

Tableau 10 : Les variables de la clef d'appariement																								
Variables/Codes géographiques de résidence		Variables socio-économiques et mode de transport																	Poids					
V	M	R	C	A	C	I	T	V	D	E	C	N	T	I	I	L	M	S	N	T	T	S	A	I
arriables	étropolitaine	égion	anton-village	rrondissement	ommune de résidence	L	RANS	OIT	IPL	MFL	S1	AS	P	Mi	ATC	PRM	OCO	EXE	PERR	YPL	YPMR	TOCD	AGEREVQ	POND
Modaliés	1	1	267	21	791	7	5	4	11	9	6	5	2	2	2	9	8	2	6	6	10	7	25	25290

Certaines de ces variables ne sont pas directement construites. Nous construisons la variable commune de résidence à partir de la variable renseignant l'IRIS qui dans sa codification donne le code Insee de la commune. Ainsi, nous connaissons la commune de résidence dans les deux fichiers pour les communes de plus de 10 000 habitants. Et, pour les petites communes nous mettons la valeur « ZZZZZ ». Nous obtenons alors 791 modalités pour décrire les communes de résidence sur les 1 281 communes d'Île-de-France.

Aussi, il est également nécessaire de construire la variable renseignant les cantons de résidence pour le fichier MOBPRO. Nous nous servons des tables des codes géographiques diffusées par l'Insee.

Enfin, nous construisons la variable renseignant le secteur d'activité en 5 postes dans le fichier INDCANTVILLE à partir du secteur d'activité en 17 classes, en regroupant des modalités

La troisième étape d'appariement est de trier par ordre alphanumérique les fichiers selon les clefs d'appariement. Nous supposons alors que la ligne N du fichier INDCANTVILLE décrit le même actif que dans la ligne N du fichier MOBPRO. Toutefois, lorsque plusieurs clefs d'appariement sont identiques nous ne pouvons pas discriminer la population : rappelons que les lignes décrivent des ménages statistiques et non des ménages réels, et que la population complète est reconstruite à l'aide de poids. Du fait de ces cas d'indétermination, il est possible

que nous faisons de mauvais appariements. Nous quantifions d'abord les erreurs d'appariement éventuels pour les petites et les grandes communes. Puis, nous évaluons la méthodologie à l'aide des données confidentielles de l'Insee auxquelles nous avons eu accès. Nous effectuons le même travail d'appariement et comptons le nombre d'erreurs produites.

4.1.1 Appariement des observations de ménages actifs occupés des communes de moins de 10 000 habitants

Lorsque plusieurs clefs sont identiques il peut arriver d'apparier un actif CANTVILLE avec un actif MOBPRO ne correspondant pas dans la réalité aux mêmes individus (tableau 11). Ainsi, nous pouvons intervertir des actifs dans des communes auxquelles ils n'appartiennent pas dans la réalité. Toute la question est de savoir dans quels cas cela peut être bloquant pour notre objectif de désagrégation spatiale de la mobilité, et dans quels cas cela ne l'est pas. Nous souhaitons naturellement établir les correspondances entre fichiers de la manière la moins fallacieuse possible lorsque nous sommes en présence de plusieurs clefs d'appariement identiques pour des ménages distincts (nous disons que nous sommes en présence de clefs multiples).

Tableau 11 : Exemples de cas d'appariements pour les petites communes										
INDIVIDUS CANTVILLE					ACTIFS MOBPRO					
	IDENTIFIANT	ID_ MEN	CANTON (cantville)	Commune résidence	Clefs	Clefs	Commune résidence	Commune d'emploi	ID_ MEN	IDENTIFIANT
Clefs uniques	A	1	Canton de Crécy-la-chapelle	?	clef1	clef1	Crécy-la-chapelle	Meaux	1	A
	B	2	Canton de Crécy-la-chapelle	?	clef2	clef2	Crécy-la-chapelle	Paris	2	B
Clefs Multiples	C	3	Canton de Crécy-la-chapelle	?	clef3	clef3	Esbly	Torcy	3	C
	D	4	Canton de Crécy-la-chapelle	?	clef3	clef3	Montry	Montry	4	D

Par exemple, si nous traitons d'une modalité de clefs d'appariements dites multiples et que nous sélectionnons les observations du fichier INDCANTVILLE et MOBPRO il est alors retourné plusieurs individus INDCANTVILLE et plusieurs actifs MOBPRO. Il faut choisir qui va avec qui. Selon les choix d'appariement nous allons pouvoir intervertir des individus INDCANTVILLE dans de mauvaises communes au sein des cantons (tableau 11).

Le tableau suivant dénombre les effectifs d'observations d'actifs occupés dont la clef est dite multiple et permet de quantifier les erreurs potentielles pour les petites communes d'Île-de-France (tableau 12). Précisions que dans ce décompte les individus hors-ménages sont comptabilisés.

Tableau 12 : Décompte des clefs d'appariements des communes de moins de 10 000 habitants

		Total	75	77	78	91	92	93	94	95
TOTAL	Total d'observation d'actifs occupés petites communes	147 486	0	61 244	28 326	36 479	167	587	2 231	18 452
Clefs simples	Nombre d'observations a clefs uniques	139 784	0	58 606	26 711	34 037	161	555	2 097	17 617
Clefs multiples	Nombre d' observations a clefs multiples	7 702	0	2 638	1 615	2 442	6	32	134	835
	% observations a clefs multiples	5,22 %	0	4,3 %	5,7 %	6,69 %	3,59 %	5,45 %	6,01 %	4,53 %
Doublons pouvant générer des erreurs dans la matrice	Nombre d' observations a clefs multiples avec des localisations différentes	1 419	0	451	323	493	2	2	26	122
	% observations a clefs multiples avec des localisations différentes	0,96 %	0	0,7 %	1,14 %	1,35 %	1,2 %	0,34 %	1,17 %	0,66 %
	Nombre d' observations a clefs multiples et des localisations différentes mais avec un membre du ménage actif à clef simple	1272	0	400	278	451	2	2	24	115
Cas que nous pouvons traiter	% d' observations avec des clefs multiples et des localisations différentes que nous pouvons traiter	89,64 %	0	88,69 %	86,01 %	91,48 %	100 %	100 %	92,31 %	94,26 %
Cas que nous ne traitons pas	Nombre d' observations d'actifs dont tous les autres actifs occupés du ménage ont une clef multiple	66	0	20	17	26	0	0	0	3
	Nombre des autres observations de ménages avec des clefs multiples et des localisations différentes	81	0	31	28	16	0	0	2	4
	% d' observations de ménages avec des clefs multiples et des localisations différentes	10,36 %	0 %	11,31 %	13,93 %	8,52 %	0 %	0 %	7,69 %	5,74 %

Source : Fichiers détail INDCANTVILLE et MOBPRO 2012

Les cas problématiques sont relativement peu nombreux : 5,2 % des observations décrites dans ces deux fichiers ont une clef multiple. Procédons à un examen des erreurs commises en cas de mauvais appariement. Il faut distinguer les actifs ayant des clefs multiples avec des localisations domicile-travail identiques, des actifs ayant des clefs multiples avec des localisations différentes.

Dans le premier cas nous allons potentiellement associer des actifs au sein de ménages auxquels ils n'appartiennent pas dans la réalité. Pour autant nous ne considérons pas cela comme une source d'erreur. Nous gardons les mêmes effectifs d'observations par communes

de résidence et d'emploi. Les modes de transport étant les mêmes, les parts modales par communes ne changent également pas. Les secteurs d'activité en 5 postes demeurent identiques. Les paires d'origines-destinations domicile-travail sont identiques selon les choix d'appariement. Nous considérons que ces cas ne sont pas problématiques ; ce sont simplement des indéterminations statistiques. Ils concernent 4,26 % des observations d'actifs occupés de plus de 15 ans.

En revanche lorsque les actifs occupés ont une clef multiple et que les localisations domicile-travail ne sont pas identiques alors des erreurs dans la matrice origine-destination peuvent être introduites. Comme dans la clef d'appariement nous connaissons seulement les cantons de résidence nous pouvons en cas de clefs multiples apparier un individu INDCANTVILLE avec un actif MOBPRO d'une autre commune au sein du canton. Techniquement, nous aurons deux communes de résidence renseignées pour ces appariements. Cette source d'erreur est identique pour les localisations d'emploi. En cas de clefs multiples il est possible d'intervertir les communes de travail. Nous disons que les actifs ont des localisations domicile et/ou de travail différentes. En moyenne, nous avons 0,96% de chance de se trouver dans de tel cas en Île-de-France. Heureusement, dans certains cas nous disposons d'heuristique nous permettant de réduire ces erreurs, que nous décrivons ci-après.

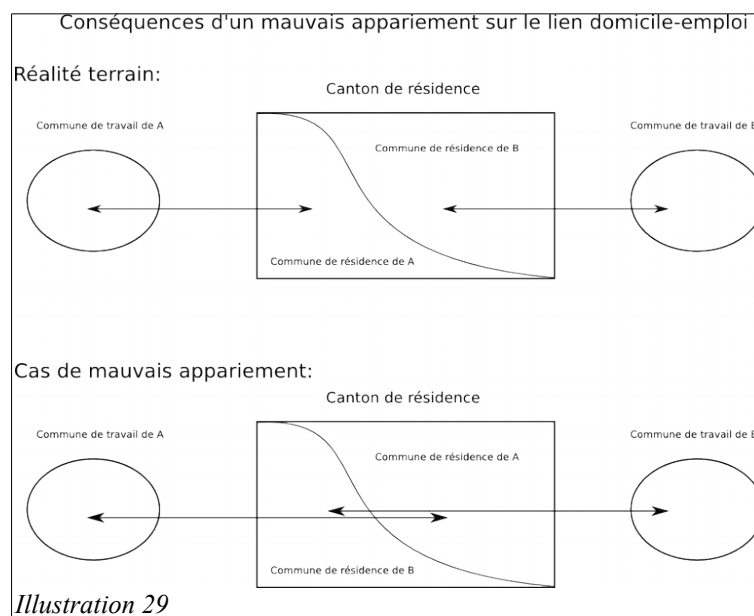
Cas que nous traitons :

Dans le cas d'actifs occupés avec des clefs multiples s'il y a un autre membre du ménage actifs occupé dont la clef n'est pas multiple nous pouvons mobiliser la connaissance de sa commune de résidence pour aider à effectuer les choix d'appariement des membres du ménage ayant une clef multiple. Parmi les 0,96 % d'observations avec une clef multiple ayant des localisations différentes nous pouvons traiter 90 % de ces cas.

Cas que nous traitons pas :

Quelques cas assez rares concernent des ménages où toutes les observations des actifs occupés ont des clefs multiples. Dans ces cas les choix d'appariement se font en s'assurant que les observations mises en ménage ont bien les mêmes communes de résidence. Toutefois, nous pouvons alors intervertir des ménages au sein des communes où ils ne résident pas dans la réalité. Certes, les membres d'un même ménage sont localisés dans la même commune mais nous ne savons pas si c'est la commune du ménage correspondant à la réalité. Dans ce cas nous gardons les mêmes effectifs d'observations par communes de résidence et d'emploi.

Les parts modales des modes de transport ne changent également pas. Les secteurs d'activité en 5 postes demeurent identiques. Mais les paires d'origines-destinations domicile-travail dans la matrice peuvent être modifiées. Les stratégies de localisation résidentielle en lien avec la distance au lieu d'emploi ne sont pas respectées. Il y aura des erreurs dans la matrice origine-destination. La distance métrique de ces erreurs est l'écart entre les lieux de résidence au sein des cantons (illustration 29).



Les ménages composés d'un seul actif occupé sont également sources d'erreur dans la matrice lorsqu'ils ont des clefs multiples. Nous pouvons intervertir les communes de résidence au sein des cantons. Et comme il n'y a qu'un seul actif occupé nous disposons pas d'heuristique permettant de procéder à un appariement veillant à respecter le lien domicile-travail.

En tout, les observations d'actifs occupés avec des clefs multiples ne pouvant pas être traitées sûrement concernent environ 10 % des 0,96 % d'observations d'actifs problématiques, soit environ 0,096 % des observations d'Île-de-France. En utilisant les fichiers détail diffusés publiquement nous ne pouvons pas savoir si nous effectuons les bons appariements. Avant de procéder à l'évaluation avec les données confidentielles nous quantifions les sources d'erreurs potentielles pour le cas des communes de 10 000 habitants et plus.

4.1.2 Appariement des observations de ménages actifs occupés des communes de 10 000 habitants et plus

Le cas des grandes communes est différent de celui des petites communes, puisque cette fois-ci la variable de localisation résidentielle la plus fine est dans le fichier INDCANTVILLE. Nous gardons la localisation résidentielle décrite dans le fichier INDCANTVILLE et ajoutons dans le fichier la variable décrivant la commune de travail.

Lorsque nous sélectionnons un individu INDCANTVILLE et que nous sélectionnons les actifs MOBPRO ayant la même clef d'appariement il nous faut faire des choix (tableau 13). Du point de vue du fichier INDCANTVILLE nous pouvons attribuer une mauvaise commune de travail à cet individu que nous sommes en train de traiter.

Le cas des clefs multiples est plus ou moins gênant selon que les actifs ont des communes de travail différentes ou non.

Tableau 13 : Exemples de cas d'appariements pour les grandes communes												
INDIVIDUS CANTVILLE							ACTIFS MOBPRO					
	Numéro de ligne du fichier	IDENTIFIANT	ID_MEN	CANTON	IRIS	Clef	Clef	COMMUNE RESIDENCE	COMMUNE DE TRAVAIL	ID_MEN	IDENTIFIANT	Numéro de ligne du fichier
Clefs simples	1	77023481 1 1 96 404 -- 76	7702 3706	7702	770530 201	M 11 77053 7702 ZZZZZ 1 1 0 11 15 5 OQ 1 2 1 1 32 1 1 2 11 21 95 2,763476254 08475	M 11 77053 7702 ZZZZZ 1 1 0 11 15 5 OQ 1 2 1 1 32 1 1 2 11 21 95 2,76347625408 475	77053	77053			1
	2	77023706 1 2 39 303 -- 729	7702 3706	7702	770530 102	M 11 77053 7702 ZZZZZ 1 1 0 13 16 6 GU 1 1 2 1 32 2 1 2 12 22 35 1.076304712 60665	M 11 77053 7702 ZZZZZ 1 1 0 13 16 6 GU 1 1 2 1 32 2 1 2 12 22 35 1.07630471260 665	77053	77053	?	?	2
Clefs multiples ayant les mêmes localisations	3	77994479 1 1 39 671 -- 865	7799 4479	7799	772880 602	M 11 77288 7799 ZZZZZ 4 5 1 18 16 3 GU 1 2 1 1 21 1 2 2 41 10 35 1.091713355 29003	M 11 77288 7799 ZZZZZ 4 5 1 18 16 3 GU 1 2 1 1 21 2 2 41 10 35 1.09171335529 003	77288	76497	?	?	3
	4	77992487 1 1 39 214 -- 917	7799 2487	7799	772880 602	M 11 77288 7799 ZZZZZ 4 5 1 18 16 3 GU 1 2 1 1 21 1 2 2 41 10 35 1.091713355 29003	M 11 77288 7799 ZZZZZ 4 5 1 18 16 3 GU 1 2 1 1 21 2 2 41 10 35 1.09171335529 003	77288	76497	?	?	4
Clefs multiples n'ayant pas les mêmes localisations de travail	5	77372930 2 2 32 551 -- 629	7737 1291	7737	772940 203	M 11 77294 7737 ZZZZZ 3 4 2 17 16 4 GU 1 2 1 2 22 2 3 1 41 10 30 1.044603820 93334	M 11 77294 7737 ZZZZZ 3 4 2 17 16 4 GU 1 2 1 2 22 2 3 1 41 10 30 1.04460382093 334	77294	75120	?	?	5
	6	77373114 2 2 35 599 -- 750	7737 3114	7737	772940 203	M 11 77294 7737 ZZZZZ 3 4 2 17 16 4 GU 1 2 1 2 22 2 3 1 41 10 30 1.044603820 93334	M 11 77294 7737 ZZZZZ 3 4 2 17 16 4 GU 1 2 1 2 22 2 3 1 41 10 30 1.04460382093 334	77294	93053	?	?	6

Le décompte des cas problématiques de clefs multiples ne se fait pas de la même manière que pour les petites communes (tableau 14). Nous différencions cette fois-ci les appariements gênants pour les localisations résidentielles et pour les localisations d'emplois. Contrairement au cas des petites communes, l'information géographique des localisations à garder est dans les deux fichiers. En effet, pour les petites communes nous prenons les deux localisations du fichier MOBPRO. Pour les grandes communes, nous gardons la localisation résidentielle

renseignée dans le fichier INDCANTVILLE (l'IRIS de résidence). Nous prenons seulement la commune d'emploi du fichier MOBPRO. Nous sommes alors obligés de faire la distinction entre les observations avec des clefs multiples ayant des lieux de résidence et/ou des lieux d'emploi différents. Ce sont deux indicateurs sur le potentiel d'erreur suivant l'utilisation de tel ou tel fichier. Comme nous utilisons le fichier INDCANTVILLE nous utilisons l'indicateur sur les emplois. Ainsi, les communes d'emploi des individus INDCANTVILLE localisés aux IRIS peuvent être interverties⁸⁰.

Tableau 14 : Décompte des clefs d'appariements des communes de 10 000 habitants et plus									
	IDF	75	77	78	91	92	93	94	95
Total des observations d'actifs occupés des grandes communes	1 807 669	425 807	128 772	190 473	148 014	288 068	238 195	22 7317	161 023
Nombre d'actifs a clef unique	1 474 963	377 495	120 271	17 4505	136 015	261 096	220 483	208 533	151 070
Nombre d'observations à clefs multiples	158 201	48 312	8 501	15 968	11 999	26 972	17 712	18 784	9 953
% d'observations à clefs multiples	8,75 %	11,35 %	6,6 %	8,38 %	8,11 %	9,36 %	7,44 %	8,26 %	6,18 %
Si utilisation INDCANTVILLE									
Nombre d'observations à clefs multiples avec des communes de travail différentes	15 084	7 609	683	1 571	930	2009	680	1001	601
% sur nb total d'observations	0,83 %	1,79 %	0,53 %	0,82 %	0,63 %	0,7 %	0,29 %	0,44 %	0,37 %
Si utilisation MOBPRO									
Nombre d'observations à clefs multiples avec des IRIS de résidence différents	1767	0	367	510	321	178	212	179	263
% sur nb total d'observations	0,1 %	0 %	0,28 %	0,27 %	0,22 %	0,06 %	0,09 %	0,08 %	0,16 %
Source : Insee, fichiers détail INDCANTVILLE et MOBPRO, 2012									

80 Si à la suite de l'appariement on utilise le fichier MOBPRO alors les IRIS au sein des communes de résidence et/ou les actifs au sein des ménages peuvent être intervertis.

Les observations dont les clefs sont multiples représentent pour les résidents des grandes communes 8,75 % des observations. Les observations avec une clef multiple ayant des localisations d'emploi différentes concernent 0,83 % des observations des grandes communes d'Île-de-France. Et, contrairement aux petites communes nous ne disposons pas d'heuristique pour aider à procéder à l'appariement. Lorsque nous sélectionnons un individu INDCANTVILLE et qu'il est retourné plusieurs actifs MOBPRO nous ne pouvons pas regarder dans le fichier MOBPRO la commune de travail du conjoint. Ainsi, en utilisant le fichier INDCANTVILLE, 0,83 % des individus des ménages peuvent éventuellement travailler dans des communes différentes de celle dans lesquelles ils travaillent en réalité.

En conclusion de la présentation des sources d'erreurs potentielles, en effectuant l'appariement des fichiers détail INDCANTVILLE et MOBPRO librement accessibles, les taux d'erreur potentiels d'appariement sont de 0,096 % et 0,83 % pour les petites et grandes communes. Ces taux d'erreurs sont maximaux car le travail d'appariement utilise des méthodes stochastiques pour résoudre ces cas problématiques : il supposerait que l'on se trompe systématiquement lors des tirages au sort, ce qui est peu probable. Afin de quantifier précisément les erreurs d'appariement nous utilisons les données confidentielles de l'Insee.

4.1.3 Évaluation de l'appariement avec les données du CASD

« Le centre d'accès sécurisé aux données (CASD) du Genes (Groupe des écoles nationales d'économie et statistique) est un équipement conçu pour permettre aux chercheurs de travailler sur des données individuelles très détaillées, et donc soumises à la confidentialité, dans des conditions de sécurité élevées. » (CASD, présentation⁸¹) Par l'intermédiaire des services du CASD nous avons eu accès à distance aux données complètes des bulletins du recensement (le fichier confidentiel). Nous avons à disposition un seul fichier contenant l'ensemble des variables qui sont éclatées en fichiers détail pour le grand public. Dans un même fichier nous avons les communes de travail et les IRIS-Communes de résidence et les identifiants permettant de retrouver les observations appartenant au même ménage.

À partir de ce fichier du CASD nous reconstituons deux fichiers identiques aux fichiers détails INDCANTVILLE et MOBPRO tel qu'ils sont diffusés. La seule différence est que

81 CASD. Présentation [en ligne] (Consultée le 14/06/2016) <<https://casd.eu/>>

dans ce fichier nous ne disposons pas des observations hors-ménages qui sont pourtant présentes dans le dénombrement des actifs ayant des clefs multiples. Précisons que la part d'individus hors-ménage ayant des clefs multiples est à peu près égale à celle des individus en ménage.

Dans ces deux fichiers que nous construisons nous préservons les variables que l'appariement des fichiers détail grand public tente de gagner : les identifiants individuels, les identifiants des ménages, les communes de résidence des petites communes et les communes de travail. Ensuite, nous effectuons le travail d'appariement. À la suite de ce processus, nous comparons les identifiants des données confidentielles à ceux obtenus par la méthode d'appariement sur ce même jeu de donnée (tableau 15). Le tableau suivant dénombre les erreurs d'appariements brutes puis les erreurs d'appariements avec des localisations différentes, celles qui modifient la matrice de flux.

Tableau 15 : Résultats de l'appariement des données du CASD ⁸²										
		IDF	75	77	78	91	92	93	94	95
	Total d'observations d'actifs en ménage	195156	425807	190016	218800	184493	288235	238782	229548	179475
Erreurs d'appariements « réelles »	Nombre mauvais correspondance des identifiants	85751	25300	5636	9072	7395	14065	9052	9839	5392
	%	4,39 %	5,94 %	2,97 %	4,15 %	4,01 %	4,88 %	3,79 %	4,29 %	3 %
Erreurs d'appariements dues à des localisations différentes pour les grandes communes	Nombre actifs mauvaises grandes commune de résidence	0	0	0	0	0	0	0	0	0
	Si utilisation INDCANTVILLE									
	Indicateur n°1 : Nombre d'individus mauvaise commune d'emploi	8 888	1 296	488	1 336	1 249	1 421	1 490	972	636
	%	0,45 %	0,3 %	0,26 %	0,61 %	0,68 %	0,49 %	0,62 %	0,42 %	0,35 %
	Si utilisation MOBPRO									
Indicateur n°2 : Nombre d'actifs mauvais IRIS	962	0	168	270	165	71	97	89	102	
%	0,05 %	0 %	0,09 %	0,12 %	0,09 %	0,02 %	0,04 %	0,04 %	0,04 %	0,06 %
Erreurs d'appariements dues à des localisations différentes pour les petites communes	Nombre d'actifs mauvaise petite commune de résidence	122	0	55	27	24	0	0	0	16
	%	0,01 %	0 %	0,03 %	0,01 %	0,01 %	0 %	0 %	0 %	0,01 %

Source : Données confidentielles du CASD 2012

Pour chaque actif occupé nous comparons les identifiants officiels avec ceux construits par la méthode d'appariement. Les mauvais appariements sont aux alentours de 5 % (4,39%). Mais pour les grandes communes les erreurs d'appariement modifiant la matrice origine-destination ne représentent plus que 0,45 % et 0,01 % pour les petites communes. Les cas problématiques ont été résolus environ pour moitié.

Ainsi malgré la présence de clefs multiples l'appariement n'est pas systématiquement mauvais. L'évaluation que nous venons de présenter confirme que nous pouvons utiliser le fichier INDCANTVILLE apparié à MOBPRO par notre méthode comme source

82 Les effectifs totaux ne sont pas égaux au tableau du chapitre 3.3. Nous n'avons pas utilisé le fichier décrivant les individus hors-ménage et il faut enlever les actifs occupés de moins de 15 ans pour retrouver les totaux du tableau 3.3.

d'information enrichie et fiable pour la suite de notre travail de désagrégation spatiale.

Conclusion :

Suite à ce travail d'appariement nous obtenons dans le fichier INDCANTVILLE les communes de travail des actifs occupés avec une incertitude jugée acceptable. Pour les actifs occupés des petites communes nous obtenons aussi les communes de résidence. Nous transmettons ensuite les communes de résidence aux autres membres des ménages des petites communes avec un actif occupé. Nous avons ainsi localisé au lieu de résidence les observations n'étant pas des actifs occupés sur le support IRIS-Communes, au lieu du support IRIS-Cantons.

Précisons tout de même que nous n'avons pas effectué cette descente d'échelle pour une partie de la population. Par construction, nous ne pouvons pas connaître la commune de résidence des ménages des petites communes n'ayant aucun actif occupé (les retraités par exemple), puisqu'ils ne sont pas décrits dans le fichier MOBPRO, qui est notre source d'information pour affiner la localisation du canton vers la commune. Nous les laissons tout de même dans le fichier apparié sans pour autant procéder aux étapes suivantes. Ils représentent 42 305 observations de ménages sur les 1 845 217 observations de ménages d'Île-de-France soit 2,3 % des observations de ménages. En effectif, ils représentent 172 405 ménages. Rappelons qu'en 2006, le recensement indique que près de 178 000 ménages n'ont aucun actif occupé dans les petites communes de moins de 10 000 habitants. En 2011, ils sont 191 544. Les résultats de nos traitements semblent cohérents.

La prochaine étape consiste à transférer les localisations résidentielles des observations de ménages du support IRIS-Communes au carroyage de 200 m de l'Insee. En effet, bien que le support cible de notre désagrégation soit le carroyage INSPIRE de 1 km, nous utilisons comme étape intermédiaire le support plus fin de 200 m, qui a l'avantage de correspondre peu ou prou à l'échelle des bâtiments. Les données à 200 m seront ensuite ré-agrégées, pour retrouver une représentativité statistique.

4.2 Ventilation des ménages aux lieux de résidence

4.2.1 Ventilation du support IRIS-Communes au carroyage de 200m de l'Insee

Nous allons maintenant procéder à la ventilation des lieux de résidence du support IRIS-Communes au carroyage de 200m de l'Insee pour les ménages décrits dans le fichier INDCANTVILLE apparié à MOBPRO. Nous nous inspirons de la méthodologie de la Direction Générale de l'Énergie et du Climat (DGEC) qui a commandité un rapport portant sur la méthodologie de répartition spatiale de la population (Létinois, 2012). Ce rapport débute en nous informant qu'une législation européenne sur la surveillance de la qualité de l'air requiert *une cartographie des zones géographiques du dépassement d'une valeur limite et l'estimation du nombre d'habitants exposés au dépassement*. Ce rapport a pour objet de proposer un guide méthodologique spécifique au calcul de la répartition spatiale des populations pour par la suite estimer un taux d'exposition.

La méthodologie mise en place consiste à affiner les localisations résidentielles du recensement de la population vers le carroyage de population de 200m de l'Insee. Ce carroyage informe sur les effectifs de population d'individus et de ménages pour chaque carreau à partir des données d'imposition. Le rapport du DGEC insiste sur le fait que les effectifs de population du recensement et des carroyages ne sont pas identiques. Les résultats du recensement sont plus complets. « *Un simple cumul sur la Métropole fait apparaître un manque de 800 000 habitants sur le carroyage INSEE* » (Létinois, 2012, p.47). En Île-de-France, la différence entre le carroyage de 2010 et le recensement de 2012 est de 327 849 ménages. De fait, l'idée est d'utiliser les données carroyées renseignant sur l'effectif de population comme coefficient pondérateur pour ventiler les populations des IRIS-Communes sur chacun des carreaux de 200m et non pas comme valeur à atteindre pour chaque carreau.

Notre travail se distingue de la méthodologie proposée dans ce rapport en ventilant les ménages et non pas les individus. De fait nous ne ventilons pas les individus qualifiés de hors-ménage par le recensement⁸³. Aussi nous ajoutons deux étapes supplémentaires. Premièrement nous allons traiter le cas des IRIS de moins de 200 habitants. Deuxièmement

83 Même si nous avons procédé à leur appariement dans l'étape n°1.

nous allons quelquefois procéder à une séparation des clones (chapitre 3.1) afin de gérer la question des poids de sondage des observations de ménage.

Etape n°1 :

Pour des raisons de secret statistique, nous ne connaissons pas les IRIS de résidence des habitants des IRIS de moins de 200 habitants. Par exemple, les codes INSEE de ces IRIS sont tronqués de la manière suivante « 77284XXXX ». La première partie du code renseigne la commune de résidence, la suite devant identifier l'IRIS au sein de la commune n'est pas renseignée.

Pour retrouver ou au moins estimer les IRIS de résidence des ménages habitant dans des IRIS de moins de 200 habitants, nous nous servons de la couche d'information géographique des IRIS fourni par l'IGN. Ici, nous utilisons la couche des IRIS de l'année 2011 avec le millésime de 2012. Pour chaque IRIS 2011, nous comptons la population d'individu y résidant en 2012. De cette manière nous identifions les IRIS de moins de 200 habitants. Une fois identifié les IRIS de moins de 200 habitants nous sélectionnons les observations ayant un code IRIS tronqué et nous les distribuons aléatoirement dans ces IRIS précédemment sélectionnés. Nous remarquons que si dans une commune il n'y a qu'un seul IRIS de moins de 200 habitants nous retrouvons par construction le bon IRIS. Cette opération concerne 2 182 ménages francilien.

Etape n°2 : Ensuite nous procédons à l'étape de ventilation proprement dite. Une intersection géographique est effectuée entre les contours IRIS-Communes et le carroyage 200m. Le résultat du processus est un ensemble de polygones (morceaux) comportant chacun une double information : le numéro du carreau et le code de l'IRIS-Commune.

Etape n°3 : Puis, les populations P_c des mailles sont répartis dans chaque morceau de carreau en pondérant cette population par le ratio surfacique du morceau par rapport à la surface totale du carreau⁸⁴. Rappelons, la population est ici le nombre de ménage renseigné par le carroyage.

84 Nous reprenons la formulation proposée dans le rapport de la Direction Générale de l'Énergie et du Climat (Létinois, 2010)

$$P_{c,m} = \frac{P_c \times S_{c,m}}{\sum_n S_{c,n}}$$

$P_{c,m}$: population carroyée calculée sur le morceau m du carreau c

P_c : population carroyée du carreau c

$S_{c,m}$: surface du morceau de carreau m du carreau c

$S_{c,n}$: surface du morceau de carreau n du carreau c

Etape n°4 : La population « carroyée » de chaque carreau est utilisée comme pondérateur pour estimer la population de ménage du recensement à placer dans chacun des carreaux ou morceaux de carreaux contenus dans les IRIS-communes.

$$Pr_{c,m,i} = Pr_i \times \frac{P_{c,m,i}}{\sum_{d,n} P_{d,n,i}}$$

$Pr_{c,m,i}$: population estimée du recensement dans le morceau m du carreau c de l'IRIS-Communes i

Pr_i : population du recensement de l'IRIS-Communes i.

$P_{c,m,i}$: population carroyée calculée sur le morceau m du carreau c de l'IRIS-Commune

$P_{d,n,i}$: population carroyée calculée sur le morceau n du carreau d de l'IRIS-Commune i

Etape n°5 : Avec ces trois dernières étapes nous disposons d'un effectif de ménage à localiser dans chaque IRIS-Commune. L'étape 4 consiste à effectuer un tirage de Monte-Carlo pour approcher les effectifs de ménages estimés en tirant au sort des ménages et en les plaçant afin de respecter les effectifs de ménages préalablement estimés. Toutefois, ces ménages ont des poids de sondage ce qui empêche d'atteindre précisément les estimations.

Pour approximer le nombre de ménages par carreau estimé avec les poids de sondage nous séparons dans certains cas les clones des observations de ménages (chapitre 3.1). Lorsque un ménage est placé dans un carreau ou du morceau de carreau et que celui-ci fait dépasser

l'effectif estimé alors nous séparons les clones. Une observation identique au ménage est créée et les poids de sondage sont divisés afin de respecter l'estimation de l'étape n°3. Le poids de sondage restant est attribué aux clones séparés et ces nouvelles observations sont placées dans un autre carreau en veillant toujours à ne pas dépasser les effectifs estimés.

A l'issue du tirage de Monte-Carlo nous avons attribué un identifiant carreau à partir des morceaux de carreau pour chaque ménage.

Nous allons maintenant procéder à l'examen des résultats de la ventilation sous contrainte des lieux de résidence.

4.2.2 Évaluation de la ventilation sous contrainte des lieux de résidence des ménages

L'examen de l'étape de ventilation se déroule en deux étapes. Nous regardons premièrement si nous avons bien attribué un identifiant carreau à toutes les observations de ménage du fichier apparié. Ensuite, nous regardons si les observations sont bien réparties dans l'ensemble de l'espace francilien c'est-à-dire si nous avons bien des observations dans l'ensemble des carreaux de 200m et de 1km.

Tableau 16 : Décompte des observations de ménages ventilées sur le carroyage de 200m									
	Total	75	77	78	91	92	93	94	95
Effectif initial de ménages	5 011 983	1 159 088	520 817	557 593	481 834	697 304	588 200	561 544	445 598
Nombre initial d'observations de ménages	1 845 217	454 917	162 807	192 798	162 113	270 630	228 435	216 568	156 949
Effectif initial d'individus en ménages	11 661 516	2 192 890	1 329 833	1 383 449	1 207 473	1 557 931	1 507 610	1 316 715	1 165 615
Nombre initial d'observations d'individus en ménage	4 231 656	861 225	410 811	470 926	401 812	599 614	580 640	501 920	404 708
Effectif d'hors-ménages	236 532	47 790	23 487	28 633	30 162	28 546	31 148	25 118	21 645
Nombre d'observations de hors-ménages	69 495	13 465	6 916	7 929	8 240	7 861	10 547	7 022	7 515
Effectif de ménages ventilés	4 838 050	1 158 364	449 277	522 685	435 526	696 987	587 301	558 635	429 272
Nombre d'observations de ménages ventilées	1 889 340	45 9940	164 918	200 260	163 144	277 341	236 238	223 444	164 055
Effectif de clones séparés	143 157	8 204	33 580	26 066	22 058	10 696	12 236	12 171	18 143
Nombre d'observations de clones séparés	87 126	5 364	19 304	15 822	13 200	6 812	8 048	7 562	11 014
Effectif ménages non ventilés	173 932	724	71 539	34 908	46 307	317	899	2 909	16 326
Nombre d'observations de ménages non ventilés	43 003	341	17 193	8 360	12 169	101	245	686	3 908
Nombre de ménages non ventilés de petits IRIS	1 526	540	157	198	224	83	160	22	138
Nombre d'observations de ménages non ventilés de petits IRIS	698	341	45	49	76	38	56	5	88
Nombre de ménages non ventilés car pas d'actifs occupés (petites communes)	172 405	0	71 382	34 709	46 083	233	738	2 886	16 188
Nombre d'observations de ménages non ventilés car pas d'actifs occupés (petites communes)	42 305	0	17 148	8 311	12 093	63	189	681	3 820

Source : Données du recensement désagrégées. IFSTTAR DEST/ LVMT, Insee, 2012

Suite à l'appariement des fichiers de détails INDCANTVILLE et MOBPRO nous avons près de 2 millions d'observations représentant une population de près de 5 millions de ménages. Nous enlevons de l'analyse les 230 000 individus dits hors-ménage. À peu près 43 000 observations n'ont pas reçu d'identifiant carreau : ils n'ont pas été ventilés. Pour 42 305 d'entre eux il s'agit de ménages dont aucun membre n'est un actif occupé et habitant dans une commune de moins de 10 000 habitants. Pour ces derniers, l'étape d'appariement n'a pas pu apporter la commune de résidence (rappelons que pour ces observations nous connaissons seulement le canton de résidence et nous avons fait le choix de ne pas les localiser plus finement à cette étape). Nous avons ventilé 4 838 050 de ménages.

Environ 700 observations n'ont reçu aucun identifiant de carreau sur les 2 182 ménages de ménages des IRIS de moins de 200 habitants, ces observations représentent 1 500 personnes.

De manière générale le processus fonctionne correctement, avec seulement 3,4 %⁸⁵ des observations non ventilées. Nous passons à la deuxième étape d'évaluation des lieux de résidence. Nous observons maintenant si les observations recouvrent bien l'ensemble de l'espace francilien (tableau 17). Et nous nous interrogeons également sur la distribution des profils socio-économiques non directement pris en compte dans la méthode de ventilation sous contrainte des effectifs de ménages.

85 Le calcul est effectué avec les effectifs de ménages et non les effectifs d'observations.

Tableau 17 : Remplissage moyen de la ventilation par effectif de ménage des carreaux de 200m									
		Total (absolue)	Ménage de 1personnes dans les carreaux (%)	Ménage de 5p et plus (%)	Ménages propriétaires (%)	Ménages locataires (%)	Logement individuel (%)	Logement collectif (%)	
Total des carreaux d'IDF (72370 carreaux)	Effectif total du carroyage et moyennes par profils	4 684 133	23.6	10.6	70.6	29.9	74.8	25.1	
	Effectif ventilé et moyennes par profils	4 838 050	23.1	10.5	66.7	30.6	63.5	35.4	
	Nombre de carreaux vides de 200m après ventilation	5 094	-	-	-	-	-	-	-
	Nombre de ménages du carroyage dans les carreaux non remplis	9 365	-	-	-	-	-	-	-
	Effectif moyen du carroyage des carreaux non remplis par la ventilation	2	-	-	-	-	-	-	-
Petits carreaux (moins de 100 ménages)	Effectif moyen des carreaux peu peuplés (60703 carreaux)	21	21	10	76	24	85	14	
	Effectif moyen des carreaux peu peuplés après ventilation (55823 carreaux)	21	20	10	71	26	71	27	
Carreaux moyens (de 100 à 500 ménages)	Effectif moyen des carreaux moyennement peuplé (9990 carreaux)	213	35	9	43	56	18	81	
	Effectif moyen des carreaux moyennement peuplés après ventilation (9448 carreaux)	228	33	8	47	49	28	69	
Grands carreaux (plus de 500 ménages)	Effectif moyen des carreaux grandement peuplés (1673 carreaux)	756	45	7	33	67	1	98	
	Effectif moyen des carreaux grandement peuplés après ventilation (2034 carreaux)	844	43	6	35	60	6	91	

Source : Données du recensement désagrégées. IFSTTAR DEST/ LVMT, Insee 2012

La ventilation sous contraintes n'a pas rempli près de 5 000 carreaux. Les données du carroyage de 200 m indiquent pourtant que 9 000 ménages fiscaux se trouvent dans ces carreaux, avec en moyenne 1,8 ménages par carreau. Il s'agit de carreaux en périphérie des zones urbaines denses, voire dans l'espace rural dans le cas des départements de la grande couronne.

En observant les taux de remplissage moyen selon les effectifs de ménages par carreau nous remarquons des différences. D'abord les petits carreaux de moins de 200 ménages ont des taux de remplissage similaires entre les données du carroyage de 200 m et les résultats de la ventilation. Les carreaux de moins de 200 ménages contiennent en moyenne 21 ménages dans les deux jeux de données. En revanche, pour les carreaux que nous appelons moyens (entre 200 et 500 ménages) et pour les grands carreaux (plus de 500 ménages) les résultats de la ventilation présentent des effectifs plus importants que le carroyage de 200 m. Respectivement, les carreaux moyens des données carroyées abritent en moyenne 213 ménages contre 228 ménages. Les carreaux Insee grandement peuplés contiennent 756 ménages contre 844 ménages avec la ventilation. Nous rappelons que les données du recensement contiennent plus de ménages que les données carroyées. Le carroyage Insee de 200 m contient près de 4 684 133 ménages fiscaux, le recensement 5 011 938 ménages. La méthode de ventilation sous contrainte place en priorité le surplus de ménages dans les grands carreaux.

Nous regardons maintenant comment la ventilation spatiale a réparti les observations de ménages selon les profils socio-économiques. En effet, l'étape de ventilation sous contraintes par effectif de ménage ne prend pas en compte directement les profils socio-économiques. Pour regarder la répartition des profils socio-économiques nous comparons les variables décrivant les ménages des données carroyées avec leur corollaire du recensement (effectif de ménages de 1 personne, de 5 personnes et plus, nombre de ménages propriétaires, de locataires, nombre de logement individuel et collectif).

Nous commençons par la variable indiquant le nombre de ménages de 1 personne. En moyenne, les carreaux Insee contiennent 23 % de ménages de 1 personne. Les résultats de la ventilation par effectif de ménage retournent également 23 %. En observant ces proportions selon la taille des carreaux la ventilation est très proche des proportions du carroyage de 200m peu importe la typologie. Les carreaux contenant moins de 100 habitants contiennent

dans les deux jeux de données une proportion de 20 % de ménages de 1 personne. Les carreaux moyennement ou très peuplés contiennent environ 34 et 44 % de ménages de 1 personne.

Ensuite, les proportions de ménages de 5 personnes et plus sont également similaires entre les deux jeux de données. En moyenne les carreaux contiennent 10 % de ménages de 5 personnes et plus. Selon la catégorie des carreaux les proportions sont également proches. Les petits carreaux contiennent 10 % de grands ménages, les carreaux moyens en contiennent 9 % et les grands carreaux en zones plus denses en contiennent environ 7 %.

En revanche, la répartition spatiale des ménages propriétaires est un peu moins bonne. Les données carroyées Insee indiquent une proportion moyenne de 70 % de ménages propriétaires par carreau. Les résultats de la ventilation donne 66 %. En observant la distribution par type de carreau nous remarquons que les petits carreaux sont sous-estimés par la ventilation (76% contre 71%). Et les carreaux de moyennes et de grandes tailles sont sur-évalués d'environ 3 points par la méthode de ventilation. Nous en déduisons que la répartition des ménages propriétaires est structurée. La propension à être propriétaire de son logement est plus forte dans les zones peu denses à l'intérieur des IRIS-communes (carreaux de petits effectifs). La ventilation ne considérant pas directement cette variable (chapitre 3.2.1), elle amoindrit quelque peu cette répartition à l'intérieur des IRIS-Communes.

Mécaniquement, les proportions des locataires sont asymétriques par rapport aux répartitions des propriétaires. En moyenne les proportions sont bien identiques entre les jeux de données. Toutefois en observant les proportions par type de carreaux nous remarquons que dans les carreaux peu peuplés les locataires sont surestimés de 2 points. Et dans les carreaux moyennement et très peuplé ils sont sous-estimés de 7 points. En zones plus urbaines il y a plus de ménages locataires que ne le suggèrent les résultats de la ventilation.

Les variables indiquant si les ménages habitent dans des logements individuels ou collectifs sont sujettes aux plus grands écarts. La proportion de logements individuels est sous-évaluée de près de 14 points dans les carreaux de faibles effectifs. Et, la proportion de logements individuels est surestimée par la ventilation dans les carreaux moyennement et grandement peuplés. Les écarts sont respectivement de 10 et 5 points.

Les proportions de ménages habitant dans des logements collectifs sont également mal représentées dans les espaces peu habités (moins de 100 ménages). La ventilation surestime

les proportions de logements collectifs pour les petits carreaux. Pour les carreaux de taille moyennes la ventilation sous-estime à l'inverse les logements collectifs. En revanche, dans les zones plus denses la ventilation retourne une proportion de ménages habitant dans des logements collectifs proche du carroyage Insee de 200m. Ces écarts peuvent être expliqués soit par la non prise en compte de ces variables dans le processus de ventilation sous contrainte, soit par les biais des données carroyées (méthode de floutage).

En conclusion de cette analyse de la ventilation sous contraintes des ménages au lieu de résidence nous pouvons affirmer que les variables socio-économiques non prises en compte dans la méthode sont assez bien réparties dans l'espace. Rappelons tout de même que nous avons environ 5 000 carreaux non remplis par la ventilation sur 72 370 qui contiennent 9 365 ménages.

Nous allons maintenant ré-agréger les deux sources de données sur le carroyage INSPIRE de 1km et effectuer le même travail de comparaison des répartitions spatiales des observations de ménages, pour vérifier que nous retrouvons les profils socio-économiques qui ont été un peu perturbés à l'échelle du carreau de 200 m (tableau 18). La ré-agrégation consiste à attribuer les identifiants carreau de 1km contenant les carreaux de 200m dans lesquels ont été ventilés les observations. L'étape de ré-agrégation ne pose pas de traitement particulier puisque les deux supports respectent les normes de construction INSPIRE (chapitre 2.1.2.2).

Tableau 18 : Remplissage moyen de la ventilation par effectif de ménage après ré-agrégation au carroyage de 1km		Total (absolue)	Ménage de 1 personne dans les carreaux (%)	Ménage de 5p et plus (%)	Propriétaires (%)	Locataires (%)	Logement individuel (%)	Logement collectif (%)
Total des carreaux d'IDF (8539 carreaux)	Effectif total du carroyage et moyennes par profils	4 676 890	22.7	10.7	70.2	29.9	80.1	19.8
	Effectif ventilé et ré-agrégé et moyennes par profils	4 837 360	27.8	17.4	75.2	35.6	77.0	40.5
	Nb de carreaux vides initial du carroyage de 1km	278	-	-	-	-	-	-
	Nb de carreaux non rempli par la ventilation/ré-agrégation	783	-	-	-	-	-	-
	Nb de ménages du carroyage dans les carreaux non remplis	952	-	-	-	-	-	-
	Effectif moyen du carroyage des carreaux non remplis par la ventilation	1,8	-	-	-	-	-	-
Petits carreaux (moins de 300 ménages)	Effectif moyen des carreaux peu peuplé (6213 carreaux)	49	20	10	74	25	91	8
	Effectif moyen des carreaux peu peuplé après ré-agrégation (5907 carreaux)	45	27	21	81	32	88	29
Carreaux moyens (de 300 à 600 ménages)	Effectif moyen des carreaux moyennement peuplés (536 carreaux)	430	22	10	72	27	72	27
	Effectif moyen des carreaux moyennement peuplés après ré-agrégation (428 carreaux)	427	21	10	67	30	65	34
Grands carreaux (plus de 600 ménages)	Effectif moyen des carreaux grandement peuplés (1516 carreaux)	2729	30	10	51	48	35	64
	Effectif moyen des carreaux grandement peuplés après ré-agrégation (1421 carreaux)	3086	31	9	51	46	35	63
Source : Données du recensement désagrégées. IFSTTAR DEST/LVMT, Insee 2012								

L'effectif total des ménages ré-agrégés est un peu plus faible que l'effectif ventilé aux carreaux de 200m (4 837 360 ménages au lieu de 4 838 050 soit une perte de 690 ménages).

Quelques carreaux de 200m ne sont pas compris dans le carroyage de 1km. La création de carreaux de 1km supplémentaire permettrait de les considérer. A l'inverse, après la ré-agrégation nous obtenons également des carreaux de 1km vides. Une partie des 783 carreaux recouvrant les espaces habités de l'Île-de-France sont en bordures des limites de la région. Ils ne contiennent pas de carreaux de 200m et sont donc par construction vides de population (278 carreaux). Les 505 carreaux restants contiennent pourtant 952 ménages d'après les données carroyées Insee agrégées au carreau de 1km. En moyenne, ces carreaux vides abritent 1,8 ménages. Cette perte s'explique par la ventilation à 200m qui n'a pas rempli tous les carreaux.

Sans distinction des profils socio-économiques les effectifs moyens des deux jeux de données sont relativement proches pour les carreaux de petites tailles (moins de 300) et de moyennes tailles (entre 300 et 600 ménages). Les carreaux de plus grandes tailles sont légèrement surestimés par la ventilation/ré-agrégation. Le surplus du recensement est localisé en priorité dans les carreaux plus grandement peuplés à l'instar du carroyage de 200m.

L'examen des répartitions par profils socio-économiques donne des résultats différents que ceux du carroyage de 200m. Après ré-agrégation, les carreaux moins peuplés (moins de 300 ménages) ont des taux de remplissage plus éloignés que ceux du carroyage de 200m. À 200m, les taux sont identiques avec la ré-agrégation nous obtenons un écart de 4 points. Les proportions de ménages de 5 personnes et plus et ceux des logements collectifs sont très éloignées des résultats du carroyage de 1km. La ventilation / ré-agrégation surestime respectivement de 11 et 20 points les proportions de ces deux variables. En revanche, les écarts entre les deux jeux de données deviennent plus faibles pour la catégorie des carreaux de moyennes tailles (entre 300 et 600 ménages) excepté pour la variable des logements individuels. Enfin, les carreaux grandement peuplés des deux jeux de données ont des proportions très similaires entre les deux sources de données.

En conclusion, les résultats de la ventilation / ré-agrégation sont satisfaisantes. Les écarts entre les données carroyées et le fichier détail apparié sont suffisamment faibles. La ventilation ayant pour seul critère les effectifs de ménages réparti de manière convenable les variables socio-économiques. Les écarts entre les proportions sont suffisamment faibles pour considérer que nous avons préservé la représentativité socio-économique des données (illustration 30 et 31).

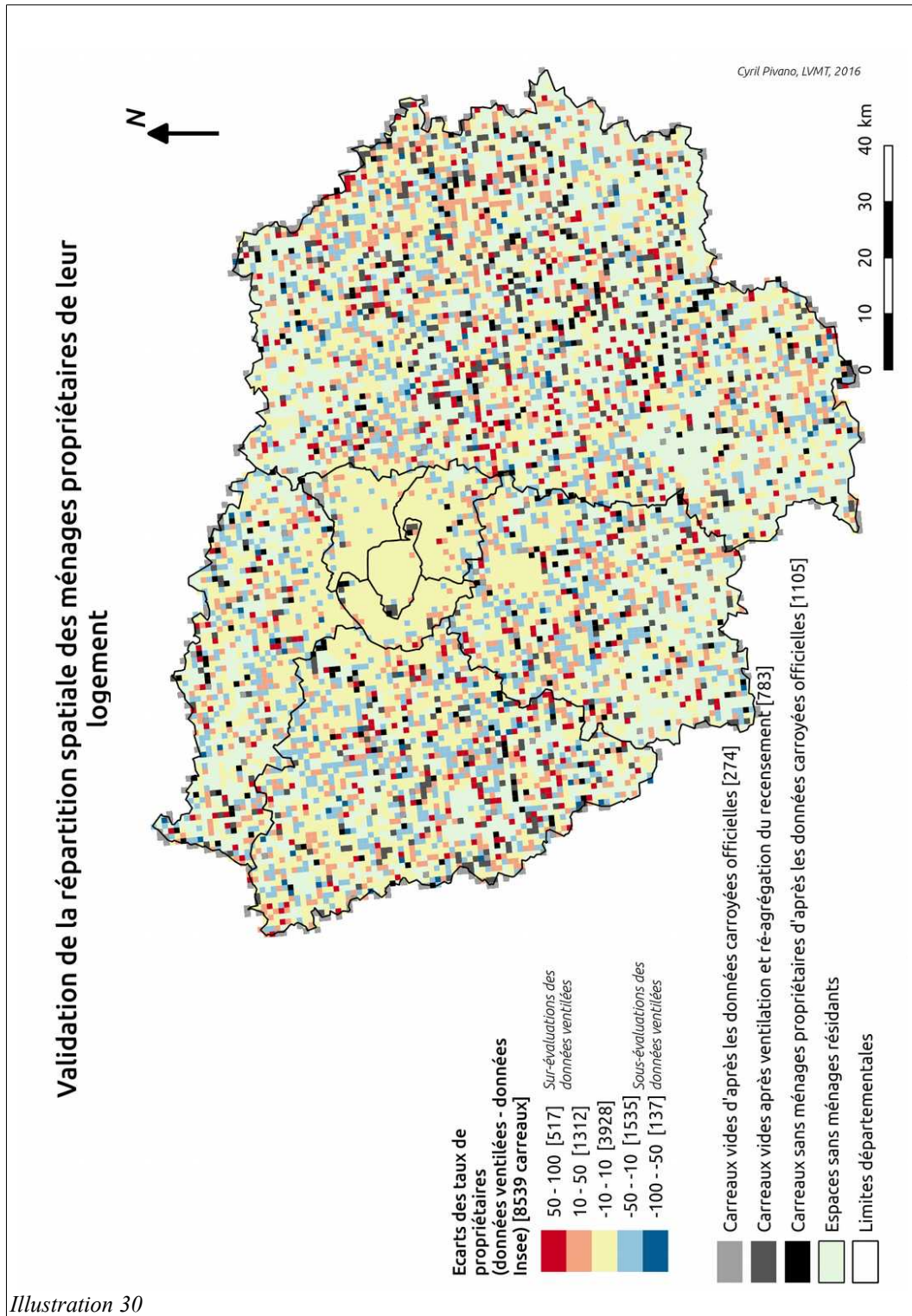


Illustration 30

La carte n°31 montre qu'avec des classes de 20 points les répartitions spatiales des proportions des propriétaires sont très proches.

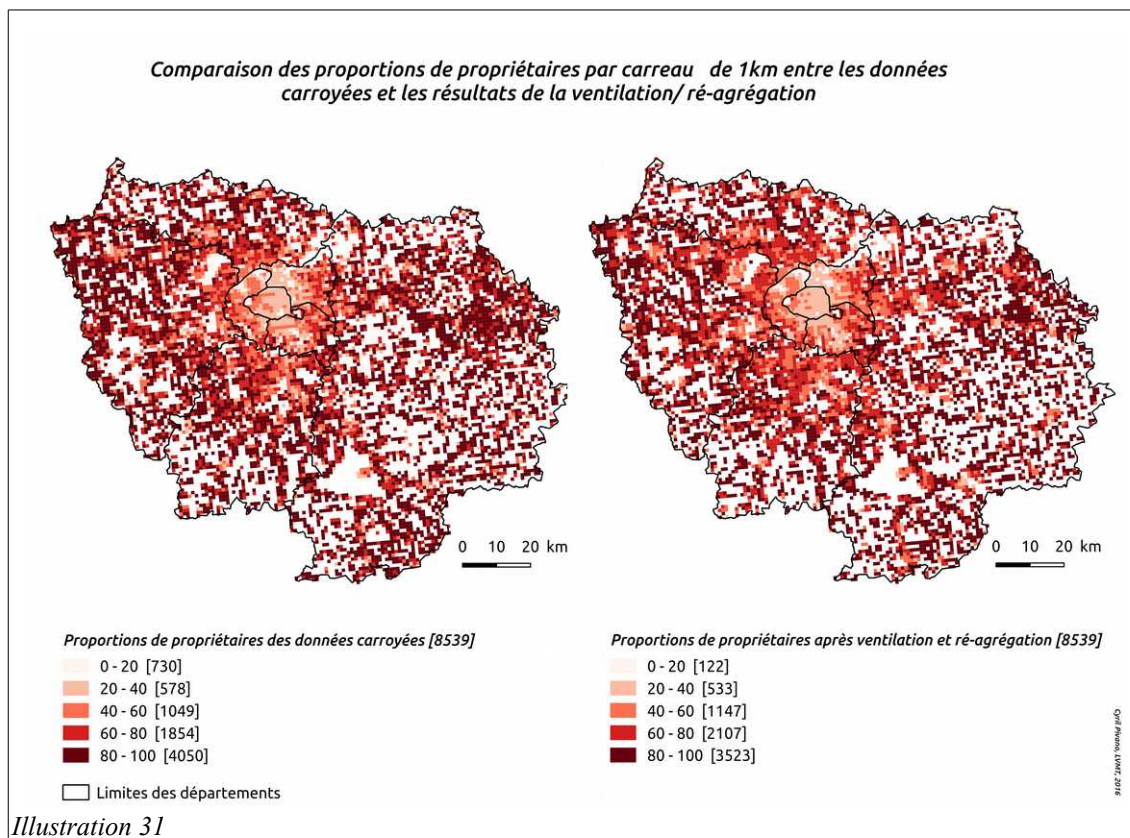


Illustration 31

Ce travail de comparaison a pu être effectué puisque les données de population carroyées informent sur la présence des ménages propriétaires. Nous pouvons bien entendu effectuer ce travail de cartographie pour l'ensemble des variables du recensement non décrites dans les données carroyées. Nous avons cependant aucun moyen de vérifier la pertinence des résultats. Notre travail de désagrégation des données de mobilités consiste pourtant à s'appuyer sur la répartition spatiale des secteurs d'activité aux lieux de résidence. Nous allons nous appuyer sur cette variable pour ventiler spatialement les lieux d'emploi.

4.3 Ventilations des lieux d'emploi

A l'aide du recensement rénové nous connaissons la commune dans laquelle l'établissement de l'employeur est domicilié. Nous allons affecter à chaque actif occupé habitant et travaillant en Île-de-France une entreprise au sein de leur commune de travail. Sur les 5 625 159 actifs travaillant dans la région nous traitons de 5 277 072 d'actifs y habitant et travaillant, soit 93,8 %.

Le processus de descente d'échelle consiste à choisir un établissement économique du répertoire SIRENE localisé à l'adresse postale. La descente d'échelle de la commune à l'adresse est plus abrupte que la ventilation résidentielle dans les grandes communes qui passe par la localisation à l'IRIS. Mais, nous allons cette fois-ci nous appuyer sur 3 variables de contrôle : les secteurs d'activité, les effectifs de salariés, et selon les modes de transport renseignés par les observations du recensement les distances aux stations de travail ou les distances domicile-travail déclarées effectuées en marche à pied qui, nous l'avons vu dans le chapitre 3, constituent des indicateurs de plausibilité de la navette domicile-travail.

4.3.1 Les secteurs d'activité

Grâce à l'appariement des fichiers détails INDCANTVILLE et MOBPRO nous disposons du secteur d'activité en 17 postes et non plus en 5 postes comme diffusé initialement dans le fichier MOBPRO (tableau 19).

Tableau 19 : Correspondance entre les différentes typologies des secteurs d'activités		
Secteurs d'activité en 5 postes	Secteurs d'activité en 17 postes	Nombre d'établissements SIRENE en IDF
AZ : Agriculture, sylviculture et pêche	AZ : Agriculture, sylviculture et pêche	7 906
BE : Industrie manufacturière, industries extractives et autres	C1 : Fabrication de denrées alimentaires, de boissons et de produits à base de tabac	7 281
	C2 : Cokéfaction et raffinage	49
	C3 : Fabrication d'équipements électriques, électroniques, informatiques ; fabrication de machines	3 753
	C4 : Fabrication de matériels de transport	527
	C5 : Fabrication d'autres produits industriels	29 185
	DE : Industries extractives, énergie, eau, gestion des déchets et dépollution	6 756
FZ : Construction	FZ : Construction	86 653
GU : Commerce, transports et services divers	GZ : Commerce ; réparation d'automobiles et de motocycles	186 840
	HZ : Transports et entreposage	33 319
	IZ : Hébergement et restauration	52 649
	JZ : Information et communication	60 057
	KZ : Activités financières et d'assurance	45 671
	LZ : Activités immobilières	41 844
	MN : Activités scientifiques et techniques ; services administratifs et de soutien	216 137
	RU : Autres activités de services	120 364
OQ : Administration publique, enseignement, santé humaine et action sociale	OQ : Administration publique, enseignement, santé humaine et action sociale	121 311
ZZ : Sans objet ou non renseigné	ZZ : Sans objet ou non renseigné	0
Total		1 020 312

Source : Nomenclature des emplois, SIRENE 2010

Le répertoire SIRENE indique le secteur d'activité de la nomenclature NAF2 en 734 classes à l'aide de la variable nommée APTET700. Il nous faut alors construire la variable des secteurs d'activités en 17 classes dans le répertoire SIRENE. La conversion de la nomenclature en 734 postes en 17 classes est réalisée à l'aide des tableaux de correspondances diffusés par l'Insee. De cette manière, pour chaque actif occupé habitant et travaillant en Île-de-France nous allons sélectionner les entreprises du même secteur d'activité décrit en 17 classes de leurs communes de travail. Par construction, plus le nombre d'établissements économiques d'un

secteur d'activité est rare dans la commune moins il y aura d'éventail de possibilités pour placer les observations d'actifs occupés.

4.3.2 Les effectifs de salariés

La suite de la ventilation consiste à effectuer un tirage de Monte-Carlo avec remise pondéré par les classes d'effectifs à l'aide de la variable EFETCENT du répertoire SIRENE. Cette variable correspond à l'effectif de salarié à la centaine près de l'établissement millésimé au 31 décembre. Comme le répertoire SIRENE renseigne sur des classes d'effectifs nous choisissons une valeur pour représenter chaque classe. Nous choisissons les bornes hautes des classes auxquelles il est ajouté un effectif de 1. Nous ajoutons un non-salarié aux entreprises (un chef d'entreprise) (tableau 20).

Tableau 20 : Choix des bornes des classes d'effectifs de SIRENE	
Variable EFFECENT	Valeur représentant les classes
NN : Unités non employeuses (pas de salarié au cours de l'année de référence et pas d'effectif au 31/12). Cette tranche peut contenir quelques effectifs inconnus	1
0 : 0 salarié (n'ayant pas d'effectif au 31/12 mais ayant employé des salariés au cours de l'année de référence)	1
1 : 1 à 2 salariés	3
3 : 3 à 5 salariés	6
6 : 6 à 9 salariés	10
10 : 10 à 19 salariés	20
20 : 20 à 49 salariés	50
50 : 50 à 99 salariés	100
100 : 100 à 199 salariés	200
200 : 200 à 299 salariés	300
.../... de 100 en 100 jusqu'à 999 999 salariés	400,500,etc
Approximation du nombre d'emplois	7 780 244
Effectif MOBPRO dans ménages ordinaires habitant et travaillant en Île-de-France	5 277 072
Sources : SIRENE,2010 ; MOBPRO, 2012	

Avec cette transformation de classes d'effectifs en nombre de salariés nous disposons de plus de salariés SIRENE que d'actifs occupés du recensement.

4.3.3 Les distances en marche à pied

Pour les actifs occupés déclarant utiliser les transports en commun pour aller travailler nous ajoutons une variable de contrôle : la distance en marche à pied à parcourir entre les établissements de travail et les stations de transport en commun. Parmi les établissements économiques de même secteur d'activité que ceux des actifs que nous traitons nous allons choisir un établissement économique étant à proximité d'une station de transport en commun. Pour ce faire, nous calculons la distance minimale à vol d'oiseau aux stations de transport en commun pour chaque établissement économique. Nous calculons ensuite un poids de tirage selon ces distances minimales à vol d'oiseau. Les poids de tirage sont calibrés en fonction de l'histogramme (d'un pas de 100 m) des distances entre les stations de travail et les lieux d'emploi de l'EGT 2010 présentés au chapitre 3 (illustration 28). Par exemple, si un établissement économique est à 400 m d'une station, nous calculons le pourcentage de franciliens de l'EGT travaillant à 400 m d'une station, et ce pourcentage devient le poids de tirage des établissements pour la variable de contrôle relative à la distance aux stations de transport en commun.

Pour les observations des actifs déclarant travailler à pieds nous utilisons l'histogramme des distances entre le domicile et le lieu de travail des observations de l'EGT 2010 déclarant aller travailler dans un lieu fixe (illustration 29)

Nous présentons maintenant comment nous combinons les trois variables de contrôle que nous venons de présenter⁸⁶. Premièrement, nous sélectionnons uniquement les établissements économiques de même secteur d'activité que les observations. Deuxièmement, nous construisons un arbre de tirage à l'aide des poids de tirage des deux variables de contrôle restantes (effectifs de salariés et des distances) (tableau 21). Par exemple, s'il y a 20 établissements de N salariés nous construisons alors une liste de 20*N lignes. Pour chaque ligne représentant un salarié nous affectons le poids de tirage relatif à la distance. Puis nous construisons la courbe des fréquences cumulées de cette deuxième colonne (opération de réduction). Troisièmement, nous procédons à un tirage au sort d'une valeur comprise entre 0 et 100. En fonction de la valeur du tirage au sort nous sélectionnons l'entreprise équivalente. En prenant l'exemple de l'arbre de tirage de Monte-Carlo du tableau suivant, si nous tirons au

86 Pour les observations déclarant utiliser un véhicule personnel nous utilisons que les variables de contrôle relative aux secteurs d'activité et au effectif de travailleurs des établissements.

sort le nombre 21 alors nous considérons que l'observation que nous sommes en train de traiter sera localisée à l'établissement de travail dont l'identifiant est le chiffre 2.

Tableau 21 : Exemple de construction de l'arbre de tirage de Monte-Carlo			
Identifiant des établissements SIRENE	Liste des salariés	Poids de tirage relatif à la distance	Fréquence cumulée
1	1	5,5 %	5,5
1	1	5,5 %	11
1	1	5,5 %	16,5
2	1	4,2 %	20,7
2	1	4,2 %	24,9
3	1	3,7%	27,6
...
20	1	2,17 %	100

Nous répétons ce tirage au sort pour l'ensemble des observations des actifs occupés. Nous rappelons que nous effectuons un tirage avec remise, ainsi l'arbre de tirage est construit une seule fois pour chaque commune. Les temps de calculs sont amoindris. Mais la répartition des actifs dans les établissements peut être lacunaire. Des établissements économiques du répertoire SIRENE peuvent ne accueillir pas assez ou même aucune observations du recensement.

4.3.4 Évaluation de la ventilation des lieux de travail

Le premier élément à vérifier après la ventilation des lieux de travail est le nombre d'actifs occupés du recensement qui ne sont pas localisés. Suite à la ventilation résidentielle et à la séparation des observations, nous avons 2 015 266 observations représentant les 5 277 072 actifs occupés des ménages ordinaires habitant et travaillant dans la région Île-de-France⁸⁷. La ventilation des lieux de travail n'a pas pu localiser 4 111 de ces observations parce que nous ne retrouvons pas d'établissements économiques SIRENE de même secteur d'activité au sein des communes de travail.

Ensuite, comme l'étape de ventilation sous contrainte est un tirage avec remise nous devons compter les effectifs d'actifs occupés pour chaque établissement économique SIRENE afin d'observer si nous avons bien rempli les établissements (tableau 22). Pour effectuer ce

⁸⁷ Dans les fichiers détail officiel du recensement nous avons 1 914 648 observations représentant 5 277 072 actifs occupés de ménages ordinaires habitant et travaillant dans la région Île-de-France.

décompte nous avons agrégés les classes d'effectifs des établissements en 8 classes.

Tableau 22 : Dénombrement des effectifs ventilés à l'adresse postale de travail							
Classes	"0 salariés"	"de 1 à 20 salariés"	"de 21 à 50 salariés"	"de 51 à 100 salariés"	"de 101 à 500 salariés"	"de 501 à 1000 salariés"	"de 1001 à 5000 salariés"
Nombre d'établissements total	673 937	309 850	21 347	7 775	6 236	732	413
Nombre d'établissements avec aucune observations ventilées	517 801	107 725	1 439	513	418	60	27
Pourcentage de remplissage par rapport aux bornes hautes des classes d'effectifs SIRENE	75.5	69.0	66.3	67.8	65.3	62.6	62.3
Source : Données du recensement désagrégées IFSTTAR - DEST/ LVMT, Insee 2012							

Premièrement, plus les établissements sont petits en nombre de salariés plus ils sont enclins à rester vide. Les établissements avec aucun salarié, de type auto-entreprises, sont bien plus sensibles que les plus grandes entreprises. Les établissements de 3 à 20 salariés sont aussi mal remplis par le tirage avec remise.

La dernière ligne du tableau informe sur le pourcentage de remplissage. Attention, il n'est pas calculé avec les deux premières lignes du tableau. Pour chaque établissement SIRENE nous regardons la part que représentent les effectifs ventilés par rapport aux bornes hautes des classes renseignées dans le répertoire SIRENE⁸⁸. Ensuite nous faisons la moyenne de ces taux de remplissage en reprenant les 8 classes. Bien que les petits établissements avec aucun salarié aient le plus souvent des effectifs de 0 observation ventilée le taux de remplissage moyen est de 75 %. En effet dès que nous plaçons une observation nous remplissons totalement l'effectif de travailleurs. Et en raison du poids de sondage nous dépassons systématiquement les effectifs renseignés par le répertoire SIRENE⁸⁹. Il y a donc un effet de compensation entre les petits établissements vides et les petites établissements en sureffectif. Un tirage sans remise pourrait améliorer les taux de remplissage mais les temps de calculs pour l'Île-de-France seraient de plus d'une semaine rien que pour cette étape, c'est pourquoi nous choisissons de ne pas effectuer de tirages sans remises. Nous allons voir que l'agrégation au carroyage de 1km apporte des résultats différents.

88 Il se peut que nous avons des taux de remplissage plus grand que 100.

89 Nous ne procédons pas à la séparation des clones

Nous observons ensuite sur ce tableau que les très grandes entreprises, plus de 5 000 salariés, sont en sous-effectifs.

Rappelons que l'objet de notre travail est de proposer une information sur un carroyage de 1km. Les résultats nous intéressant sont donc ceux agrégés sur ce carroyage. Pour chaque carreau de 1km du carroyage INSPIRE nous comptons les effectifs de salariés du répertoire SIRENE et ceux ventilés (tableau 23). Puis nous calculons les taux de remplissage et faisons la moyenne par classe d'effectif. Pour des objectifs de comparaison nous reprenons les mêmes classes que pour le dénombrement des effectifs par établissements à l'adresse postale.

Tableau 23 : Taux de remplissage des carreaux de 1 km de travail								
Classes des carreaux	"0 salariés"	"de 1 à 20 salariés"	"de 21 à 50 salariés"	"de 51 à 100 salariés"	"de 101 à 500 salariés"	"de 501 à 1000 salariés"	"de 1001 à 5000 salariés"	"Plus de 5000 salariés"
Effectif par classes	433	1843	694	482	1057	468	864	289
Pourcentage de remplissage moyen	128.6	105.5	83.3	82.2	79.3	75.9	74.3	70.9

Source : Données du recensement désagrégées. IFSTTAR DEST/ LVMT, Insee 2012

Nous observons que les taux de remplissage à l'échelle d'1 km sont proches de 100 %. Alors qu'à l'échelle des établissements les petits établissements étaient en sous-effectif (75%), après ré-agrégation les carreaux ayant peu d'emplois sont cette fois-ci en sureffectifs (128%).

Les carreaux contenant plus de salariés ont des taux de remplissage de meilleurs qualités (plus proche de 100). Mais ils demeurent en sous-effectif. Rappelons que nous nous servons des bornes hautes des classes d'effectif de SIRENE. En prenant les bornes moyennes nous aurions mécaniquement des taux de remplissage plus près de 100 %.

Nous concluons tout de même que le tirage de Monte-Carlo avec remise est suffisant. Il est donc inutile d'effectuer un tirage de Monte-Carlo sans remise coûteux en temps de calculs si l'on ré-agrège à 1km.

Passons à l'examen des distributions des distances en marche à pied.

Évaluation des variables de contrôle relatives aux distances en marche à pied

L'étape de ventilation des lieux de travail s'appuie sur des données de contrôle relatives à la distance. Il convient maintenant d'observer les différences entre les distributions de distance contrôle de l'EGT et les distributions des distances simulées. Nous regardons premièrement les distances aux stations de travail, puis, nous examinons les distances domicile-travail effectuées en marche à pied.

La distance aux stations de travail :

La comparaison entre les distributions des distances à la station de travail EGT et les distribution des distances minimales à vol d'oiseau du lieu de travail choisi par la ventilation ne peut pas se faire sur l'ensemble des observations. Nous allons regarder seulement les observations décrivant des individus travaillant en petite couronne. En effet, pour mesurer la distance aux stations de transport en commun des établissements nous avons utilisé une couche d'information géographique des stations de transport en commun dit lourd de l'Île-de-France (métros, RER, transilien). Nous n'avons pas de données décrivant les stations de bus de ville. Ainsi, dans les zones rurales de Seine-et-Marne par exemple, des actifs déclarant utiliser les transports en commun pour aller travailler peuvent être localisés à plusieurs dizaines de kilomètre d'une station de transport en commun lourd. Ils utilisent probablement un bus. D'un point de vue technique cela n'est pas gênant dans l'algorithme de ventilation. La ventilation respecte tout de même la commune de travail déclarée et le secteur d'activité. En revanche la distance mesurée entre le lieu de travail et la station de travail n'est pas cohérente et n'est pas à considérer dans les analyses. C'est pourquoi nous analysons ici seulement les distances des actifs travaillant dans la petite couronne francilienne (illustration 32 et tableau 24). Dans ce type d'espace nous supposons qu'il est plus cohérent d'analyser les distances aux stations de transport en commun dit lourd. Nous enlevons Paris de l'analyse en raison de son exceptionnel niveau de desserte en transport en commun.

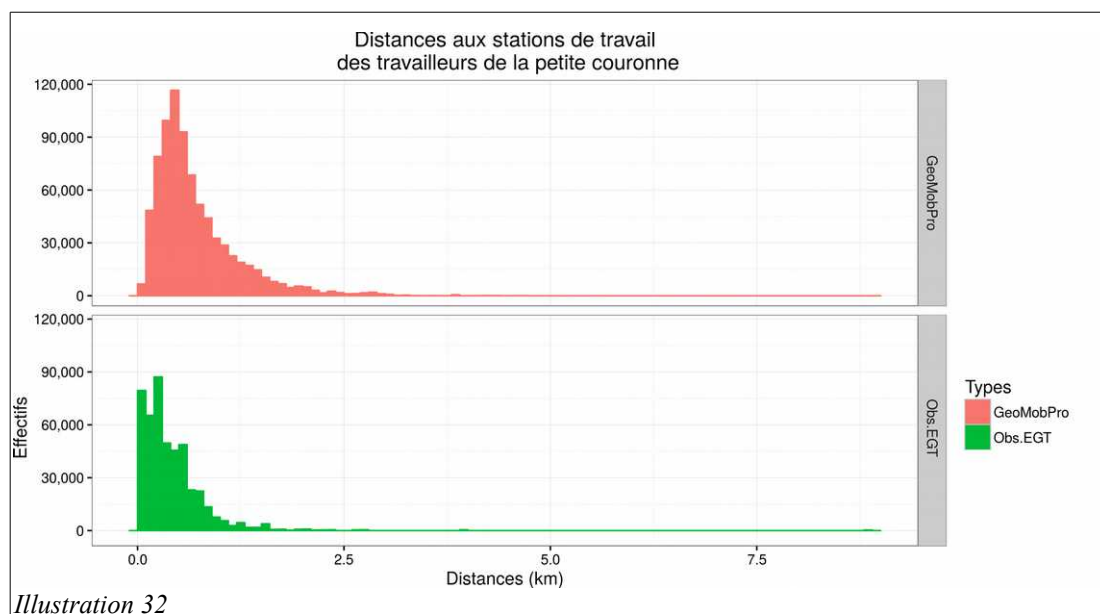


Illustration 32

Décile	0 %	10 %	20 %	30 %	40 %	50 %	60 %	70 %	80 %	90 %	100 %
GeoMobPro	0.06	0.24	0.32	0.40	0.47	0.54	0.65	0.79	1.00	1.36	5.98
Obs. EGT	0.06	0.06	0.10	0.14	0.22	0.30	0.36	0.50	0.58	0.80	8.86

Source : Données du recensement désagrégées. IFSTTAR DEST/ LVMT, Insee 2012, EGT 2010

Pour analyser les distances entre le lieu de travail et la station de transport en commun la plus proche à vol d'oiseau nous comparons la distribution des distances issue de la ventilation avec les résultats de l'Enquête Globale Transport d'Île-de-France. Sur le tableau 24 les premiers déciles des observations du recensement désagrégées (GeoMobPro) sont supérieurs aux résultats de l'EGT. Mise à part les très petites distances la suite de la distribution des distances a une allure d'une courbe exponentielle décroissante.

Toutefois, lorsque nous ré-agrégeons les résultats de la ventilation sur le carroyage de 1km nous obtenons une meilleure représentativité des petites distances (tableau 25). Mesurées avec les coordonnées à l'adresse les observations du premier décile des observations parcourent moins de 240 m et après ré-agrégation ils parcouraient moins de 160 m. Alors que les résultats de l'EGT indiquent que le premier décile des observations parcourent des

distances inférieures, proche de 50m⁹⁰.

Tableau 25 : Distributions des déciles des distances sur le carroyage de 1km entre les stations de travail et les lieux de travail en petite couronne (km)											
Décile	0 %	10 %	20 %	30 %	40 %	50 %	60 %	70 %	80 %	90 %	100 %
GeoM obPro	0.00	0.16	0.24	0.29	0.35	0.39	0.43	0.47	0.50	0.55	1.13
Obs. EGT	0.06	0.06	0.10	0.14	0.22	0.30	0.36	0.50	0.58	0.80	8.86

Source : Données du recensement désagrégées. IFSTTAR DEST/ LVMT, Insee 2012, EGT 2010

Les distances domicile-travail en marche à pied :

De la même manière que pour les distances entre le lieu de travail et les établissements économiques la ventilation des lieux d'emploi des actifs occupés du recensement déclarant aller travailler à pied a du mal à respecter les premiers déciles des distances (tableau 26). Les plus grandes distances sont quant à elles mieux respectées.

90 La valeur 0,06 indique que le lieu de travail et la station emprunté sont dans le même carreau de 100m.

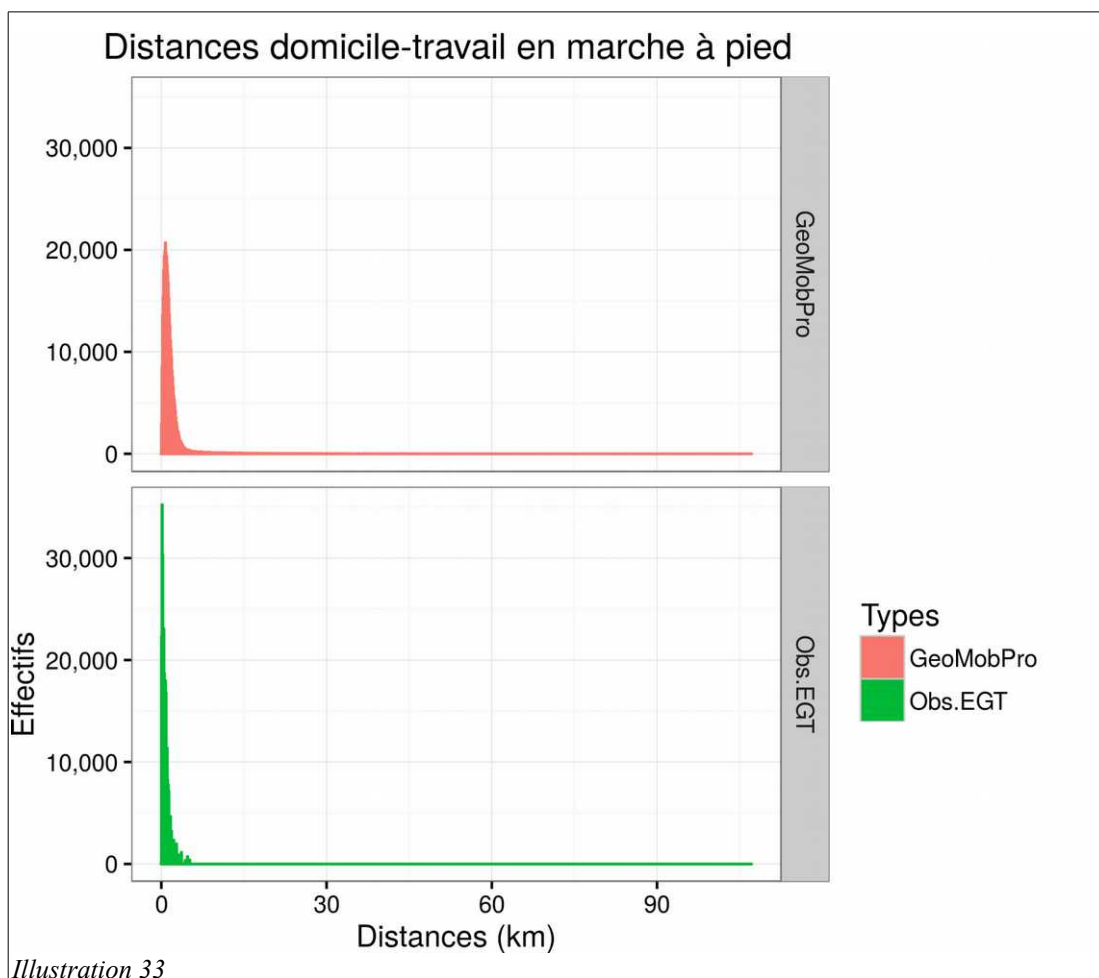


Illustration 33

Tableau 26 : Distributions des déciles des distances effectuées en marche à pied pour aller travailler par l'ensemble des franciliens

Décile (km)	0 %	10 %	20 %	30 %	40 %	50 %	60 %	70 %	80 %	90 %	100 %
GeoMobPro	0.0	0.38	0.59	0.77	0.97	1.17	1.40	1.69	2.12	3.00	107.00
Obs. EGT	0.0	0.10	0.14	0.28	0.40	0.51	0.67	0.82	1.12	1.61	5.11

Source : Données du recensement désagrégées. IFSTTAR DEST/ LVMT, Insee 2012, EGT 2010

Nous remarquons que dans les deux enquêtes les distances en marche à pied peuvent être de plusieurs kilomètres. Dans le cas des données du recensement nous retrouvons le manque de connaissance sur les sans déplacements et la bi-résidentialité (chapitre 3.2.4). Pour les données EGT nous avons quelques observations parcourant 5 kilomètres. Par exemple, une observation de l'EGT déclare habiter vers le quai de Bercy et aller travailler dans le 1^{er} arrondissement de Paris. Cette observation parcourt plus de 5 kilomètres à pied. On peut se

demander si cette distance est cohérente, si elle est parcourue tous les jours. N'ayant aucune donnée de contrôle pour comprendre ces grandes distances déclarées effectuées en marche à pied nous ne pouvons pas les traiter.

Après ré-agrégation les résultats de la ventilation des petites distances sont plus proches du premier décile de l'EGT (tableau 27). Pour 10 % des observations du recensement marchant pour aller travailler les distances sont inférieures à 1 km. Bien entendu il s'agit de distances infra-zone. La valeur indiquée dans le tableau est de 0 mais nous savons qu'il s'agit de distances en réalité inférieures à 1km.

Tableau 27 : Distributions des déciles des distances au carreau de 1km effectuées en marche à pied pour aller travailler de l'ensemble des franciliens											
Décile (km)	0 %	10 %	20 %	30 %	40 %	50 %	60 %	70 %	80 %	90 %	100 %
GeoMobPro	0	0	0.99	0.99	1.00	1.41	1.41	1.99	2.23	3.16	107.31
Obs. EGT	0	0.10	0.14	0.28	0.40	0.51	0.67	0.82	1.12	1.61	5.11
Source : Données du recensement désagrégées. IFSTTAR DEST/ LVMT, Insee 2012, EGT 2010											

Pour les plus grandes distances l'opération de ré-agrégation n'apporte pas de meilleur résultats.

Conclusion du chapitre 4

Nous venons d'exposer les étapes du processus de désagrégation spatiale des données du recensement appliqué à l'Île-de-France et de procéder à des éléments d'évaluation.

La première étape du processus de désagrégation consistant à appairer les fichiers détail MOBPRO et INDCANTVILLE du recensement apporte une plus grande description des profils socio-économiques des navetteurs domicile-travail. Notamment nous avons une plus grande description des secteurs d'activité des actifs. D'une variable décrivant les secteurs d'activité en 5 postes nous avons obtenu une description en 17 classes. De plus, cette étape a permis d'affiner les localisations résidentielles des résidents des communes de 10 000 habitants et plus. De la commune de résidence nous sommes passés à l'IRIS de résidence.

La deuxième étape du processus de désagrégation consiste à ventiler les lieux de résidence des observations du support IRIS-Communes sur le carroyage de 200m de l'Insee. Pour ce faire, nous avons utilisé une seule variable de contrôle : l'effectif de ménage. Nous nous sommes ensuite interrogés sur la distribution des variables de contrôle non prise en compte dans la ventilation sous contrainte. La répartition spatiale des profils socio-économiques semble de bonne qualité. Toutefois les répartitions des variables socio-économiques sont de meilleures qualités lorsque nous ré-agrégeons les résultats sur le support carroyé INSPIRE de 1km. En particulier pour les variables décrivant le type de logement (logement individuel ou collectif). Cela nous fait penser que la distribution des secteurs d'activité au lieu de résidence, qui est aussi non considérée dans la ventilation sous contraintes, a un certain niveau de représentativité statistique.

La troisième étape a ventilé les lieux de travail des actifs occupés résidant en Île-de-France. Nous avons également procédé à des éléments d'évaluation de la répartition spatiale des localisations de travail produites. La ré-agrégation au carroyage INSPIRE de 1km permet de retrouver une meilleure représentativité statistique.

Notons tout de même que la reproductibilité de cette méthode est limitée par les changements de procédé de diffusion des données carroyées de l'Insee (encadré 1). Nous avons utilisé les effectifs de ménages renseignés par le carroyage de 200m diffusé en janvier 2013. Aujourd'hui, cette information est disponible sur un support rectangulaire. Nous avons utilisé les données du carroyage de 200m pour des raisons de commodité. Pour reproduire cette

méthode avec les données rectangulaires il faudrait prévoir une projection des effectifs de ménages sur la grille de 200m. Toutefois, en cas d'accord avec l'Insee et du comité du secret statistique les bases de données internes et confidentielles peuvent être utilisées. Dans ce cas il faut s'assurer que le fichier produit serait anonyme. Ce travail dépasse le cadre de cette thèse.

Nous allons procéder maintenant à des analyses descriptives de la mobilité domicile-travail en nous appuyant sur les données désagrégées que nous venons de produire.

Chapitre 5 : Exploitations des données de mobilité désagrégées du recensement sur le support carroyé de 1km

Chapitre 5 : Exploitations des données de mobilité désagrégées du recensement sur le support carroyé de 1km

Ce dernier chapitre propose quelques exploitations pouvant être réalisées avec les données du recensement localisées sur le carroyage INSPIRE de 1km. Nous insistons sur le fait que l'objet de ce chapitre n'est pas d'effectuer les analyses complètes des cas d'étude proposés. La littérature sur l'emploi et l'habitat de la région Île-de-France est très riche. Nous partons de quelques grandes problématiques d'aménagement de la région Île-de-France décrites dans les schémas d'aménagements et cherchons à les illustrer avec nos données.

Brièvement, les questions d'aménagement parisien traitent du desserrement urbain à plusieurs échelles. La structure urbaine française est macrocéphale : l'espace français est dominé par le pôle parisien. Paris doit garder un statut de pôle économique de rang européen tout en ne freinant pas le développement de pôles urbains secondaires sur le reste du territoire national. A l'échelle régionale, les schémas d'aménagement ont comme objectif d'améliorer l'équilibre habitat-emploi. Aujourd'hui, les emplois sont trop concentrés dans l'espace francilien. De plus, les schémas d'aménagements mettent en évidence une disparité est-ouest. La petite couronne et la frange ouest de la grande couronne accueillent plus d'emplois et des populations plus aisés. Les objectifs des projets d'aménagement régionaux sont :

« une organisation polycentrique des espaces urbains, structurée par différents sites de dynamique urbaine, aux tailles et fonctions variées : les centres d'envergure européenne, les secteurs de redéveloppement de la proche couronne, les villes nouvelles et nouveaux sites d'urbanisation, les villes de la couronne rurale;

- une valorisation du milieu rural et une irrigation de l'agglomération par les espaces naturels ;

- une organisation des transports confortant les priorités de l'aménagement régional. » (DRIEA, Le SDRI de 1194 quel bilan ?, p.1⁹¹)

Pour tendre vers un meilleur équilibre habitat-emploi, des pôles secondaires pré-existants peuvent être renforcés par des politiques d'incitation financière qui favoriseraient

91 DRIEA, Le SDRI de 1194 quel bilan ? [en ligne] (Consultée le 15/06/2016) <http://www.driea.ile-de-france.developpement-durable.gouv.fr/IMG/pdf/151004-region-etat-cesr-pour-colloque-211004_cle53bb51.pdf>

l'implémentation des activités économiques privées, et par l'implémentation de service public qui inciterait au développement de ces pôles dits secondaires. L'identification de ces pôles sur lesquels ancrer des politiques d'aménagement est une condition pour la réussite des projets d'aménagement. Il faut également ajouté que le développement du polycentrisme de la région doit s'appuyer sur des infrastructures de transport propres.

C'est dans l'esprit de ces grandes questions d'aménagement du territoire, que nous présentons quelques types d'application de nos données de mobilité désagrégées au carreau de 1km.

Le premier sujet abordé est l'observation des petites distances, puisque les politiques régionales mettent l'accent sur les liens de proximité. Nous commençons par nous demander quel est l'apport du support carroyé de 1km pour la mesure des distances. Par construction, nous obtenons une meilleure connaissance des petites distances comparé à l'utilisation du support communal. Nous poursuivons en examinant les couples domicile-travail infra-communaux et les couples dont l'origine et la destination se situent sur deux communes limitrophes (5.1).

Après avoir étudié les différences de mesure des distances domicile-travail selon les choix de support, nous reprenons la proposition du CNIS sur la définition d'indicateurs de distances domicile-travail. Cependant, au lieu des distances médianes par carreau, nous mesurons le premier et dernier décile des distances domicile-travail. Les carreaux générant de relativement petits et grands déplacements sont ainsi cartographiés (5.2).

Pour analyser les espaces recouverts par les carreaux nous cartographions ensuite un indicateur d'attraction-répulsion. Nous faisons la différence entre le nombre d'actifs travaillant dans les carreaux et le nombre d'actif habitant dans les carreaux (5.3).

Les pôles d'emploi secondaires étant relativement plus petits que les centres d'envergures européens il est utile d'effectuer des zooms géographiques pour les étudier. A l'aide de trois cas d'études géographiques nous montrons en quoi les données de mobilité désagrégées apportent des éléments de connaissances nouveaux pour étudier de petits territoires. Par exemple, nous effectuons le portrait d'une zone aéroportuaire. Ce type d'espace accueille de nombreux établissements économiques. Les données de mobilité professionnelle du recensement sur le support communal ne permettent pas d'identifier les flux attirés par

l'infrastructure de transport. Les données désagrégées permettent alors de découper le territoire à façon et de zoomer sur l'infrastructure aéroportuaire.

Nous présentons également deux autres cas d'étude. Nous choisissons d'étudier un espace paupérisé (la Plaine-Saint-Denis) et un espace gentrifié (Vincennes et le quartier du Bas-Montreuil). Ces deux zones sont situées à cheval sur plusieurs communes (5.4).

Enfin, nous nous intéressons à la problématique de la planification des transports. Dans cette dernière section nous montrons l'apport du choix du carroyage de 1km pour déterminer les périmètres restreints nécessaires aux politiques d'aménagement de transport. Nous prolongeons cette analyse en construisant une matrice de flux, afin d'identifier des zones où il peut être mené des politiques d'incitation à la mutualisation des moyens de transport (co-voiturage et transports en commun) (5.5).

5. 1 Analyse des distances domicile-travail à vol d'oiseau selon le choix du support

5.1.1 Les distances domicile-travail à vol d'oiseau

Le premier apport des données désagrégées est de pouvoir mesurer des petites distances. Nous commençons donc par observer les différences de mesures de distances selon plusieurs supports spatiaux (tableau 28). Nous comparons les distances domicile-travail de l'EGT supportant l'information sur un carroyage de 100m. Puis, nous les comparons avec les distances des données du recensement mesurées sur le support de construction des données désagrégées⁹², le support carroyé de 1km, et le support communal avec comme point de référence les centroïdes et les chefs-lieux déterminés par l'IGN. Les indicateurs de comparaison sont les déciles des distances. En effet, les distributions des distances sont rarement des distributions gaussiennes. Les moyennes résumant des distributions trop diverses pour être suffisantes.

Tableau 28 : Comparaison des mesures des distances domicile-travail selon le choix du support											
Déciles des distances (km)	Min	10 %	20 %	30 %	40 %	50 %	60 %	70 %	80 %	90 %	Max
EGT (100m) (9024 obs)	0	0.92	2.02	3.42	5.04	6.90	9.21	12.23	16.44	23.30	88.11
Carroyage (200m etX/Y)	0	0.84	1.49	2.46	4.06	6.06	8.39	11.32	15.43	22.45	127.36
1km	0	0.99	1.41	2.23	4.12	6.07	8.25	11.19	15.29	22.37	127.89
Centroïdes communes	0	0	0	2.23	4.01	6.13	8.45	11.45	15.60	22.68	127.43
Chefs-lieux IGN	0	0	0	3.11	5.17	7.40	9.97	13.13	17.33	24.44	128.97
<i>Effectif total 5 140 550 actifs occupés avec 1 930 206 observations</i>											
<i>Source : Données du recensement désagrégées. IFSTTAR - DEST/LVMT, Insee 2012, EGT 2010</i>											

92 Le carroyage de 200m pour les lieux de résidence et les coordonnées X/Y pour les lieux de travail.

Lorsque les distances sont à zéro nous ne connaissons pas les mesures avec précision, nous savons qu'elles sont inférieures aux « rayons moyens » de la maille du support. Par construction, plus le support est fin, plus nous obtenons une meilleure connaissance des petites distances. Avec le support communal nous ne connaissons plus les distances pour 30 % des observations d'actifs occupés du recensement. Le dossier complet de la région Île-de-France diffusé sur le site de l'Insee indique bien que 29,4 % des actifs travaillent dans leur commune de résidence en 2012 (Insee. Dossier complet de l'Île-de-France⁹³). À partir d'environ 2,2km les mesures des petites distances ne peuvent plus être effectuées à l'aide du support communal car les distances concernent alors majoritairement des déplacements infra-communaux. Au chef-lieu nous retenons 3,1km comme valeur seuil.

93 Insee. Dossier complet de l'Île-de-France. [en ligne] (Consultée le 25/06/20156)
<http://www.insee.fr/fr/themes/dossier_complet.asp?codgeo=REG-11>

5.1.2 Les distances infra-communales

Les distances domicile-travail internes aux communes rentrant dans la catégorie des petites distances, nous effectuons le même travail d'analyse des distributions des déciles des distances domicile-travail à vol d'oiseau (tableau 29).

Tableau 29 : Comparaison des mesures des distances infra-communales selon le choix du support											
Déciles des distance (km)	Min	10 %	20 %	30 %	40 %	50 %	60 %	70 %	80 %	90 %	Max
<i>EGT (100m)</i> (1298 obs)	0.0	0.10	0.22	0.41	0.58	0.76	0.94	1.20	1.55	2.00	7.21
Fin (200m etX/Y)	0.0	0.36	0.56	0.73	0.91	1.09	1.29	1.53	1.84	2.33	11.24
1km	0.0	0.0	0.99	0.99	1.00	1.00	1.41	1.41	2.00	2.23	11.98
Centroides communes	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Chefs-lieux IGN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Effectif total : 1 381 087 actifs occupés avec 521 239 observations Source : Données du recensement désagrégées. IFSTTAR DEST/ LVMT, Insee 2012, EGT 2010											

Les deux supports (de 200m et celui des adresses postales) sur lesquels nous avons ventilés les données du recensement permettent de retrouver des distances infra-communales. C'est également le cas pour le support de 1 km qui agrège cependant les distances inférieures à 1 km. Au moins 10 % des observations du recensement effectuant des trajets internes aux communes sont concernées par les distances infra-carreaux. Les déplacements entre carreaux se juxtaposant (par construction, les distances de ces déplacements sont de 1km ou de 1,4km) concernent près de 70 % des observations désagrégées travaillant dans leur commune de résidence.

L'illustration 34 cartographie les couples domicile-travail infra-communaux sur le carroyage de 1km.

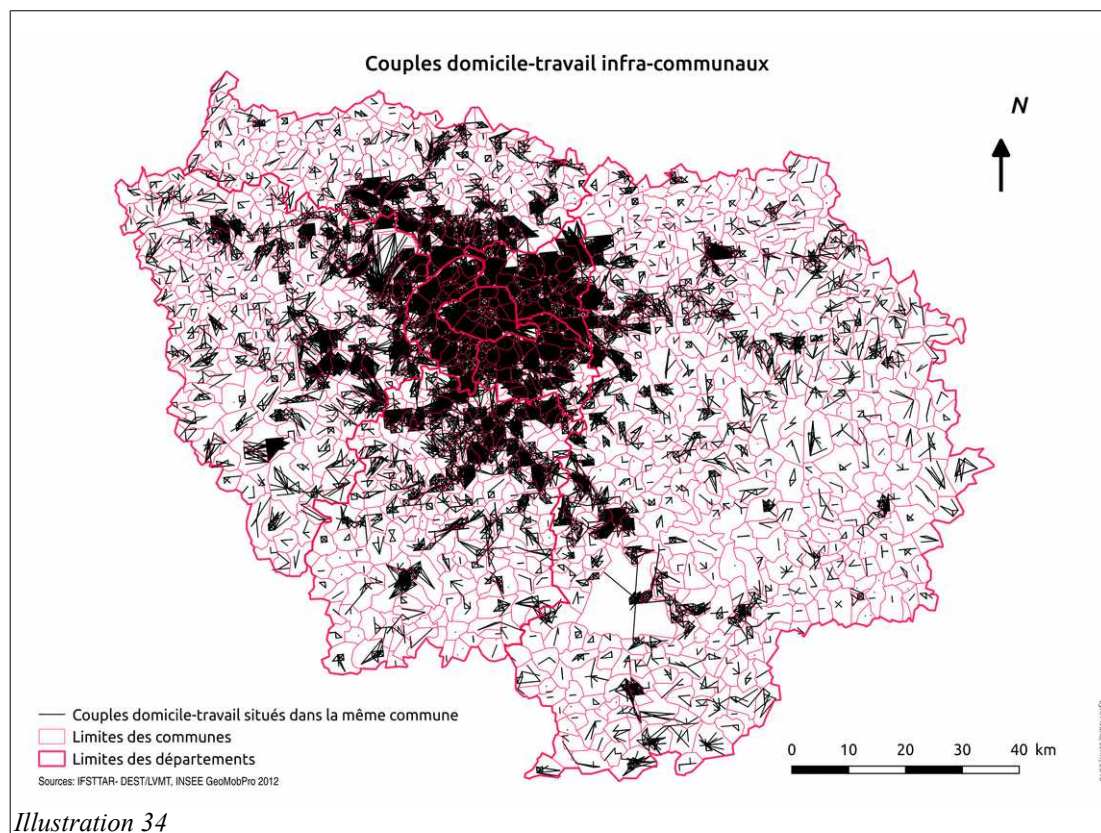


Illustration 34

Nous observons bien la centralité de Paris et de sa petite couronne. La cartographie des déplacements internes aux communes permet également de mettre en lumière des zones plus périphériques à Paris et sa petite couronne mais générant de nombreux déplacements infra-communaux. Il est encore trop tôt pour définir s'il s'agit de pôles secondaires sur lesquels ancrer des politiques d'aménagement.

Continuons d'abord nos analyses sur les petites distances.

5.1.3 Les déplacements entre communes limitrophes

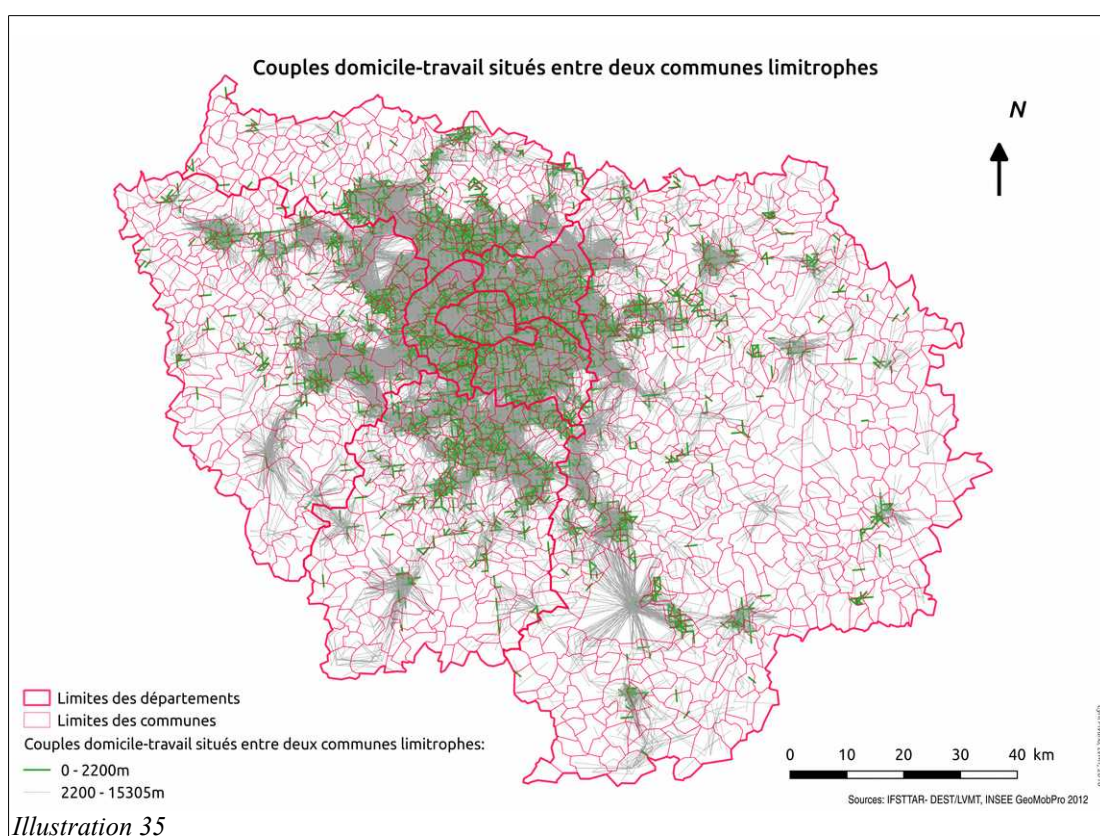
Examinons maintenant les distances parcourues entre deux communes limitrophes. Les localisations domicile-travail concernant des communes ayant des limites communales se juxtaposant peuvent cacher deux réalités géographiques. Soit il s'agit d'une conurbation urbaine, et les tâches de bâti des deux communes se rejoignent. Il s'agit alors d'une seule et même agglomération au sens de l'Insee. Soit il s'agit d'espaces plus ruraux, les centres bâtis des deux communes sont éloignés, et nous sommes en présence de deux agglomérations. Dans ce cas les distances domicile-travail effectuées pour rejoindre les deux communes sont alors suffisamment grandes pour être observées avec un support relativement grossier. Mais s'il s'agit de conurbation urbaine alors l'utilisation d'un support grossier va rallonger les distances. Nous avons besoin d'un support fin pour étudier les conurbations urbaines.

Déciles des distance (km)	Min	10 %	20 %	30 %	40 %	50 %	60 %	70 %	80 %	90 %	Max
EGT (100m) (1283 obs)	0.1	0.11	1.52	1.81	2.14	2.46	2.88	3.32	3.77	4.80	34.65
Fin (200m etX/Y)	0.0	1.32	1.75	2.10	2.42	2.75	3.12	3.54	4.10	4.97	16.06
1km	0.0	1.41	1.99	2.23	2.23	2.83	3.16	3.60	4.12	5.00	15.30
Centroïdes communes	0.2	1.74	2.08	2.33	2.64	2.86	3.10	3.42	3.70	4.30	13.11
Chefs-lieux IGN	0.45	1.62	1.91	2.22	2.47	2.73	3.05	3.32	3.78	4.54	15.61
Effectif total 609 688 actifs occupés avec 230 266 observations Source : Données du recensement désagrégées. IFSTTAR DEST/ LVMT, Insee 2012, EGT 2010											

Le support du carroyage de 1km donne une valeur minimale de 0. Il y a donc des observations habitant et travaillant dans un même carreau lui-même traversé par une limite communale. Mais les distances entre communes limitrophes sont tout de même plus grandes que celles des distances infra-communales. Le premier décile *trans-communal* est de 1,4km

celui des distances infra-communales est de 0km. Il y a donc beaucoup moins de petits déplacements que dans le cas des distances infra-communales. En effet, dans le cas de conurbation urbaine nous savons que ce sont des espaces urbanisés sans qu'il s'agisse des centres historiques (en minorant l'espace de Paris et ses communes limitrophes). Ces espaces moins attractifs que les centre-villes concentrent moins de très petites distances infra-communales.

De manière générale, les petites distances limitrophes permettent d'observer les franges urbaines se rejoignant, et donc les espaces ayant une certaine attractivité. L'illustration 35 met en lumière les conurbations urbaines en distinguant les petits et grands déplacements. Nous reprenons la valeur seuil de 2,2km du tableau 28.



5.2 Cartographie des lieux générant de grands et de petits déplacements

Nous allons maintenant analyser les petits et grands déplacements domicile-travail en spatialisant l'information. Nous analysons les territoires qui génèrent des déplacements domicile-travail sous l'angle de l'emploi : c'est-à-dire que nous nous intéressons aux pôles d'emploi et que nous essayons de déterminer si ces pôles génèrent de grands ou petits déplacements. Pour chaque carreau de 1km nous collectons donc les distances des déplacements ayant pour destination ces carreaux. Nous calculons ensuite le premier décile pour observer les lieux générant de relatifs petits déplacements (illustration 36) puis le 9^o décile pour observer les lieux générant de plus grands déplacements (illustration 37).

La carte du premier décile : Les lieux économiques générant de petits déplacements

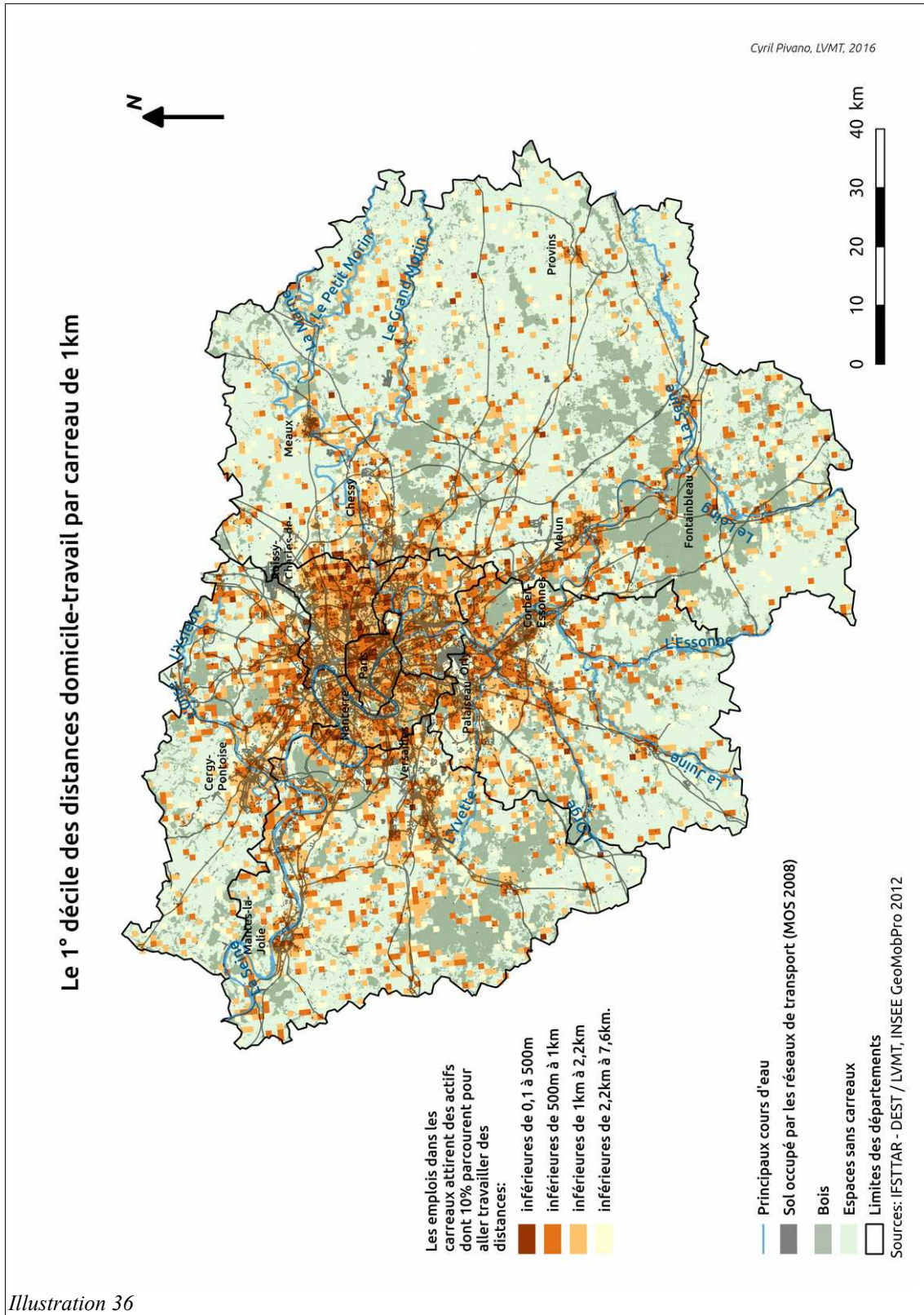


Illustration 36

Nous observons comme attendu que les zones denses (le cœur de l'agglomération parisienne) génèrent de moins grandes distances que les zones peu denses. Cependant, la finesse du support spatial permet de faire apparaître que les arrondissements du nord-est parisien génèrent relativement plus de petites distances que la rive droite aval (9^e arrondissement). Nous retrouvons une disparité est-ouest au sein de la petite couronne. Les emplois du département de Seine-Saint-Denis et du Val-de-Marne génèrent aussi de plus de petites distances que les emplois des Hauts-de-Seine. Les actifs occupés des Hauts-de-Seine habitent plus loin de leur emploi que les actifs occupés travaillant en Seine-Saint-Denis, bien que les emplois soient plus nombreux à l'ouest de Paris.

En grande couronne les lieux générant de petits déplacements sont les noyaux d'urbanisation historiques (Meaux, Melun, Fontainebleau, Mantes-la-Jolie). Les villes-nouvelles n'apparaissent pas clairement comme générant de petits déplacements. Enfin, les emplois des carreaux éloignés des petits noyaux de la grande couronne génèrent de plus grands déplacements. Toutefois faisons attention aux faibles effectifs dans ces territoires dans l'interprétation des déciles.

Nous effectuons désormais le même travail de cartographie pour le 9^e décile des distances domicile-travail. Nous observons maintenant les carreaux dont les actifs occupés parcourent de relatives grandes distances pour rejoindre leur emploi (illustration 37).

La carte du neuvième décile : Les lieux économiques générant de grands déplacements

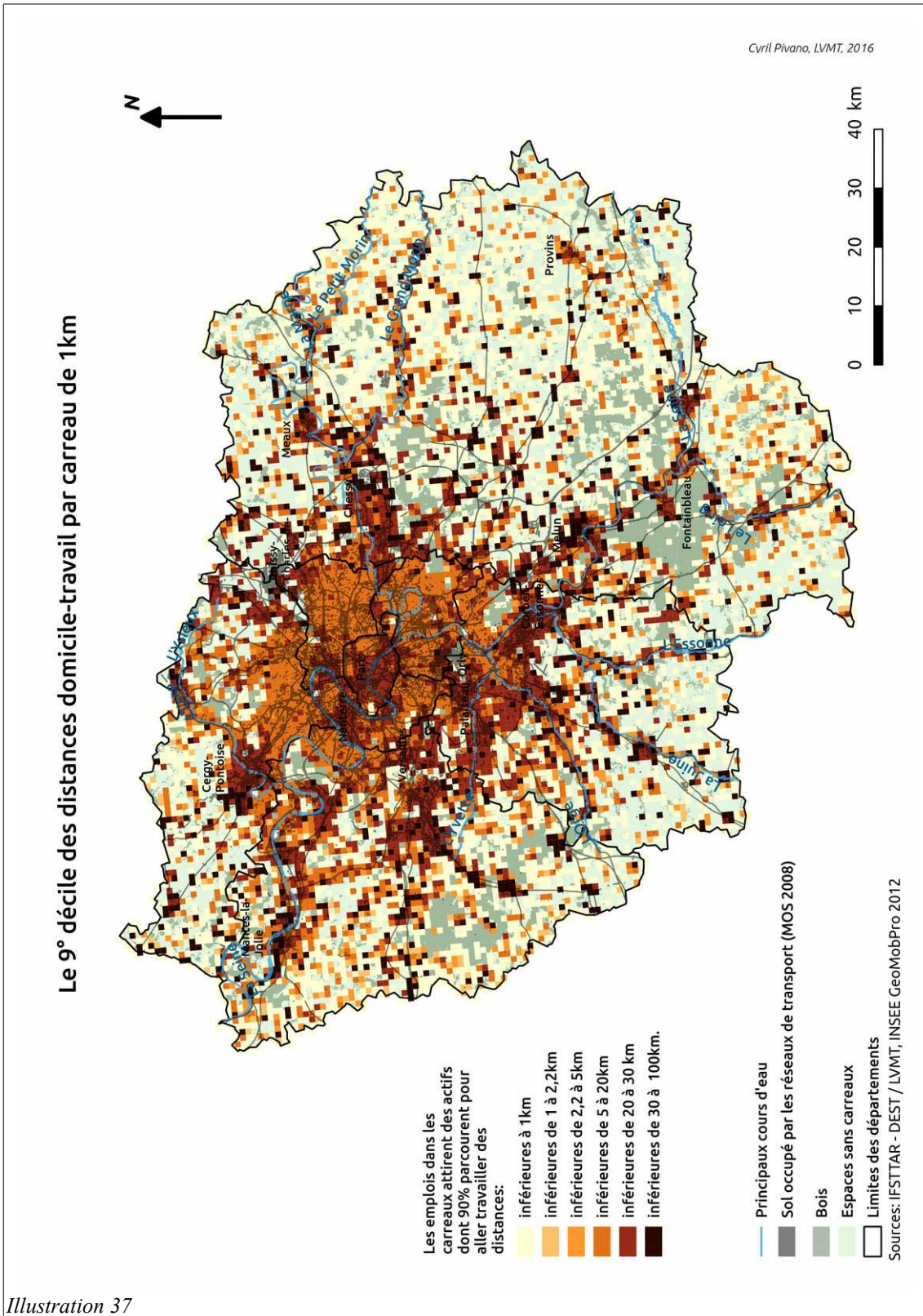


Illustration 37

La carte du 9° décile est plus lisible, les carreaux de même catégorie ont tendance à être regroupé spatialement. Trois zones peuvent être délimitées :

- le cœur de l'agglomération parisienne et la plaine de Nanterre (marron).
- Un anneau encerclant cette première zone. Cet anneau est plus épais que les départements de la petite couronne.
- Puis des radiales le long des principaux axes d'urbanisation au long des vallées (principaux cours d'eau).

Avec la carte du premier décile nous avons vu que les petits déplacements internes à Paris sont plus petits qu'ailleurs. La commune de Paris (exception faite des arrondissements du nord-est) génère également de plus grands déplacements, ainsi que la zone d'emploi de la Défense.

Nous observons ensuite qu'en petite couronne élargie les grands déplacements sont plus petits (anneau orange sur la carte). Nous en déduisons que Paris attire des actifs de toute part de la région Île-de-France alors que les emplois de la petite couronne ont des bassins de main d'œuvre relativement de plus petites taille.

En grande couronne, nous distinguons les zones plus urbanisées le long des cours d'eau et des axes de transport. L'indicateur du 9° décile met en lumière les pôles d'emploi plus secondaires de la région (Fontainebleau, Melun, Corbeil-Essone, Orly-Rungis, Massy-Palaiseau, Versailles, Mantes-la-Jolie, Cergy-Pontoise, la plaine Saint-Denis, Roissy-Charles-de-Gaulle, la vallée de la Marne de Paris à Chessy-Marne-la-Vallée, Meaux et Provins). Ces pôles comprenant cette fois-ci les villes-nouvelles attirent des actifs venant de relativement loin.

Si l'indicateur du 9° décile donne une information sur les zones d'emploi attirant de loin des travailleurs nous ne connaissons en revanche par leur provenance.

Pour délimiter plus finement les pôles d'emploi de la région et repérer les zones émettant des travailleurs, nous utilisons un autre indicateur : le différentiel d'attraction-répulsion.

5.3 Cartographie des carreaux d'attraction et d'émission

Pour chaque carreau nous faisons la différence entre les effectifs d'actifs occupés habitant les carreaux et les effectifs travaillant dans les carreaux. Nous mettons en lumière les carreaux attirant plus qu'ils n'émettent d'actifs. Ces carreaux ont une fonction économique. Les carreaux émettant plus qu'ils ne reçoivent d'actifs ont une fonction plus résidentielle.

Cet indicateur ne peut pas être utilisé avec le support des communes. La maille doit être suffisamment petite pour différencier les espaces résidentiels et économiques.

La cartographie de cet indicateur (et la discrétisation choisie) met en lumière la concentration des lieux d'activité de la région (illustration 38). Les zones résidentielles sont bien étalées dans l'espace francilien. La carte est à lire avec les éléments de connaissances des cartes des premiers et des derniers déciles.

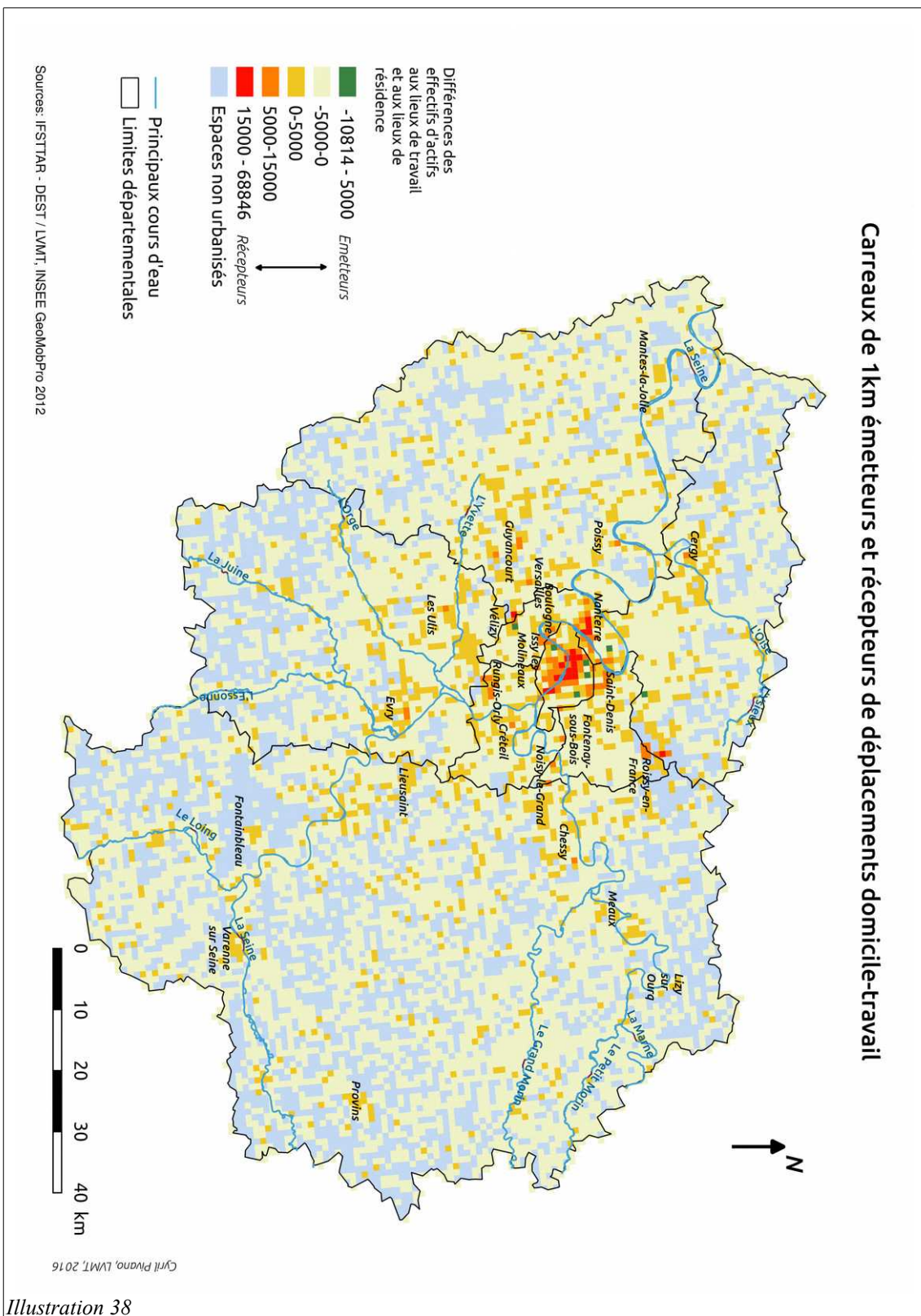


Illustration 38

La zone d'emploi principale de la région Île-de-France est la rive droite parisienne, et dans une moindre mesure sa rive gauche. C'est à l'aide de la carte du premier et dernier décile que nous observons que la zone du 9^e arrondissement génère de plus grands déplacements que la rive droite amont (10^e et 11^e arrondissement). Les actifs de la rive droite amont habitent et travaillent relativement plus au même endroit que ceux de la rive droite aval. La rive droite aval a une spécialisation économique bien plus accentuée que la rive droite amont. Les actifs y travaillant parcourent de plus grandes distances pour aller travailler.

Nous retrouvons ce phénomène de spécialisation des territoires dans la plaine de Rueil-Malmaison et de Nanterre. Cette zone génère de grands déplacements mais il ne s'agit pas d'un espace ayant une seule fonction. Ces espaces sont à différencier avec notre indicateur sur le différentiel émetteurs-récepteurs. La plaine de Rueil-Malmaison est bien plus une zone résidentielle que la zone économique de Nanterre (la Défense). Les zones sont proches mais génèrent toutes deux de grandes distances. Nous en déduisons qu'il n'y a pas de surreprésentation des résidents de Rueil-Malmaison travaillant à la Défense.

En continuant notre analyse des pôles d'emploi de la région, les zones aéroportuaires (Roissy-en-France, Orly) sont des zones génératrices de grandes distances en raison de leur forte concentration en emplois. Leur aire d'attraction sont grandes, ces activités économiques ne peuvent, en effet, employer une main-d'œuvre plus ou moins qualifiée à très petites distances. La plaine Saint-Denis est également dans la même situation. C'est une zone accueillant de nombreux emplois et générant de grandes distances domicile-travail. Les actifs y travaillant viennent relativement de loin. Mais contrairement aux zones aéroportuaires, cette zone est pourtant très proche de la poche de main d'œuvre parisienne, elle génère toutefois de grands déplacements.

Nous observons ensuite une série de pôles d'emploi en première couronne (Fontenay-sous-Bois, Noisy-le-Grand, Créteil, Vélizy et le plateau de Saclay, Guyancourt, Versailles et Boulogne). Toutes ces zones sont génératrices de grandes distances que ce soit mesuré à l'aide du premier ou du dernier décile. De manière générale les zones d'emploi du cœur de Paris et les secteurs de redéveloppement de la proche couronne génèrent de grands déplacements.

En grande couronne les zones d'emploi sont de plus petites tailles, ils sont constitués de moins de carreaux. Avec les cartes des déciles nous avons vu que ce sont des zones génératrices de grands déplacements. Les emplois de ces espaces sont occupés par des actifs

parcourant de plus grandes distances.

Tout en commençant à s'interroger sur les inégalités socio-spatiales en Île-de-France examinons les zones ayant une fonction relativement plus résidentielle (carrés verts). La carte montre que les carreaux de 1km accueillant de nombreux actifs et peu d'emplois sont peu nombreux. Il y a quelques carreaux au nord de Paris (17°, 18° et 19° arrondissements), un au nord de la boucle d'Asnières, un à Vélizy, un à Sarcelles (sud du 95), un dans le 15° arrondissement de Paris. Bien sûr un choix de discrétisation différent pourrait mettre en lumière d'autres carreaux. En élargissant le pas de discrétisation, il apparaîtrait alors d'autres carreaux déclarés comme fortement résidentiels.

Que cachent les carreaux ayant une propension à émettre plus d'actifs occupés qu'ils n'en réceptionnent comme réalité géographique ? Les carreaux de Sarcelles et du 15° arrondissement nous laisse penser que les structures géographiques de déplacements de ces territoires ne sont pas identiques. Pour en savoir plus il nous faut effectuer des zooms géographiques et disposer de profils socio-économiques pour catégoriser les territoires.

5.4 Zooms géographiques et discriminations socio-économiques

A l'aide des données traditionnelles du recensement sur le support communal nous pouvons effectuer des zooms géographiques pour étudier la mobilité. Toutefois nous ne pouvons pas descendre à une échelle infra-communale. Pourtant dans certains cas nous devons disposer de données plus fines pour pouvoir étudier de petits espaces. Ce besoin se fait également sentir lorsque nous souhaitons étudier des zones d'emploi à cheval sur plusieurs communes. Dans cette section nous illustrons l'apport des données désagrégées du recensement à travers trois cas d'études: la zone aéroportuaire d'Orly, la Plaine Saint-Denis et la zone de Vincennes, Saint-Mandé et du bas-Montreuil.

5.4.1 La zone aéroportuaire d'Orly et de Rungis

L'aéroport d'Orly a été inauguré en 1909 et est chronologiquement le premier aéroport du monde. Il est situé sur la plaine de Longboyau à cheval sur deux départements, l'Essonne (91) et le Val-de-marne (94). L'aéroport est placé sous l'autorité du préfet de ce dernier département, et est également à cheval sur neuf communes se partageant les taxes professionnelles des établissements économiques. De plus, l'aéroport a une piste et une aérogare dédiées au fret. L'établissement du marché international de Rungis en 1969 à proximité de l'aéroport renforce l'activité de fret d'Orly. C'est le plus grand marché agricole du monde. Le marché international est également à cheval sur deux communes (Rungis et Chevilly-Larue).

L'étude des déplacements générés par cette importante zone d'activité est problématique que ce soit avec l'utilisation des données du recensement ou avec les Enquêtes Globale Transport d'Île-de-France. Les données du recensement n'ont pas un support suffisamment fin pour délimiter la zone aéroportuaire. Les données EGT ont un taux d'échantillonnage trop faible et les analyses statistiques sont souvent réalisées en distinguant les déplacements des départements de la petite et de la grande couronne sur lesquelles l'aéroport est à cheval. Avec les données désagrégées du recensement nous pouvons délimiter à façon la zone aéroportuaire et construire en quelques sortes son territoire (espace approprié). L'illustration 39 est un exemple de découpage pouvant être effectué.

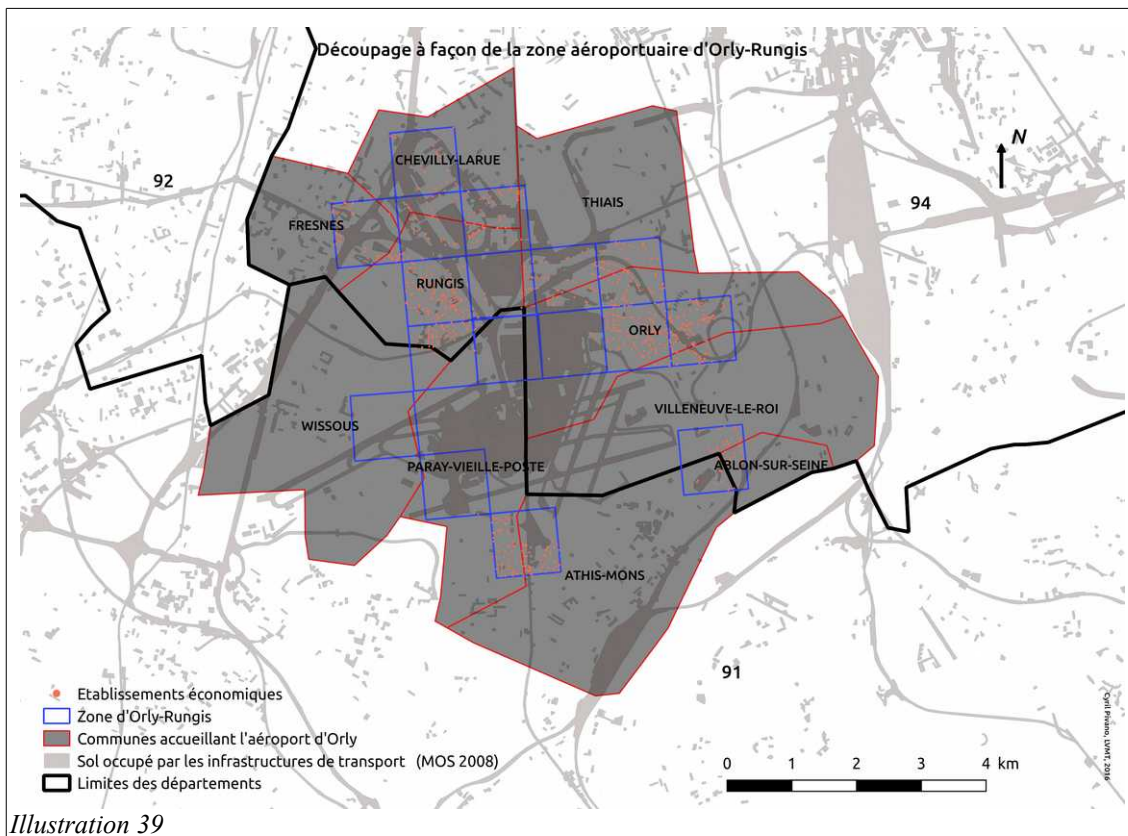
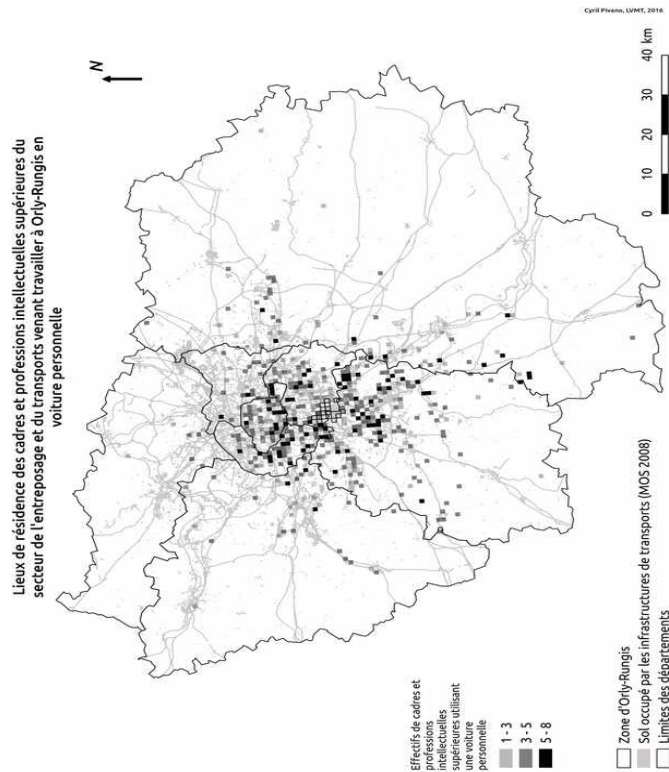
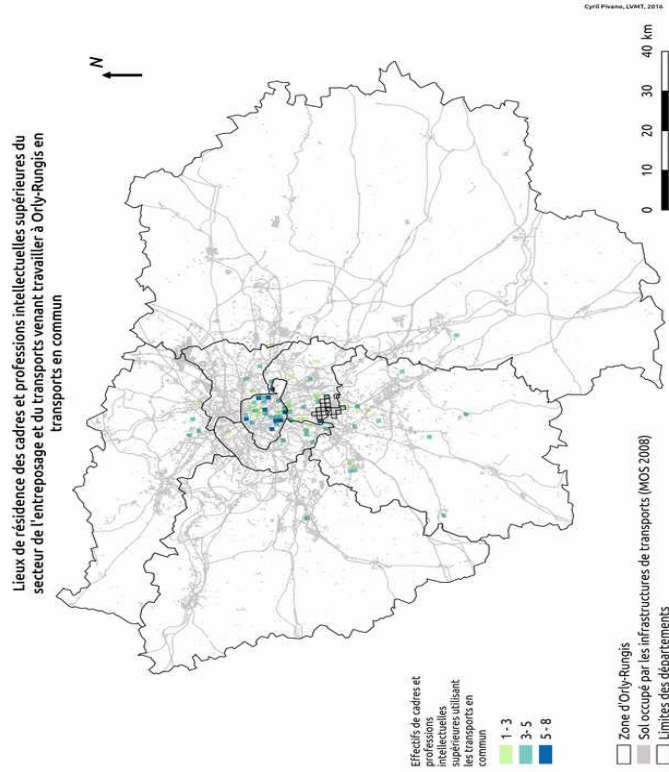


Illustration 39

À l'aide de ce découpage des secteurs d'activité en 17 postes, des catégories socioprofessionnelles et des modes de transport nous pouvons observer le bassin d'attraction de la zone aéroportuaire. Les planches cartographiques 40 et 41 illustrent un exemple de discrimination pouvant être effectuée. La planche met en opposition la répartition des lieux de résidence des cadres et professions intellectuelles supérieures travaillant dans le secteur de l'entrepôt et du transport avec celle des ouvriers, tout en distinguant les modes de transport.

Lieux de résidence des actifs du secteur du transport et de l'entreposage travaillant dans la zone d'Orly-Rungis :
les cadres et professions intellectuelles supérieures

Illustration 40



Source: Recensement 2012 désagrégé aux carreaux de 1km

Lieux de résidence des actifs du secteur de l'entreposage et de l'entreposage travaillant dans la zone d'Orly-Rungis : les ouvriers

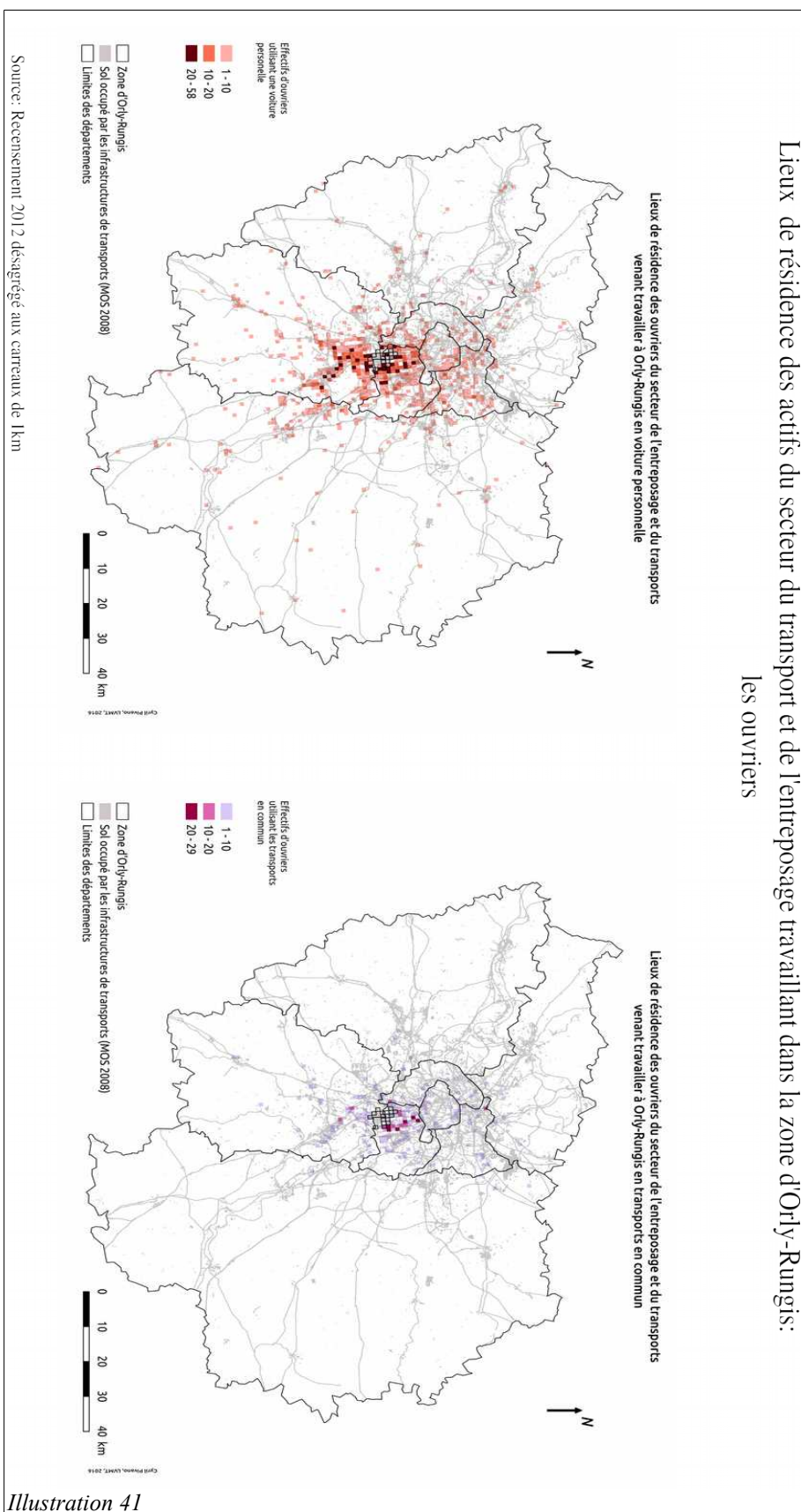
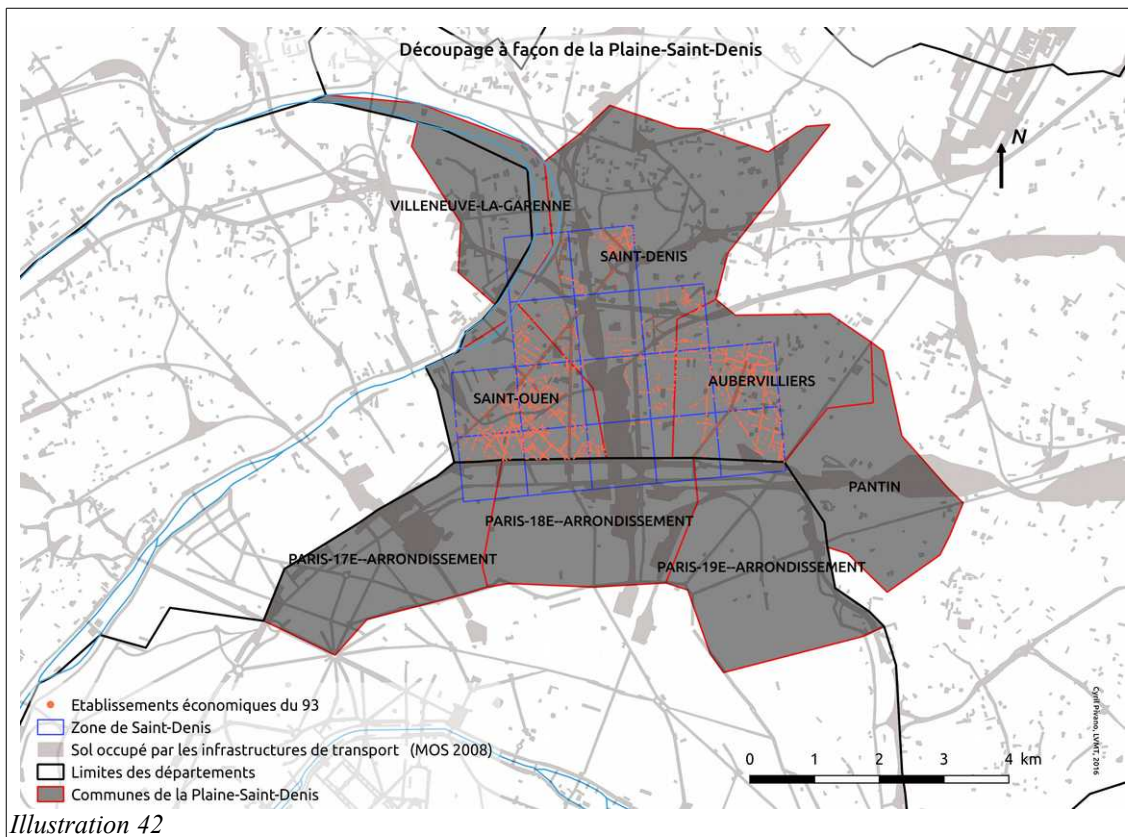


Illustration 41

Les lieux de résidence des cadres utilisant une voiture personnelle sont bien plus diffus que les lieux de résidence des cadres utilisant les transports en commun. Les cadres automobilistes proviennent de Paris, des Hauts-de-Seine, et du sud de l'Essonne. Les cadres utilisant les transports en commun habitent bien plus au sud de Paris à proximité des stations de transport desservant la zone aéroportuaire (RER B). En ce qui concerne les lieux de résidence des ouvriers travaillant dans le transport et l'entreposage, ils sont plus concentrés dans l'espace. Alors que les cadres automobilistes proviennent des Hauts-de-Seine, les ouvriers automobilistes proviennent de la Seine-Saint-Denis. Enfin, les ouvriers utilisant les transports en commun habitent dans les communes au sud de Paris, peut être parce que les prix du foncier sont plus accessibles pour eux.

5.4.2 La Plaine-Saint-Denis



La Plaine-Saint-Denis est un quartier du département de Saint-Denis partagé entre trois communes : Saint-Denis, Aubervilliers, et Saint-Ouen. Cet espace fut urbanisé au 19^e siècle en accueillant des fonctions industrielles. Après la seconde guerre mondiale cet espace devient une friche industrielle. En 1985, il est créé le syndicat intercommunal « Plaine Renaissance » devant transformer cette friche en une ville active. Aujourd'hui des activités économiques se sont ré-implantées (illustration 42). Nous observons également que la Plaine est traversée par une importante infrastructure ferroviaire amenant à la gare du Nord qui est la première gare européenne en termes de passagers.

Pour étudier l'attraction du pôle d'emploi de la Plaine de Saint-Denis nous constituons deux catégories : les cadres et professions intellectuelles supérieures (illustration 43) d'une part et d'autre part les ouvriers et les professions intermédiaires (illustration 44). Les aires d'attraction de ces deux catégories ne sont pas identiques selon les modes de transport utilisés. Les cadres utilisant une voiture personnelle habitent en majorité à l'ouest de la

Plaine-Saint-Denis. Les ouvriers et professions intermédiaires sont localisés dans les communes de l'est plus pauvres économiquement. Nous observons également une différence de répartition des lieux de résidence pour ceux utilisant les transports en commun. Les cadres utilisant les transports en commun proviennent bien plus de l'ensemble de la commune de Paris, exception faite des arrondissements plus riches de l'ouest parisien. Les catégories sociales plus défavorisées utilisant les transports publics sont plus présentes dans les arrondissements plus pauvres (17^e, 18^e, 19^e arrondissements et dans une moindre mesure le 20^e arrondissement). Elles sont également plus présentes dans les communes défavorisées de l'est de la Plaine-Saint-Denis.

L'exemple de la Plaine-Saint-Denis met en lumière un phénomène nous intéressant directement dans notre projet d'améliorer l'accessibilité en mode de transport durable. À l'aide du support carroyé nous remarquons un phénomène de fragmentation urbaine en raison des infrastructures de transport en commun ferroviaires. Les grandes emprises au sol des infrastructures ferroviaires et plus généralement de transport engendrent de la fragmentation urbaine.

Lieux de résidence des actifs travaillant dans la zone de la plaine de Saint-Denis: les cadres et professions intellectuelles supérieures

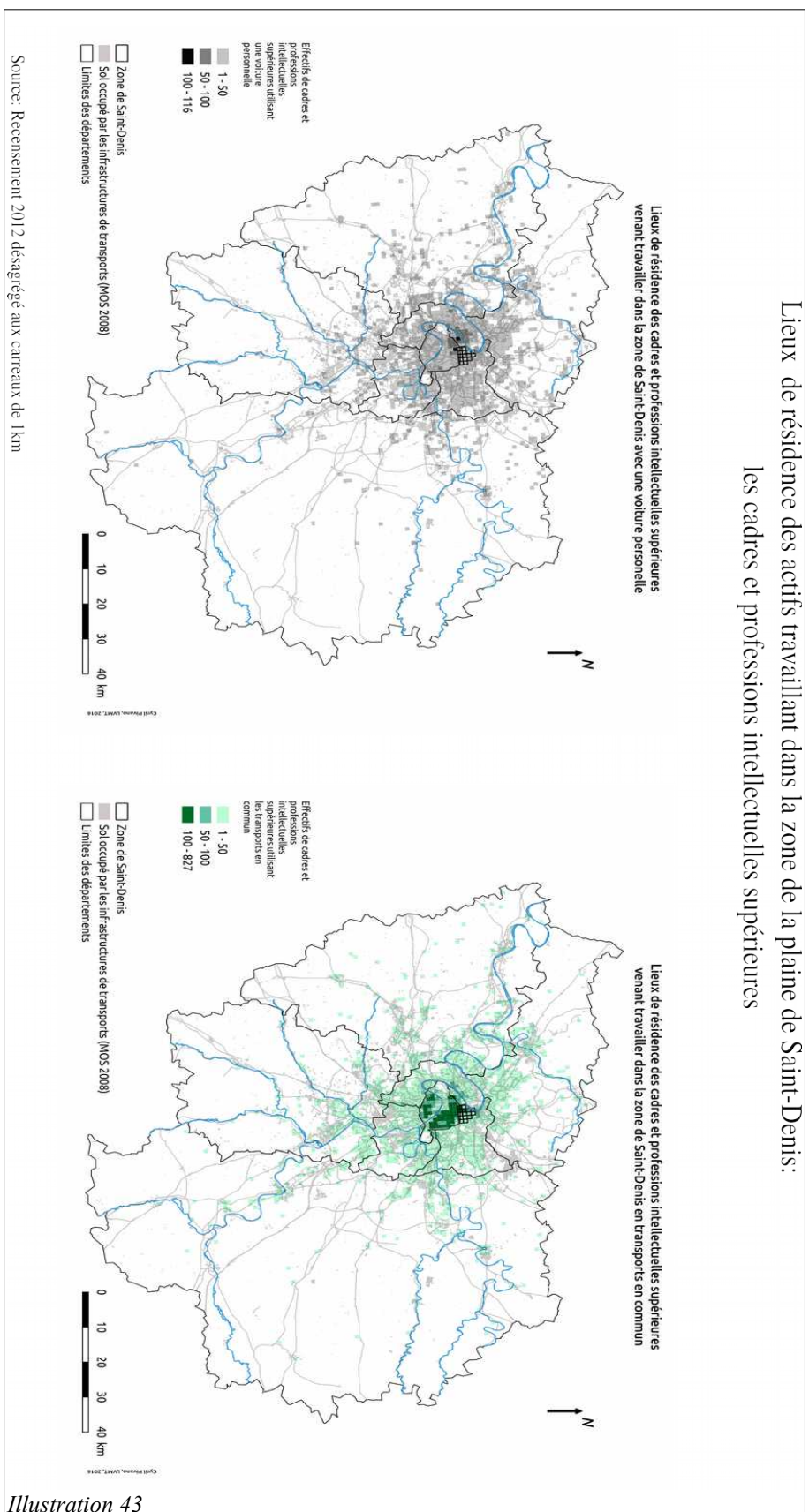
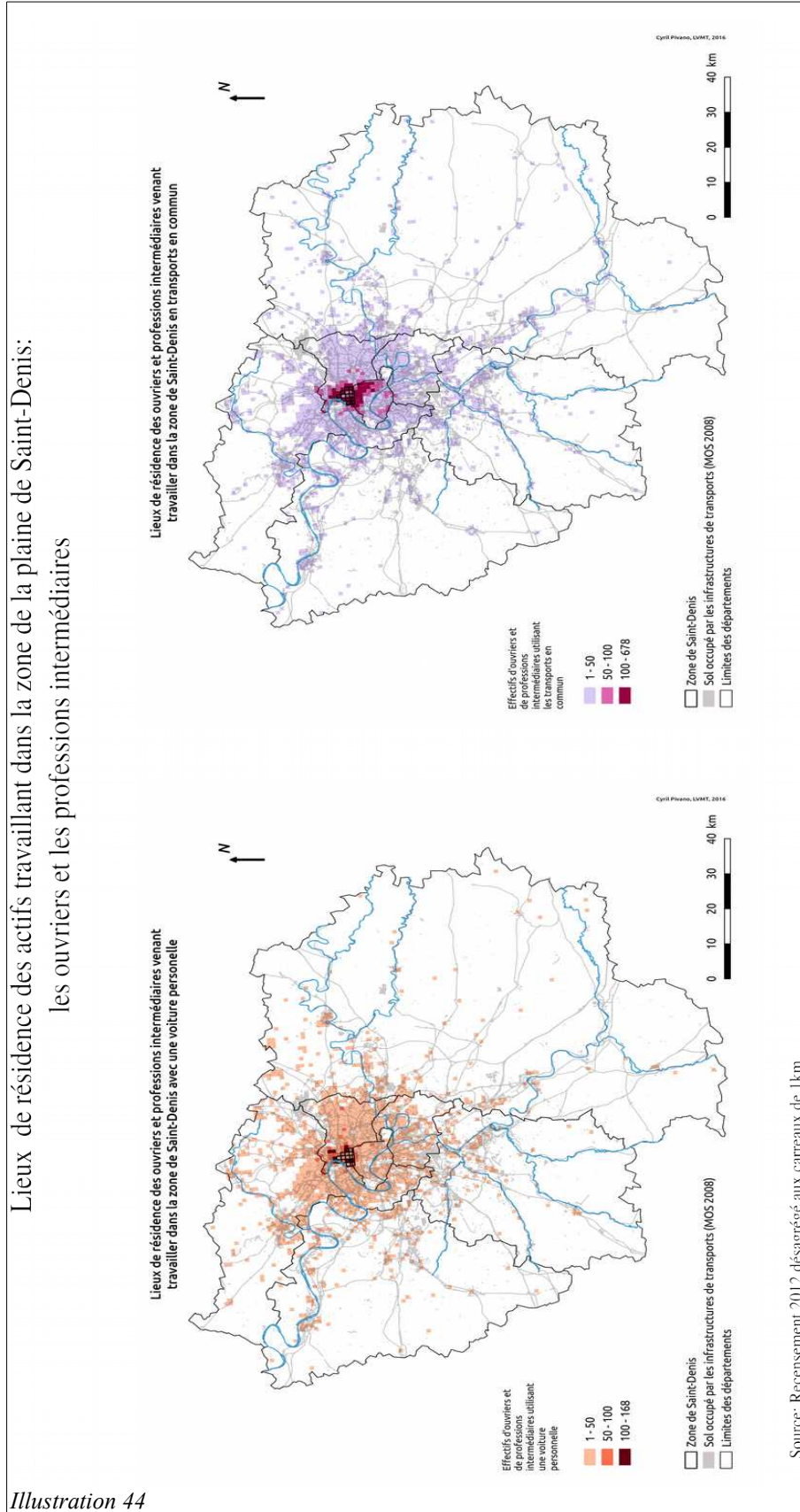


Illustration 43

Lieux de résidence des actifs travaillant dans la zone de la plaine de Saint-Denis:
les ouvriers et les professions intermédiaires



5.4.3 Vincennes, Saint-Mandé et le quartier du Bas-Montreuil

Nous terminons notre série de cas d'étude par l'examen de la zone est de Paris constituée des communes de Vincennes, de Saint-Mandé et du quartier du bas-Montreuil (illustration 45).

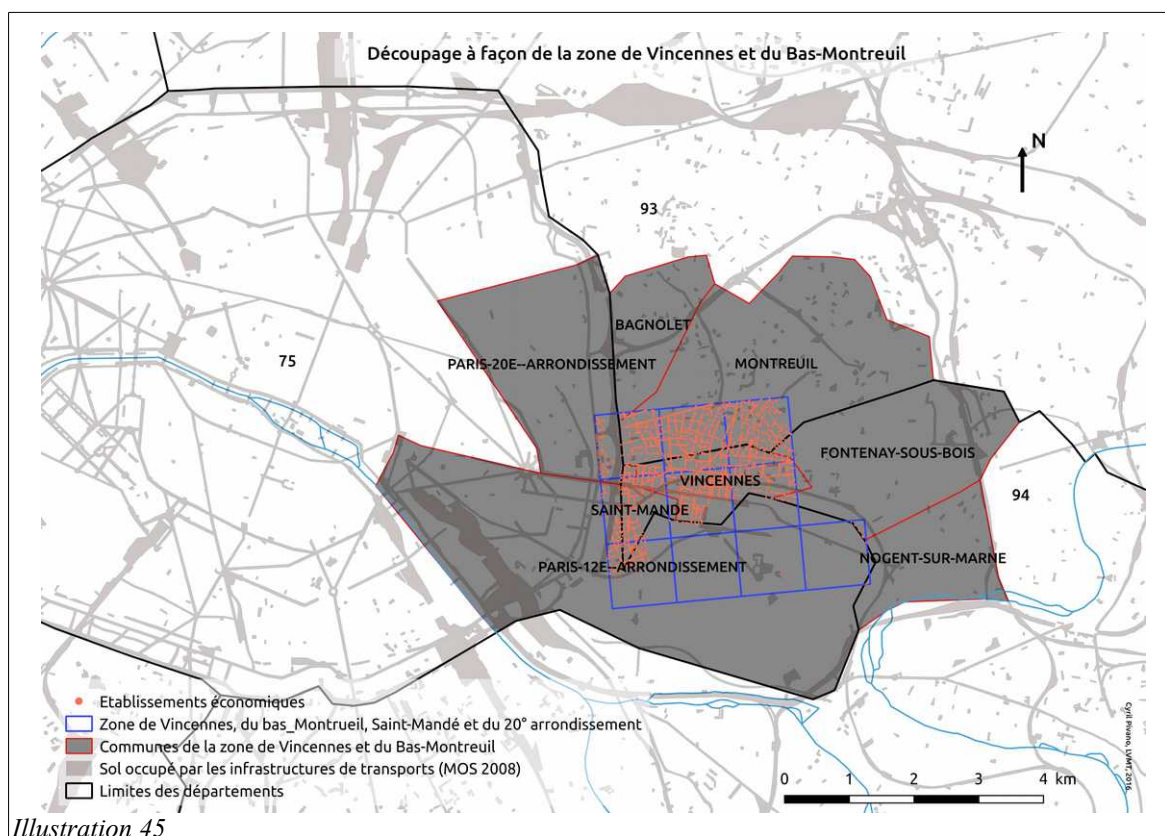


Illustration 45

Le Bas-Montreuil est un quartier de la commune de Montreuil. Ce quartier a connu un phénomène de gentrification. « *De la rue, on ne se doute de rien. Des bicoques en crépi entre deux hangars, adossées à des barres HLM, des meublés insalubres, des bazars à trois sous, des bars-PMU : le quartier du Bas-Montreuil, limitrophe de Vincennes, est une banlieue populaire, sans chic. La rue Bara abrite le premier foyer malien de France. Rue de Paris, les boucheries halal sont plus nombreuses que les bonnes fromageries. Pourtant, il suffit de pousser les grilles, d'entrer dans les cours, de visiter les anciennes usines, les vieux ateliers, pour découvrir d'autres hirondelles des faubourgs. Les ouvriers ou les artisans qui travaillaient jadis pour le faubourg Saint-Antoine ont laissé place à des artistes, cachés à*

l'abri des façades en meulière, logés dans des lofts retapés. [...] Intermittents fauchés, travailleurs indépendants avec enfants, journalistes, universitaires, chassés de Paris par la cherté des loyers, en quête d'un lieu pour répéter, dessiner, sculpter et... respirer, ils ont été des précurseurs. Ces défricheurs de marges, ces décalés – avec du temps, du goût, mais peu d'argent – ont retroussé leurs manches pour retaper eux-mêmes une maison ou un pan d'usine désaffectée. En même temps, ils ont insufflé une nouvelle manière de vivre dans leur quartier. Se cooptant entre eux pour l'achat d'une maison. Et contribuant à rendre la ville plus vivante. Mais plus chère aussi. (« Les bobos investissent la banlieue rouge de Paris », citation du journal Le Monde, 30 mai 2004, p. 11 repris de la thèse de Anaïs Collet (Collet, 2010, p.9). Le phénomène de gentrification est un processus par lequel d'anciens quartiers populaires se trouvent modifiés par un afflux de ménages de classes moyennes et une réhabilitation du bâti. Ce processus est identifié depuis longtemps et continue à affecter les centres de nombreuses métropoles. Toutefois, la gentrification peut concerner des populations et des espaces de plus en plus variés (Authier, 2008).

La question de la gentrification dans notre contexte de recherche amène à vouloir disposer de données pour observer dans le temps un même lieu et observer les évolutions des profils des résidents et des actifs y travaillant à l'aide de données. Dans sa thèse Anaïs Collet a par exemple été confrontée au changement des limites administratives pour observer plus spécifiquement le quartier dit du Bas-Montreuil (annexes 2 de sa thèse). Plus largement les travaux de la sociologie urbaine sont confrontés à l'antagonisme énoncé au chapitre 2. La mise à disposition de données neutres peut-elle aider ? Le commentaire d'ouvrage de Grafmeyer permet d'illustrer ce besoin sur le cas de l'agglomération Lyonnaise (Grafmeyer, 1991) : « *L'auteur sociologique, spécialisé des milieux urbains, s'intéresse aux relations entre résidents et lieux. Il se livre ici à une étude comparative non pas entre des quartiers, - et il s'interroge d'ailleurs sur cette notion -, mais sur trois terrains d'observation réduits, pris au sein d'îlots I.N.S.E.E. et n'ayant à priori ni identité, ni centre, ni nom propre, ni vie de quartier.[...] L'analyse s'opère à partir des données brutes de recensement I.N.S.E.E. accessibles seulement pour 1975* ». La fin du commentaire d'ouvrage précise que le géographe aurait souhaité une localisation et des illustrations plus précises.

Pour notre travail d'illustration du potentiel des données du recensement désagrégées nous reprenons la distinction entre les cadres et professions intellectuelles (illustration 46) et les ouvriers et professions intermédiaires (illustration 47).

Lieux de résidence des actifs travaillant dans la zone de Vincennes, Saint-Mandé et du Bas-Montreuil : les cadres et professions intellectuelles supérieures

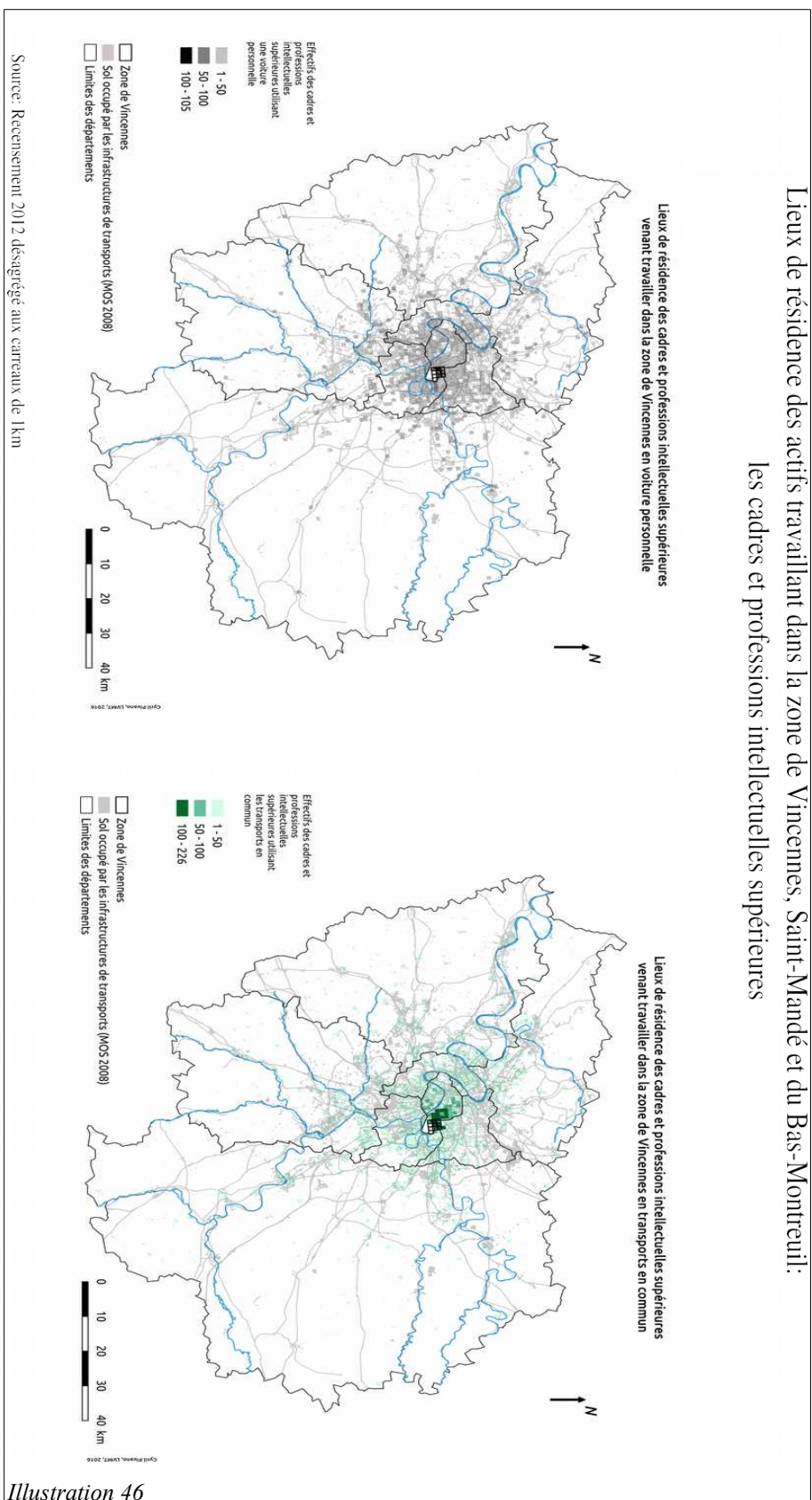
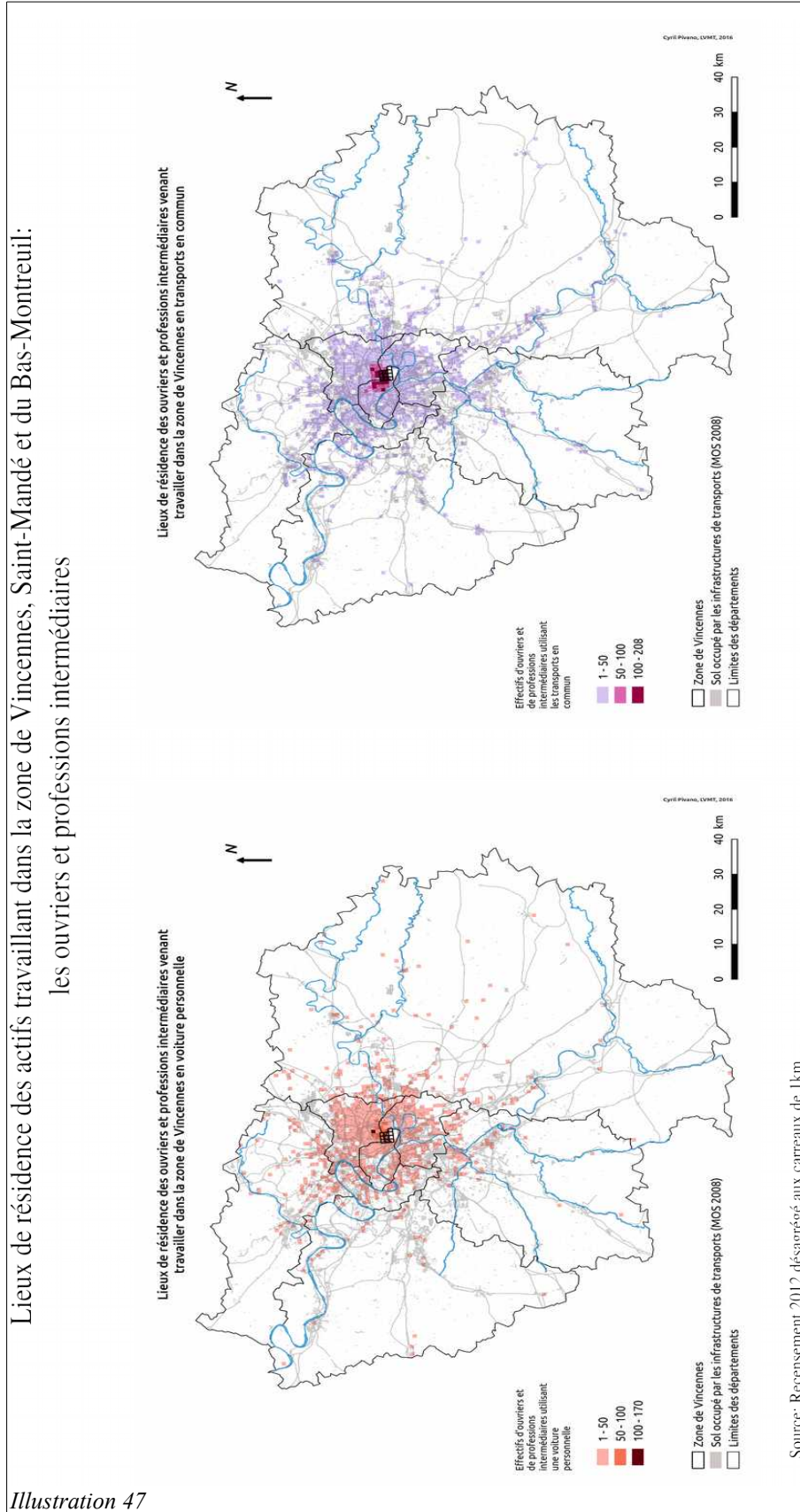


Illustration 46

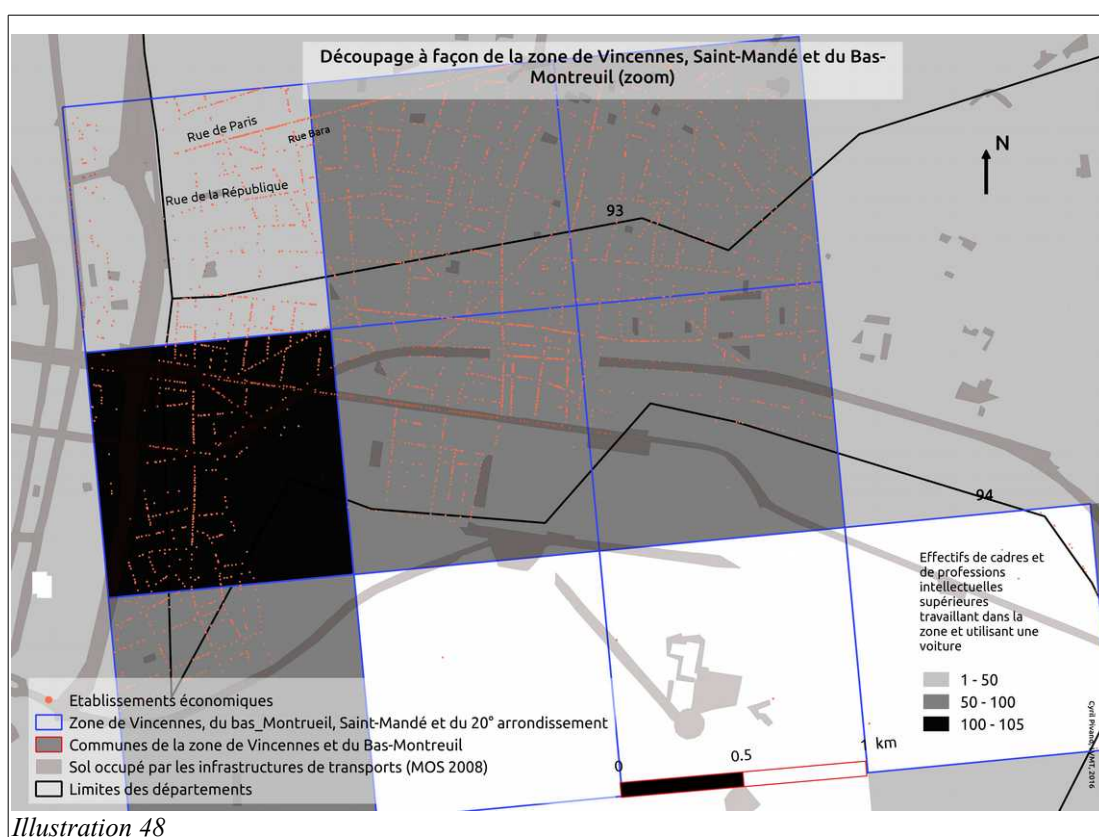
Lieux de résidence des actifs travaillant dans la zone de Vincennes, Saint-Mandé et du Bas-Montreuil:
les ouvriers et professions intermédiaires



Source: Recensement 2012 désagrégé aux carreaux de 1km

Nous observons que les classes les plus défavorisées travaillant dans la zone d'étude habitent au nord de notre zone d'étude. Les ouvriers et professions intermédiaires utilisant une voiture pour aller travailler habitent plus à l'est dans les quartiers résidentiels de Montreuil et de Fontenay-sous-Bois. Alors que les utilisateurs des transports en commun chapeautent la zone d'étude du nord-ouest au nord-est, le long de la ligne 9 du métro.

Les cadres travaillant dans la zone d'étude et utilisant les transports en commun habitent relativement plus dans la commune de Paris (20^earrondissement). Et, nous observons que certains cadres utilisent une voiture pour aller travailler dans cette zone alors qu'ils habitent à proximité. Ils effectuent donc de très petites distances (illustration 48).



Notons que les bornes de discrétisation de la figure ont été construites pour être comparées aux autres illustrations. La troisième classe devrait être regroupée avec la seconde classe. Cette carte montre que le quartier du bas-Montreuil qui est ici délimité par le carreau en haut à gauche génère moins de trafic automobile que les autres carreaux. D'une part, nous observons que la délimitation incorpore une lourde infrastructure routière (le boulevard

périphérique) et donc la population résidente est certainement plus faible que dans les autres carreaux. D'autre part le taux de chômage est peut être plus élevé qu'ailleurs. Pour en savoir plus, le géographe souhaiterait sûrement aller sur le terrain et/ou plus de données sur les carreaux. Nous laissons le soin au lecteur intéressé de continuer cette analyse. Nous allons continuer la présentation des exploitations des données du recensement désagrégées au carreau de 1km en traitant des modes de transport à une échelle plus globale.

5.5 Les modes de transport

La démarche d'évaluation des projets de transport est encadrée par la loi du 16 juin 2014. Les parts modales doivent être mesurées avant et après la mise en place du projet de transport. Nous avons souligné la difficulté de la mesure de cet indicateur. Il est relatif à l'ensemble spatial choisi pour observer les proportions d'utilisation des modes de transport. Pourtant nous ne pouvons pas connaître à l'avance les territoires concernés par l'aménagement du projet de transport. Jusqu'à quelle distance pouvons-nous dire que les territoires seront concernés par un projet de transport ? Les fiches méthodologiques du ministère du transport demandent de créer deux périmètres, l'un restreint et l'autre élargi sans toutefois donner des indications sur la superficie de ces deux zonages et la manière de découper les territoires.

Au chapitre 2 nous avons supposé que l'élaboration de données de mobilité sur un support carroyé de 1km permettait de produire des supports à façon répondant entre autre à la problématique de délimitation des périmètres d'étude des projets de transport. En effet, l'utilisation d'un support surfacique rend par essence impossible la mesure de distances infrazonales. Dans le cas de périmètres élargis regroupant les quartiers politiques des communes, nous faisons face à une trop grande incertitude sur les distances d'accès réalisables pour utiliser le projet (s'ils n'ont pas été mesurés sur un support de recueil suffisamment fin). Il faut déterminer un périmètre restreint sur lequel la part modale d'utilisation des transports en commun est bien plus fiable. Ensuite, il nous faut comparer les écarts entre les parts modales des périmètres élargis et restreints afin d'estimer un potentiel de report modal.

L'illustration 49 représente les parts modales selon des périmètres élargis (les communes) et restreints (les carreaux de 1km). Les parts modales construites cumulent les taux d'utilisation des transports en commun à l'origine et à la destination des déplacements domicile-travail.

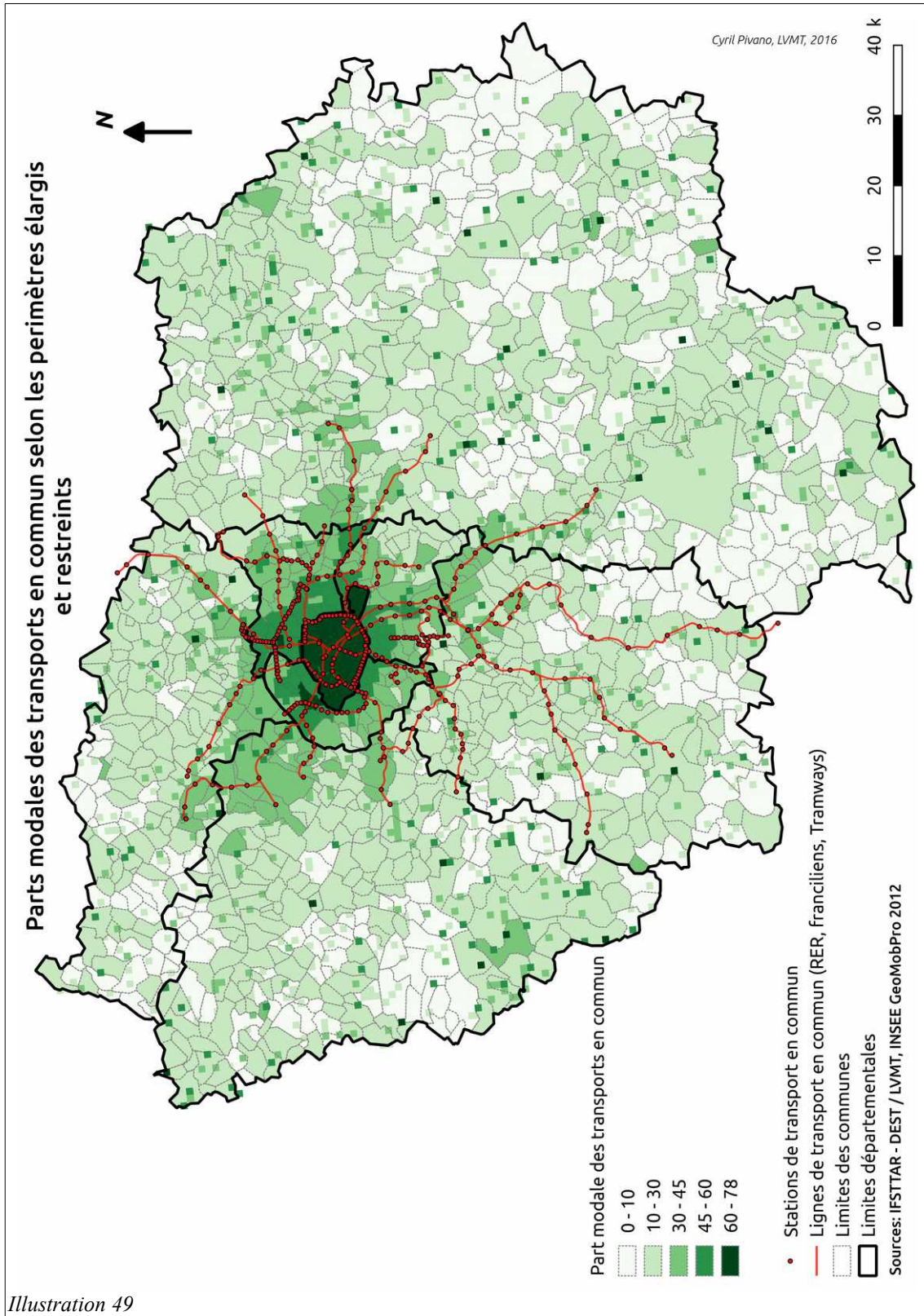


Illustration 49

En superposant les résultats des parts modales des périmètres restreints (les carreaux) sur les périmètres élargis (les communes) nous pouvons observer plus finement les concentrations d'utilisation des transports en commun au sein des communes. A l'intérieur des communes, nous différencions les carreaux ayant des parts modales de transport en commun moins importantes que dans le reste de la commune, et nous distinguons ceux qui ont une part modale plus importante. Les carreaux ayant une part modale identique à la classe communale (de même couleur) n'apparaissent pas sur la carte.

Nous observons que la propension à utiliser les transports en commun est plus forte à Paris, puis en petite couronne. En grande couronne, les taux d'utilisation sont plus forts le long des axes de transport en commun. Nous distinguons toutefois le long de ces lignes de transport des carreaux avec de fortes parts modales et d'autres non. Pour aller plus loin dans l'analyse, nous pourrions regarder la distance aux stations de transport en commun de ces carreaux pourtant proches des lignes ferroviaires. Cependant, cela nous emmène en dehors de notre champ d'analyse qui est de présenter et non d'approfondir les exploitations pouvant être faites de la désagrégation spatiale des données du recensement.

Nous continuons en observant la répartition spatiale des parts modales d'utilisation des véhicules personnels pour aller travailler (illustration 50).

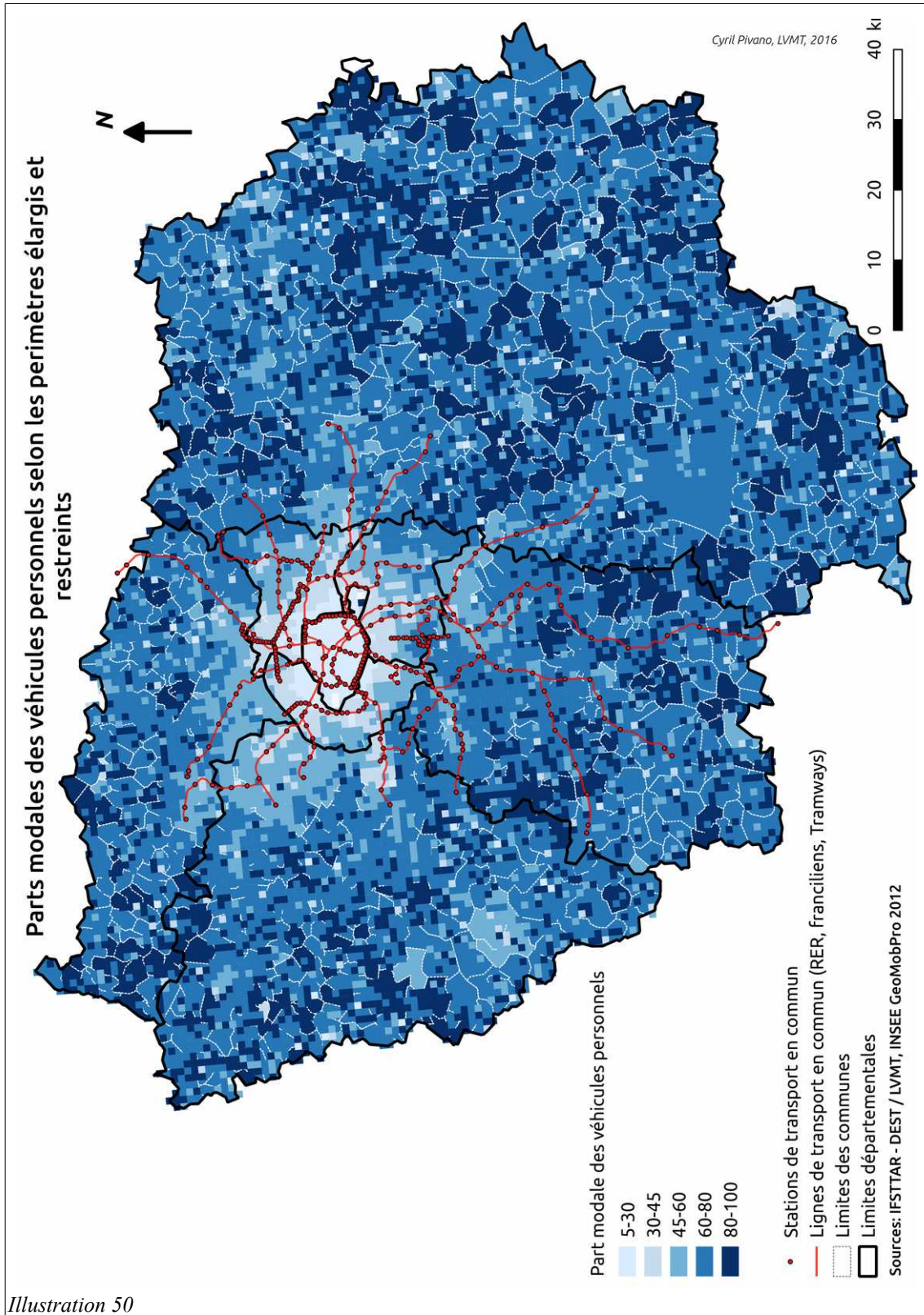


Illustration 50

A l'inverse de la répartition des parts modales de transport en commun, l'utilisation de la voiture personnelle augmente à mesure que l'on s'éloigne du cœur de l'agglomération parisienne. Néanmoins, au sein des communes périphériques nous observons tout de même des concentrations de faibles utilisations de la voiture. Ces zones sont celles où l'utilisation des transports en commun est plus importante.

A l'aide de cette information sur les parts modales, nous savons où agir pour mettre en place des systèmes de transport en commun. Nous avons donc des éléments supplémentaires pour traiter de la question des reports modaux. Dans le contexte de l'évaluation de projets de transport, un professionnel peut déterminer aux environs du projet (un carreau de 1km ou un regroupement de carreaux) les destinations des actifs s'y trouvant et estimer a priori les reports modaux induits par le nouvel aménagement. Pour ce faire, nous pouvons quantifier les automobilistes parcourant des liaisons que le nouvel aménagement de transport en commun permettra de desservir. Bien sûr l'analyse des reports modaux doit se faire en considérant les profils socio-économiques et les revenus.

Les données du recensement désagrégées n'offrent pas d'information sur les revenus. Les chercheurs peuvent néanmoins coupler les données sur les revenus diffusées par les données carroyées de l'Insee (encadré n°1) avec les données désagrégées. Les travaux de recherches comme ceux du club des ingénieurs, de J-P Nicolas (Nicolas, 2001, 2012), ou de Gallez (2000) pourraient alors déterminer des taux de dépense relatifs à la mobilité professionnelle selon les profils socio-économiques et ceux avec une plus grande échelle. Ce travail de couplage de données nécessite toutefois des hypothèses de construction supplémentaires et donc d'autres recherches méthodologiques.

Toujours à l'aide des données désagrégées, nous pouvons déterminer des regroupements d'actifs occupés ayant des origines-destinations plus ou moins communes (même couple de carreaux de 1km) afin de leur proposer de mutualiser leur mode de transport. Pour ce faire, nous prenons tous les automobilistes franciliens et construisons une matrice de flux. Pour mettre en valeur les flux pouvant être mutualisés nous sélectionnons les flux dominants (W. Tobler, 1987). La matrice de flux des actifs déclarant aller travailler en voiture sur un support suffisamment fin donne ainsi une information spatiale sur le potentiel de mutualisation des moyens de transport (transports en commun ou co-voiturage⁹⁴)

94 Les données du recensement ne précisent pas si les actifs utilisant une voiture sont des conducteurs ou des passagers. De plus, la proximité spatiale n'est pas le seul critère pour favoriser le co-

(illustration 51)

voiturage. Les contraintes horaires et les freins psychologiques sont également à prendre en compte.

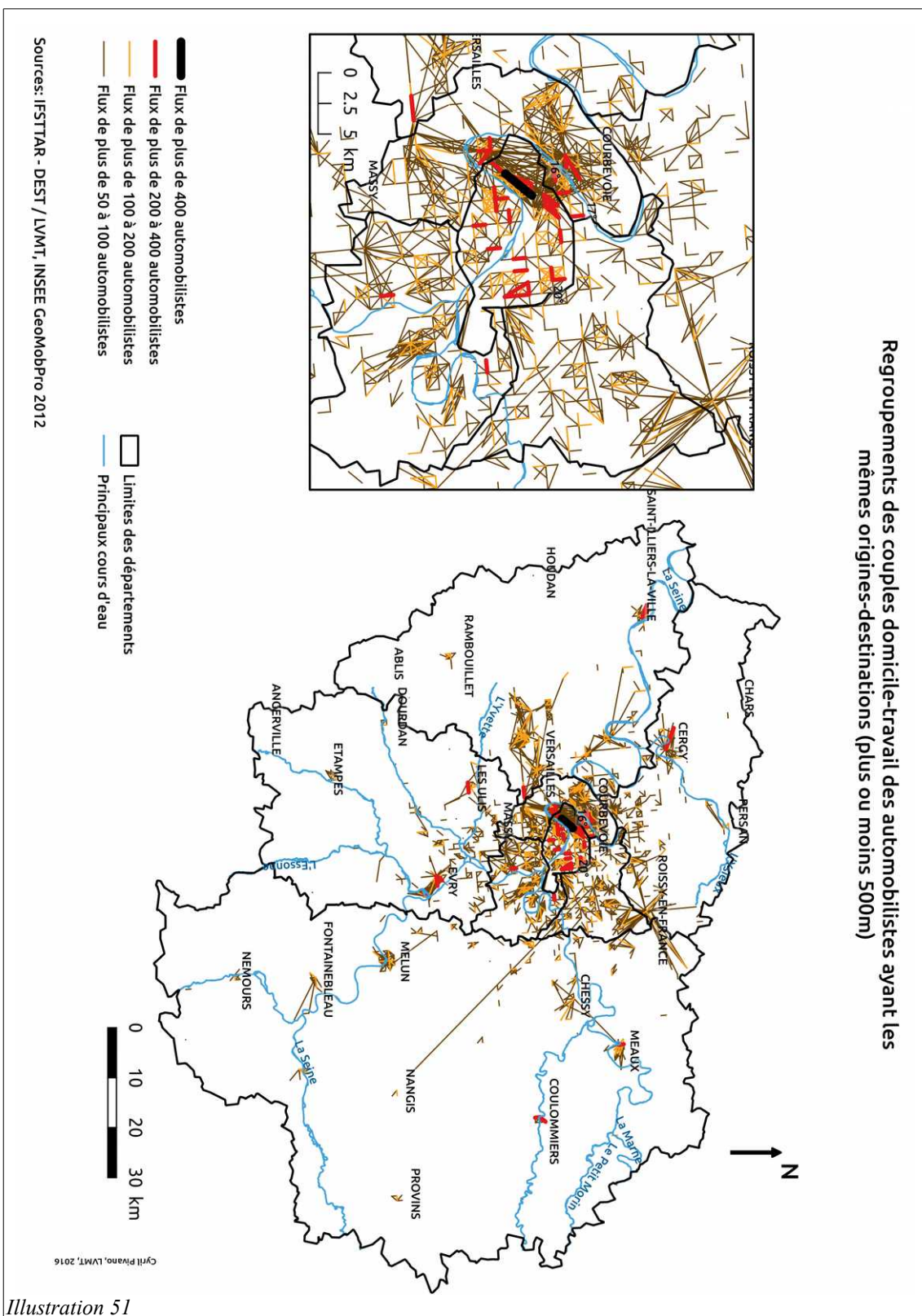
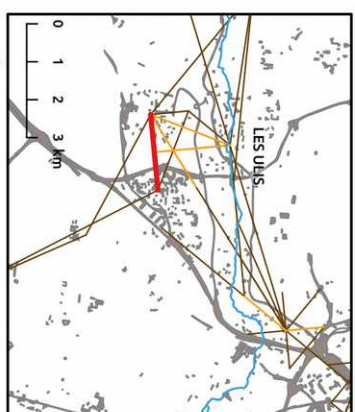
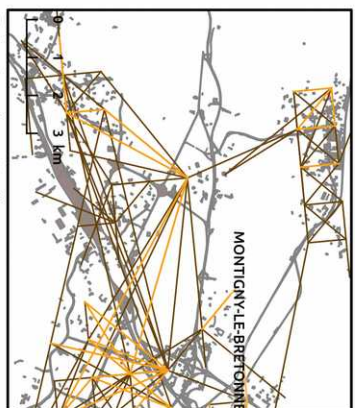
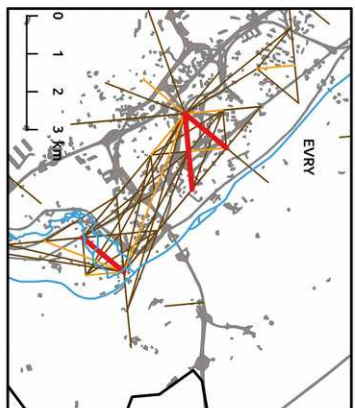
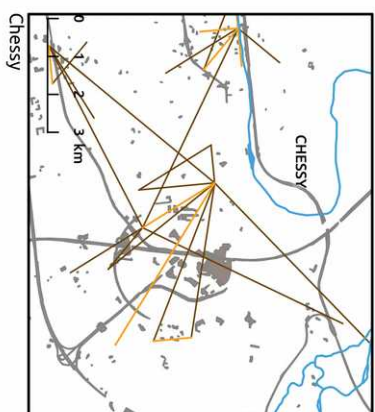
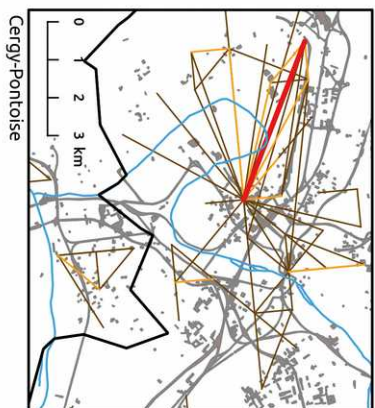
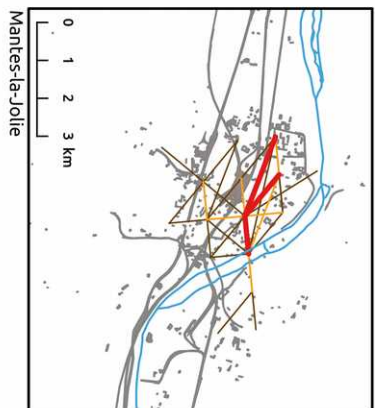


Illustration 51

Le potentiel de mutualisation des modes de transport est plus fort dans le 16^e arrondissement puis dans le reste de la commune parisienne. Les communes plus éloignées du cœur de l'agglomération parisienne ont aussi un potentiel de mutualisation des modes de transport. L'illustration 52 effectue quelques zooms à l'intérieur des villes-nouvelles.

Regroupements des couples domicile-travail des automobilistes ayant les mêmes origines-destinations (plus ou moins 500m)

Quelques zooms géographiques:



- ▬ Flux de plus de 400 automobilistes
- ▬ Flux de plus de 200 à 400 automobilistes
- ▬ Flux de plus de 100 à 200 automobilistes
- ▬ Flux de plus de 50 à 100 automobilistes

- Limites des départements
- ▬ Principaux cours d'eau
- Sol occupé par les infrastructures de transports (MOS)

Sources: IFSTTAR - DEST / LVMT, INSEE GeoMobpro 2012

Cyril Pivano, LVMT, 2016

Illustration 52

Cette méthode des flux dominants appliquée aux déplacements effectués en transport en commun permet également de repérer les liaisons les plus fréquentées et donc les plus rentables pour les gestionnaires des transports en commun (illustration 53).

Nous retrouvons sur l'illustration suivante la zone de l'est parisien qui génère de petits déplacements (cartes des déciles). C'est la zone où les liaisons similaires sont les plus importantes. A l'inverse, les liaisons de périphérie à périphérie sont encore peu nombreuses. Le projet du Grand Paris Express devrait les développer. Une étude ex-post permettra d'observer les gains obtenus par ces nouveaux aménagements.

Regroupements des couples domicile-travail des voyageurs en transports en commun ayant les mêmes origines-destinations (plus ou moins 500m)

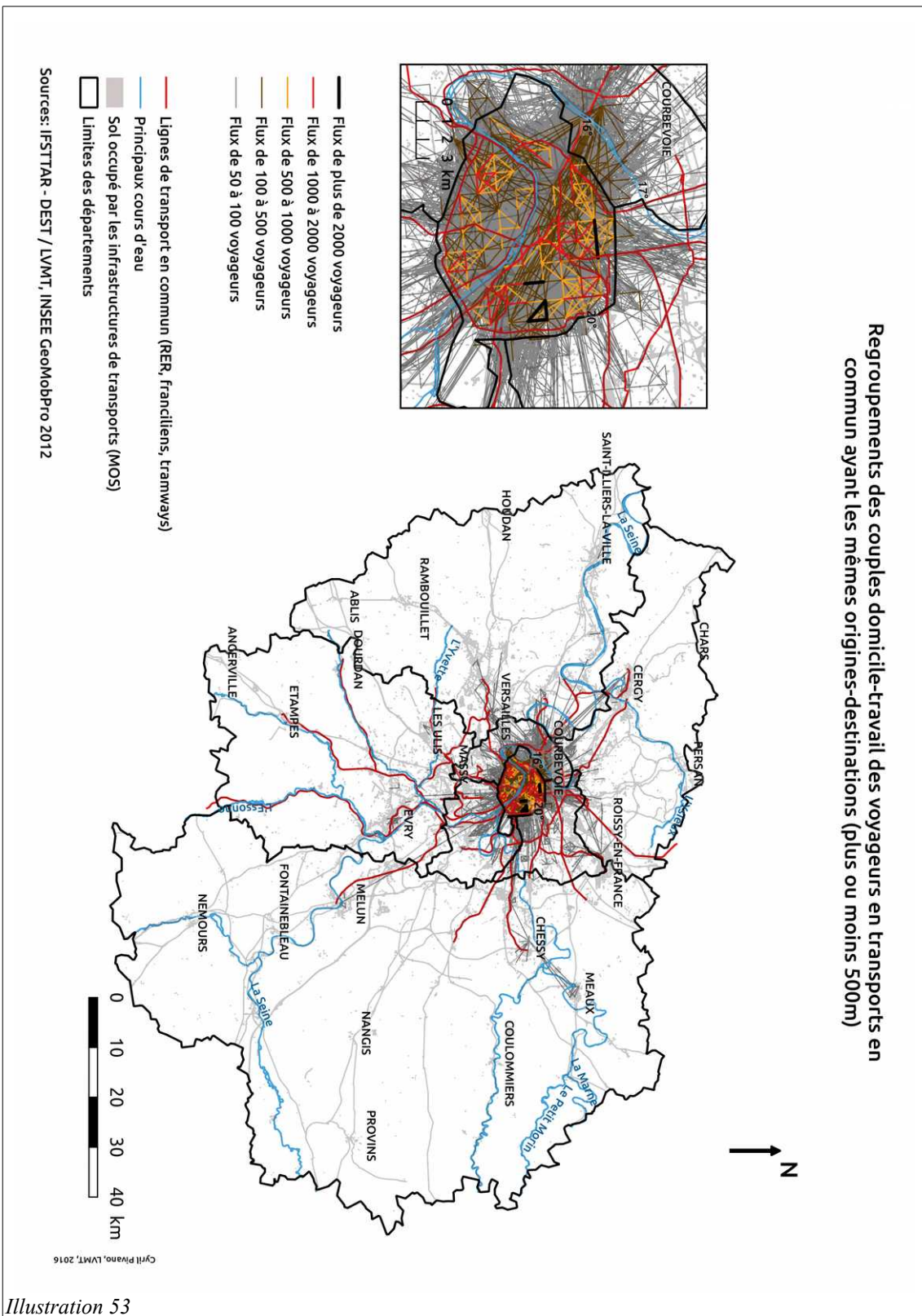


Illustration 53

Conclusion du chapitre 5

Le chapitre 5 a illustré des applications pratiques des données désagrégées sur le support carroyé de 1km. Nous pouvons observer plus finement la mobilité professionnelle sur des petits territoires et/ou à cheval sur des limites administratives.

Le support carroyé de 1km étant plus fin que celui de la commune les mesures des petites distances sont améliorées. Nous disposons d'une connaissance sur les mobilités infra-communales ; des indicateurs sur les petites et grandes distances pour chaque carreau ont été mesurés.

Les données de mobilité sur un support fin permettent également de mettre en lumière des zones attirant ou émettant des actifs occupés à l'aide d'un ratio simple (le nombre d'emplois moins le nombre de résidence). De cette manière les pôles d'emploi de la région internes aux communes sont mis en lumière.

Nous avons effectué ensuite quelques zooms géographiques pour étudier des petits territoires permettant ainsi d'observer les équilibres habitats-emplois avec une plus grande échelle que le support communal. Par exemple, nous avons pu observer localement les différentes distributions spatiales des distances domicile-travail selon les catégories socioprofessionnelles des individus pour plusieurs zones économiques (aéroports d'Orly, Plaine de Saint-Denis, Vincennes, Saint-Mandé et Bas-Montreuil).

Les données désagrégées du recensement peuvent aider à définir des périmètres restreints demandés par la législation des évaluations des projets de transport. Les zones de concentration des utilisateurs des transports en commun et les poches de dépendance à la voiture personnelle à l'intérieur des communes ont été cartographiées.

En agrégeant, les origines-destinations synthétiques domicile-travail en flux nous pouvons observer des poches de mutualisation des moyens de transport.

Ces cas d'application ne font qu'exposer des applications des données désagrégées du recensement et mériteraient d'être approfondis.

D'autres thèmes peuvent être traités, à l'aide du répertoire SIRENE géocodé à l'adresse nous pouvons mesurer des distances aux différents services (de base, de proximité, intermédiaire et supérieure). En conclusion générale, nous discutons plus généralement de notre proposition méthodologique de désagrégation spatiale des données du recensement.

Conclusion générale

Conclusion générale

La directive INSPIRE préconise de diffuser les données permettant le suivi et l'évaluation des politiques environnementales sur un support carroyé. Nous avons choisi de descendre les données de mobilité du recensement sur un support carroyé de 1km pour observer la demande de mobilité domicile-travail avec une résolution fine. De cette manière nous proposons des données répondant aux besoins des analyses allant dans le sens des politiques de facteur 4. Nous avons présenté des travaux de perspectives qui nécessitent des données finement localisées. En effet, la baisse des émissions de CO₂ demande des changements de comportements de mobilité de la part des individus. Pour déterminer les critères d'action sur lesquels jouer nous avons besoin de comprendre les relations transport-urbanisme à une échelle locale. Les leviers identifiés sont l'usage quotidien des modes doux et des transports en commun, la réduction des distances quotidiennes parcourues et la restauration des liens sociaux de proximité.

Pour mieux diriger les politiques allant dans ce sens nous avons besoin de données. Dans la perspective d'une augmentation des coûts liée à l'utilisation de véhicule personnel, elles permettent, par exemple, d'identifier les ménages les plus vulnérables. Les données finement localisées permettent également de mieux cibler l'action publique. Notamment, des périmètres restreints pour évaluer des politiques de transport doivent être définis. De manière générale, les données finement localisées permettent d'analyser les contraintes locales des individus en lien avec leur positionnement dans l'espace. Avec ces données les politiques peuvent agir en proposant des aménagements de transport et urbanistiques favorables à un changement de comportement de la part des individus.

Pourtant, les éléments de la littérature pointe du doigt le manque de données de qualité pour répondre à ces impératifs de localisations et de délimitations des territoires trop dépendant à l'utilisation de la voiture personnelle et susceptible d'être une niche économique pour les modes alternatifs à la voiture. Au chapitre 1 nous avons présenté l'appel à projet du PREDIT-4 avec notamment les deux questions suivantes :

- L'accès à de nouvelles connaissances sur nos pratiques de mobilités est-il actuellement techniquement possible, à quel coût ?
- Quelles sont les modalités pour mesurer l'évolution des pratiques des citoyens en

fonction du contexte, de l'offre et des contraintes ?

Nous avons répondu à la question sur l'accès aux connaissances des pratiques de mobilités en effectuant un panorama des données de mobilité disponibles. Les EMD, le recensement, les données de téléphonie mobile, les données des questionnaires des infrastructures de transport sont différentes sources de données ayant des avantages et des inconvénients. Pour répondre à la deuxième question, cette thèse propose une modalité de production pour mesurer les évolutions des pratiques non pas en produisant de nouvelles données mais en travaillant sur la notion des supports géographiques des données de mobilité.

Les supports géographiques des données de mobilité conditionnent les analyses pouvant être effectuées. Ils sont toutefois dépendants des choix des producteurs de données. Pour adapter les supports géographiques des données deux solutions sont envisageables. D'une part il est possible d'agrèger les données sur des supports alignés aux supports de base de diffusion. D'autre part, les études de modélisation de trafic automobile démontrent qu'il est possible d'effectuer des traitements sur les données pour les transférer sur des supports non-alignés au support de diffusion. Les politiques de diffusion des données vont dans ce sens en demandant de diffuser des données sur un support carroyé pour faciliter les combinaisons de séries de données géographiques (interopérabilité des données).

Nous avons choisi cette dernière voie en proposant de changer le support des données de mobilité du recensement en effectuant une désagrégation spatiale. Nous avons choisi de traiter du recensement en raison de son taux d'échantillonnage élevé, de la richesse des profils socio-économiques permettant de faire des analyses croisées entre les caractéristiques des actifs occupés, les localisations et les processus de choix expliquant les modes de transport pour aller travailler. Les données de mobilité du recensement sur le support carroyé INSPIRE de 1km permettent ainsi d'observer avec une résolution plus fine que le support communal les contraintes de localisations expliquant les comportements de mobilité.

Le travail de descente d'échelle nommé désagrégation spatiale est constituée de deux étapes méthodologiques. Nous avons vu qu'il nous faut disposer d'un grand nombre d'observations pour pouvoir les répartir en couvrant au maximum l'espace géographique. Ensuite, des données de contrôle doivent permettre d'effectuer des ventilations sous contraintes des données de mobilité. La difficulté étant que les données traitées sont des données bi-localisées. Nous revenons sur ces deux points.

Les données d'enquêtes ne fournissent que des échantillons d'observations. Les échantillons sont trop petits pour répartir les observations dans l'espace géographique de manière satisfaisante. Pour augmenter artificiellement le nombre d'observations des d'enquêtes du recensement, nous avons proposé le concept de clonage. Pour cela, nous utilisons les poids de sondage construit initialement pour harmoniser les différents types d'exploitations et les différences temporelles des enquêtes annuelles du recensement. Ici, nous supposons qu'une observation représente autant d'individus que son poids de sondage l'indique. Une observation est alors constituée de plusieurs clones. Et pour couvrir l'ensemble de l'espace des données de contrôle nous séparons dans certains cas les clones. Dans ces cas, les clones ont des localisations différentes. 87 126 clones ont été séparés lors du processus appliqué à l'Île-de-France.

Ensuite, pour localiser les observations d'enquêtes nous effectuons des ventilations spatiales sous contraintes. Nous avons effectué une ventilation sous contraintes pour localiser les lieux de résidence des observations de ménages du recensement, puis une seconde ventilation pour localiser les lieux d'emploi des actifs occupés de plus de 15 ans. Les données de contrôle utilisées aux étapes de ventilations ne renseignent pas directement sur les couples domicile-travail à produire. Toutefois, nous avons vu que la mobilité des individus dépend de facteurs individuels mais aussi collectifs donnant lieu à une forme urbaine structurée (concurrence foncière, ségrégation socio-spatiale, concentration des entreprises dans l'espace). Nous supposons que les données de contrôle utilisées renseignent sur cette organisation de l'espace et que les couples domicile-travail synthétiques contiennent cette information.

Avant de préciser les jeux de données utilisés dans le processus appliqué à l'Île-de-France et de commenter leur possible amélioration nous revenons sur l'appariement des fichiers détail. Nous avons procédé à l'appariement des fichiers détail anonymisés du recensement pour obtenir plus d'information décrivant les observations du fichier MOBPRO. Nous avons obtenu une plus grande précision des localisations résidentielles des résidents des grandes communes. D'une connaissance à la commune de résidence nous sommes passés à l'IRIS de résidence. De plus, nous avons obtenu une meilleure description des secteurs d'activité des actifs. D'une description en 5 classes nous disposons d'une typologie en 17 classes. Ainsi, à l'aide de l'appariement, les observations du recensement sont localisées résidentiellement au support IRIS-Communes et les lieux de travail des actifs sont localisés sur le support communal.

Conclusion générale

C'est en partant de ces deux supports que nous effectuons les deux ventilations spatiales sous contraintes :

Une première ventilation sous contraintes est effectuée pour répartir les observations du support IRIS-Communes aux carreaux de 200m. Nous avons utilisé la variable renseignant sur les effectifs de ménage par carreau pour ventiler les observations de ménage. De plus, nous avons localisé résidentiellement les autres membres des ménages décrits dans le fichier INDCANTVILLE. Rappelons que seul 172 405 ménages des petites communes avec aucun actif occupé n'ont pas été localisés sur le carroyage de 200m sur les 5 millions de ménages d'Île-de-France. Les variables du fichier INDCANTVILLE (89 variables) sont disponibles sur le carroyage de 1km de l'Insee.

Les actifs occupés ont été ensuite ventilés en se servant des données du répertoire SIRENE. Nous avons utilisé comme variable de contrôle les secteurs d'activité en 17 classes obtenus à l'aide de l'appariement, et les classes d'effectifs des établissements économiques. Aussi nous avons utilisé des distributions de distances domicile-travail effectuées en marche à pied, et des distributions de distances entre les stations de transport en commun de travail et les établissements économiques. Ces distributions de distances sont issues de l'Enquête Globale Transport de 2010.

Avec ces deux ventilations sous contraintes nous avons appliqué respectivement la structure spatiale des revenus fiscaux localisés (carroyage de population) et du répertoire SIRENE sur le support carroyé de 200m et celui des adresses postales. Puis, nous avons ré-agrégé les observations sur le carroyage INSPIRE de 1km. Nous avons ainsi affiné les localisations résidentielles de 1 889 340 ménages enquêtés représentant 4 838 050 ménages franciliens. Et, nous avons précisé les localisations des emplois pour 1 914 648 observations d'actifs des ménages ordinaires habitant et travaillant en Île-de-France qui représentent 5 277 072 actifs occupés.

Nous avons ensuite exploré le potentiel des données produites en présentant quelques applications possibles répondant au besoin en données dans le contexte du changement de paradigme des politiques de transport. Le processus de désagrégation appliqué à l'Île-de-France permet d'obtenir une image statique de la demande de déplacement domicile-travail sur un carroyage de 1km. Les données désagrégées du recensement permettent alors d'effectuer des zooms à l'intérieur des communes. Les petites distances peuvent être mieux

observées. La finesse de ce support autorise d'appliquer un indicateur d'attraction-répulsion pour différencier les espaces attractifs et émetteurs de déplacements domicile-travail. Ce support peut être également utilisé pour délimiter les périmètres restreints dans les évaluations de projet de transport. La construction d'une matrice de déplacement à l'échelle de 1km permet, de plus, d'identifier les liaisons domicile-travail les plus fréquentées et éventuellement de mener des politiques d'incitation à la mutualisation des modes de transport.

Les étapes de ventilation sous contraintes de la désagrégation peuvent néanmoins utiliser plus d'information. Par exemple, nous pouvons ventiler les lieux de résidence autrement qu'en utilisant seulement les effectifs de ménages. Mais cela nous amène à traiter des nouvelles méthodes de production des données carroyées sur lesquelles notre processus est dépendant. Les méthodes de construction des données carroyées ont évoluées. Mise à part les données renseignant sur les effectifs d'individu, les autres variables sont, à ce jour, diffusées sur des rectangles de tailles variables. C'est aux utilisateurs de ventiler les variables du support rectangulaire au support carroyé de 200m. En utilisant les nouvelles données carroyées une étape supplémentaire consiste alors à transmettre la variable « *effectifs de ménages* » des rectangles aux carreaux. Et, si nous souhaitons utiliser plusieurs variables de contrôle il convient de transmettre d'autres variables du support rectangulaire aux carreaux i.e d'effectuer une ventilation multicritère. Cette dernière réflexion nous amène à traiter des approfondissements méthodologiques possibles. Les résultats de la ventilation des lieux de résidence démontrent que l'utilisation plus simpliste des effectifs de ménages permet tout de même d'approcher les répartitions spatiales des variables non utilisées dans la ventilation sous contrainte des lieux de résidence (tableau 18). Toutefois nous pouvons envisager d'introduire plus de données de contrôle. Nous pouvons alors utiliser les variables du carroyage de 200m de l'Insee (en supposant que le passage des rectangles aux carreaux ne pose pas de problème). Nous pouvons également mobiliser les données IGN de la couche Bâti pour répartir les observations dans les bâtiments. Rappelons que plus nous affinons les localisations plus nous pouvons ré-agréger à façon pour répondre aux besoins divers des analyses de la mobilité durable.

Parmi les raffinements méthodologiques, nous n'avons pas introduit les ménages avec aucun actifs occupés des petites communes. L'appariement des fichiers détail ne permet pas de retrouver les communes de résidence, pour ceux là, nous connaissons que le canton de résidence. Il nous faudrait alors mettre au point une méthode pour construire les localisations

communales. Enfin, nous sommes confrontés au problème de la bi-résidentialité. L'exemple des enquêtés déclarant ne pas se déplacer mais dont le recensement indique que les communes de résidence et de travail sont distantes de plusieurs kilomètres est problématique. Dans cette thèse nous n'avons pas traité de ce problème. En utilisant d'autres sources de données, nous pourrions éventuellement traiter de ce manque d'information sur la bi-résidentialité et plus largement nous allons pouvoir prendre en compte l'aspect temporel de la mobilité. Le recensement ne donne qu'une image figée de la demande de déplacement, et ne donne qu'une représentation du motif de déplacement domicile-travail.

La production de données du recensement sur les carroyages répondant aux prérogatives INSPIRE peut se faire alors en utilisant des sources de données internes à l'Insee. Premièrement, les données confidentielles du CASD contiennent en un seul fichier les observations de l'exploitation complémentaire et son utilisation permettrait de ne plus procéder à l'appariement des fichiers détail, ainsi nous leverions le cas problématique des ménages sans actifs occupés des petites communes. Deuxièmement, dans une perspective de collaboration avec l'Insee il peut être envisagé d'avoir accès aux données carroyés de 200m avant l'agrégation aux rectangles. Troisièmement, l'utilisation des données des déclarations annuelles des données sociales des entreprises (DADS) peuvent être utilisées comme données de contrôle. Ce dernier point mérite d'être détaillé. Les données DADS sont une source de données renseignant de manière exhaustive les adresses des lieux de résidence et d'emploi des salariés des entreprises. A priori ces données pourraient être diffusées par l'Insee librement en les agrégeant sur un support carroyé et nous disposerions alors d'une image de la mobilité domicile-travail sur un support répondant aux prérogatives INSPIRE. Des méthodes de floutage pourraient être utilisées pour respecter l'anonymat. Toutefois, notre choix de travailler avec les données du recensement de population permet de disposer de données décrivant les actifs occupés mais également de données sur les membres de leur ménage. C'est pourquoi nous supposons que les données DADS pourraient servir bien plus de données de contrôle dans notre processus de désagrégation que de données brutes que l'on agrégerait. Le recensement apporte des informations que les DADS ne renseignent pas et cela permet d'envisager des pistes de réflexion pour traiter les autres motifs de mobilité.

Notre travail de désagrégation des données de mobilité du recensement peut *in fine* servir comme base de couplage avec d'autres sources de données de mobilité. De la même manière

que nous nous appuyons sur des données de contrôle décrivant les caractéristiques socio-économiques des observations, le couplage de données avec d'autres sources de données est plus aisé en disposant de nombreuses informations décrivant les observations. Le recensement de population nous apparaît comme la source de données la plus complète en termes de richesse de description et en termes de taux d'échantillonnage.

Une des solutions de couplage de données de mobilité qui peut être envisagée grâce aux données du recensement désagrégées au carreau de 1km est de coupler les données de traces GSM avec les données fines du recensement. Par exemple, il est possible de déterminer les antennes-relais de résidence et d'emploi des possesseurs de téléphone portable. Avec cette information les traces GSM peuvent être ventilées à l'intérieur des aires des antennes-relais pour ensuite être attribuées à des profils socio-économiques du recensement correspondant à leurs nouvelles localisations. Pour les actifs occupés il faudra veiller à faire correspondre les couples origines-destinations domicile-travail des deux sources de données. Les données GSM peuvent alors améliorer le processus de ventilation des lieux d'emploi proposée ici. De cette manière nous disposerions des profils socio-économiques du recensement et des informations sur les autres formes de mobilité et les heures de déplacements à l'aide des données GSM.

En réponse à l'interrogation posée par le PREDIT-4 sur les nouvelles modalités de production de donnée, nous voyons ainsi tous le potentiel des opérations de couplage de données. Nous proposons une modalité de production de données différentes que l'élaboration de nouvelles enquêtes statistiques. Nous nous rapprochons d'une démarche modélisatrice. Et, au lieu d'effectuer des interpolations pour répondre au manque de données dans l'espace voir même dans le temps en suivant des règles de probabilité, nous proposons de nous appuyer le plus possible sur les sources de données existantes. Les données ne sont plus simplement des observations nous permettant de mettre en lumière des occurrences, de tirer des règles de comportement générales et de les appliquer en tous point de l'espace (modèle normatif). A l'ère de la massification des données, une autre démarche peut être menée. Notre travail de désagrégation démontre qu'il est possible d'abandonner l'utilisation systématique de lois générales et de se focaliser sur les spécificités locales. La prolifération de données permet de disposer d'éléments de connaissance à grande échelle et il serait dommage de ne pas les utiliser.

Références bibliographiques

Andrienko N., Andrienko G., Stange H., Liebig T., Hecker D. (2012). Visual Analytics for Understanding Spatial Situations from Episodic MovementData. *KI - Künstliche Intelligenz*. Vol.26, n° 3. pp.241–251.

Annoni, A. (2005). European Reference Grids, proposal for a European Reference System. Proceedings of the European Reference Grid Workshop, 27-29 October 2003, Ispra (VA), Italy. European Commission, Joint Research Centre, Ispra (VA), EUR 21494 EN. p.189.

Antoni, J-P. Et Vuidel,G. (2010). Modéliser la ville. Forme urbaine et politiques de transport. *Economica: Méthodes et approches*, pp.50-77.

Apur-IAU. (2014). Abécédaire de la future Métropole du Grand Paris. STIPA. p.171

Armoogum, J., Hubert, J-P., Roux, S. (2010). Plus de voyages, plus de kilomètres quotidiens: une tendance à l'homogénéisation des comportements de mobilité des Français, sauf entre ville et campagne In : Le Jeannic Th., Roussel Ph., François D (eds). *La mobilité des Français panorama issu de l'enquête nationale transports et déplacements 2008*. Paris :La revue du CGEDD. pp.5-23.

Asakura, Y. Hato E. (2003). Tracking survey for individual travel behaviour using mobile communication instruments. *Transportation Research Part C: Emerging Technologies*, vol.12. n° 3-4. pp.273–291.

Authier, J.-Y. (2008). Les citadins et leur quartier. Enquêtes auprès d'habitants de quartiers anciens centraux en France. *L'Année sociologique*. Vol.58, n°1, pp.21-46.

Baas, K.G. (1980). Design of zonal systems for aggregate transportation planning models. *Transportation Research Record*. vol. 807, pp. 1-6.

Bacäini,F. (1995) . Les trajets domicile-travail en Ile-de-France ; contrastes entre catégories socio-professionnelles. *Economie et Statistique*. n°294-295.p.109-126.

Baro, J., Bonin O., Hubert J-P. (2016). Élaboration d'un zonage de tissus urbains : introduire de la structure dans un référentiel carroyé. *Revue Internationale de Géomatique*. Vol. 26, n°1. p.33-53.

Baro, j. (2015). Modélisation multi-échelle de la morphologie urbaine à partir de données carroyées de population et de bâti. Thèse de doctorat de sciences et technologie de l'information géographique. Paris : Université Paris Est. p.415.

Beaucire, Fr. Saint-Gerand, D. (2001). Les déplacements quotidiens, facteurs de différenciation socio-spatiale ? La réponse du périurbain en Ile-de-France. *Géocarrefour*. Vol. 76 n°4, pp.339-347.

Beckman, R. J., Baggerly, K. A., et McKay, M. D. (1996) Creating synthetic baseline populations, *Transportation Research Part A: Policy and Practice*. Vol. 30, N° 6, pp. 415-429.

Bertin, J.(1967). *Sémiologie graphique*. Paris :Mouton/Gauthier-Villars.p.431.

Bezes, P. (2009) Réinventer l'État. Les réformes de l'administration française (1962-2008). Paris, PUF : coll. « Le lien social ». p.519.

Bivand, R.S., S., Pebesam E., Gomez-Rubio,V. (2013). Hello World : Introducing Spatial Data. In : Gentleman, R. Hornik. K., Parmigiani G.G. *Applied Spatiale Data Analysis with R*. New-York:Springer. pp 1-16.

Bonin, O., Hubert, J-P. (2014). Modélisation morphogénétique de moyen terme des villes : une schématisation du modèle théorique de RITCHOT et DESMARAIS dans le cadre du modèle standard de l'économie urbaine. *Revue d'économie régionale et urbaine*. 2014/3. p.471-497

Bonnel P. (2001). Prévission de la demande de transport. Habilitation à diriger des recherches, Lyon : Université Lumière,p.409

Boulaahbal, M. (2001). Effet polarisant du lieu de travail sur le territoire de la vie quotidienne des actifs.Recherche-Transports-Sécurité.Vol. 73. p.43-63

Bouzouina, L., Quetelard, B. et Toilier, F. (2013). Emissions de CO_2 liées à la mobilité domicile-travail : une double lecture par le lieu de résidence et le lieu de travail des actifs à Lyon et à Lille. *Développement durable et territoires*.Vol. 4, n°3. p.45.

Bovy, P. H. L. and Jansen, G. R. M. (1983). Network aggregation effects upon equilibrium assignment outcomes:An empirical investigation. *Transportation Science*. Vol. 17, No. 3, pp. 240 262.

Caron, N. (2005). La correction de la non-réponse par repondération et par imputation, Série

Documents de Travail INSEE Méthodologie Statistique. n° M0502. p.48.

Caruso, G., Cavailhès, J., Peeters D., et al.(2015). Neighbourhoods make cities more sustainable! A 2D urban economics perspective. *Computers, Environment and Urban Systems*. Vol. 54. p. 82-94

Centre d'analyse stratégique (2007). Rapports énergies 2050. Paris : Centre d'analyse stratégique. p.392.

CEREMA (2015). Évaluation a posteriori des transports collectifs en site propre, note méthodologique. Lyon : Direction Technique Territoires et ville, Direction Territoriale Centre Est.p.110

Certu. (2003). Modélisation des déplacements urbains de voyageurs : guide des pratiques.p.244

Certu. (2008). L'enquête Ménages-Déplacement « standard CERTU : guide méthodologique. Paris : Edition du Certu. p.207.

Cervero, R. et Kockelman, K. (1997). Travel demand and the 3Ds : Density, diversity, and design. *Transportation Research Part D : Transport and Environment*.Vol.2, n°3, pp.199 – 219.

CGDD. (2010). La mobilité des Français, panorama issu de l'enquête nationale transports et déplacements 2008. La revue du Commissariat Général au Développement Durable. Paris : CGDD, p.226.

CGDD (2011). Le point sur la transposition de la directive européenne « Inspire »: le développement d'une infrastructure d'information géographique. *Recherche et Innovation*. n°82. p.4 [en ligne] (Consultée le 13/06/2016) < http://www.developpement-durable.gouv.fr/IMG/pdf/LPS82_-_Inspire.pdf>

CGEDD (2013). Le facteur 4 en France : la division par 4 des émissions de gaz à effet de serre à l'horizon 2050.Paris : revue du CGEDD. p.136.

Chang, K. T., Khatib, Z., and Ou, Y. (2002). Effects of zoning structure and network detail on traffic demand modeling. *Environment and Planning B: Planning and Design*. Vol. 29, No. 1, pp. 37-52.

- Choi, K. H.** (2003). Impact analysis of traffic zone size on transportation demand forecasting. Master's Thesis. Department of Transportation Engineering. Myongji University, Korea.
- Cirillo C., Cornélis E., Hubert J-P., Toint P.** (2004). Les enquêtes sur les comportements de mobilité, et après ?. *Reflets et perspectives de la vie économique*. Tome XLIII-n°4, p.111-121
- CLIP** (2012). Modes de vie et empreinte carbone : prospective des modes de vie en France à l'horizon 2050 et empreinte carbone. Paris : Pierre Barthélemy, p.132
<http://www.iddri.org/Publications/Les-cahiers-du-CLIP/Clip21_modes%20de%20vie%20prospective%202050.pdf>
- Collet, A.** (2010) Générations de classes moyennes et travail de gentrification : changement social et changement urbain dans le Bas Montreuil et à la Croix-Rousse, 1975-2005. Thèse de doctorat en sociologie. Lyon. Université Lumière-Lyon 2. p.658
- Conseil Général des Ponts et Chaussées** (2006). Démarche Prospective Transports 2050 : Elements de réflexion. La documentation Française. p.54
- Crawford, C. et Young, L.** (2004). A Spatial View of the Ecological Inference Problem. In : King, G., Rosen, O., Tanner M.A. (Eds). *Ecological Inference : New Methodological Strategies*. Cambridge University Press, pp.233-244
- Cresswell, T.** (2010). Mobilities I : Catching up. Progress. *Human Geography*. Vol. n°35. n°4. p.550-558
- Crevo, C. C.** (1991). Impacts of zonal reconfigurations on travel demand forecasts. *Transportation Research Record: Journal of the Transportation Research Board*. N°. 1305, pp. 72-80.
- Crozet, Y. Lopez-Ruiz, H.G. Chateau, B. Bagard, V** (2008). Scénarios LET-ENERDATA : Scénarios de Mobilité Durable Sous Contrainte d'un Facteur 4, Lyon : Rapport du Predit. p.220
- Cuillier, F. Hilal, M. Vienne, D.** (2009) Rapport du groupe de travail : statistiques et nouvelles tendances de localisation des populations et des activités sur le territoire. Paris: CNIS. p.63. [en ligne] (Consultée le 15/06/2016)
<http://www.cnis.fr/files/content/sites/Cnis/files/Fichiers/publications/rapports/2009/RAP_2009_117_populations_activites_territoire.PDF>

- Deming**, W. E. et Stephan, F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *The Annals of Mathematical Statistics*. Vol. 11, N° 4, pp. 427-444.
- De Palma**, A. et Zaouali, N. (2007). Monétarisation des externalités de transport : un état de l'art. Thema Working Paper. p.37.
- Desrosières**, A. (2002). Adolphe Quetelet. *Courrier des statistiques*. n°104. p.8.
- Desrosières**, A. (2003). Les qualités des quantités. *Courrier des statistiques*. n°105-106, pp.51-63.
- Deville**, F., Munizaga, M., Trépanier, M. (2012). Detection of activities of public transport users by analyzing smart card data. *Transportation Research Record: Journal of the Transportation Research Board*. n°2276. p.48-55.
- Ding**, C. (1994). Impact analysis of spatial data aggregation on transportation forecasted demand: A GIS approach. *Proceeding of the Urban and Regional Information Systems Association Conference*. 1994. Washington DC. pp.362-375.
- Duncan**, B., et Duncan, O. D. (1955). A methodological analysis of segregation indexes. *American sociological review*. Vol.20. n°2. pp. 210-217.
- Duncan**, O.D. et Davis, B. (1953). An Alternative to Ecological Correlation. *American Sociological Review*. n°18, pp.665-666.
- Dupuit**, J. (1995). De la mesure de l'utilité des travaux publics (1844). *Revue française d'économie*. Vol. 10. n°2. pp.55-94.
- Dupuy** G (1975), Une technique de planification au service de l'automobile, les modèles de trafic urbain , Document de travail, action concertée de recherches urbaines, Paris, 201. p.26
- Dupuy**, G.(1991). L'urbanisme des réseaux, théories et méthodes. Armand Colin, p.198
- Dupuy**, G.(1999). La dépendance automobile Symptômes, analyses, diagnostic, traitements. Paris : Anthropos :collection Villes. p. 160.
- Efron**, B. et Tibshirani, R.J. (1994). An Introduction to the Bootstrap. Chapman & Hall/CRC.p.456.

El Mashri, M.K. et al., (2015). Utilité des données billettiques pour l'analyse des mobilités urbaine : le cas rennais. In : Pumain, D. et Mattei, M-F. Données urbaines 7. Paris. p.77-p.86.

El Mashri M.K. (2013), Analyse et fouille de données de trajectoires d'objets mobiles. Thèse de doctorat en informatique et réseaux, Paris : Telecom Paris Tech, p.187.

Fall, B. (2012). Évaluation des performances d'un système de localisation de véhicules de transports guidés fondé sur l'association d'une technique radio ULB et d'une technique de retournement temporel. Thèse de doctorat: Electronique.Valenciennes: Université de Valenciennes et du Hainaut-Cambresis, p.174. [en ligne] <<https://tel.archives-ouvertes.fr/tel-00973788/document>>

Fen-Chong J. (2012). Organisation spatio-temporelle des mobilités révélées par la téléphonie mobile en Ile-de-France. Thèse de doctorat: Géographie. Paris: Université Paris I. p.217.

Frankhauser, P. et al. (2007). Vers des déplacements périurbains plus durables : proposition de modèles fractals opérationnels d'urbanisation. Rapport final PREDIT 3. p.115.

Frankhauser, P. (2013) Développement durable et système de peuplement : une approche conceptuelle multifractale. *50ème colloque de l'ASRDLF*, Jul 2013, Mons, Belgique.

Freedman D.A. (2001), Ecological inference and the ecological fallacy. International Encyclopedia of the Social & Behavioral Sciences. n°6. Elsevier. p.4027–4030

Fujita, M. (1989). Urban economic theory : land use and city. Cambridg Press. p.368

Fujita, M. et Thisse, J-F. (2003). Économie des villes et de la localisation. Bruxelles : De Boeck. p.562.

Fujita, M., Krugman, P.R., Venables, A.J.(2001). The spatial economy : cities, regions, and international trade.Massachusetts Institute of Technology Press. p.370.

Gallez C., Hivert L., (1998). BEED : mode d'emploi. Synthèse méthodologique pour les études « budget énergie environnement des déplacements ». Rapport de convention ADEME-INRETS. p.85.

Gallez, C. (2000). Indicateurs d'évaluation des scénarios d'évolution de la mobilité urbaine . Rapport de recherche pour le compte du PREDIT, INRETS-DEST, Arcueil, p.140.

Gascon, M.-O., Quetelard B., Patiès, C. (2009). Calcul a posteriori des distances dans les

enquêtes ménages déplacements, CERTU.p.45.

GIEC (2013). Résumé à l'intention des décideurs In : Stocker, T.F., Qin,D., Plattner G-K., et al. (eds.). Changements climatiques 2013: Les éléments scientifiques. Contribution du Groupe de travail I au cinquième rapport d'évaluation du Groupe d'experts intergouvernemental sur l'évolution du climat. Cambridge, Royaume-Uni et New York (État de New York), États-Unis d'Amérique.pp3-33.

Goodman, L.A. (1953). Ecological Regression and Behavior of Individuals. *American Sociological Review*. n°18.pp.663-664

Goodman, L.A. (1959). Some Alternatives to Ecological Correlation. *American Journal of Sociology*. n°64. pp.610-625.

Gotman A.M. (1990). Stratégies résidentielles, stratégies de la recherche, cité dans F. Dureau et al (dir.). Métropoles en mouvement. Une comparaison internationale. Paris. Anthropos, coll. Villes/IRD. p132.

Gottman, J. (1966). Essais sur l'aménagement de l'espace habité. Paris : Mouton. p.347.

Gotway, C. A., et Young, L. J. (2002). Combining incompatible spatial data. *Journal of the American Statistical Association*. Vol. 97. n°458. p. 632-648.

Grafmeyer, Y. (1991). Habiter Lyon. Milieux et quartiers du centre-ville.Lyon :Presses Universitaires de Lyon.p.150.

Guillou, M. et Colombani, P. (2008). Comité opérationnel « Recherche » du Grenelle de l'Environnement. La documentation française - rapports publics.p.87 [en ligne] (Consultée le 23/03/2016) <<http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/084000527.pdf>>

Hägerstrand, T. (1975). Space, time and human conditions.

Hivert. L., Nguyen. Q-N., Merle, N.et al. (2014) Le diagnostic énergie émissions des mobilités (Deem) : du pari méthodologique à la standardisation. In : Mobilité en transitions. De la connaissance à l'aide à la décision.CEREMA, pp. 207-220.

Hubert, H et Arabie, P.(1985). Comparing Partitions. *Journal of classification*.Vol. 2. n°1.pp.193-218

Hubert, J-P., Pistre, P., Madre, J-L. (2016). L'utilisation de l'automobile par les ménages dans les territoires peu denses : analyses croisées par les enquêtes sur la mobilité et le Recensement de la population. *Economie et statistique*. n°483-484-485. pp.179-203.

Hubert, J-P. Delisle, F. (2010). L'allongement des déplacements quotidiens contribue à l'émergence d'espaces urbains multipolaires, tandis que la mobilité baisse au centre des grandes agglomérations. In : Le Jeannic Th., Roussel Ph., François D (eds). La mobilité des Français, panorama issu de l'enquête nationale transports et déplacements 2008. La Revue du CGDD.

Jeon, J. H., Kim, D-K., Kho, S. Y., et al. (2010). Interactions of Aggregated Zoning and Network Systems: A Case Study of Seoul City. *Journal of the Eastern Asia Society for Transportation Studies*. Vol.8. p16.

Jeon, J. H., Kho, S. Y., Park, J. J., et Kim, D. K. (2012). Effects of spatial aggregation level on an urban transportation planning model. *KSCE Journal of Civil Engineering*. Vol 16. n°5. pp.835-844.

Jutand, F., (2015). Ouverture des données en transport. Paris: Union des Transports Publics et ferroviaires.p.147 [en ligne] (Consultée le 13/02/2016) <http://www.opendatafrance.net/wp-content/uploads/2015/07/vf_rapport_jutand.pdf>

Kellerer, H., Pferschy, U., Pisinger, D. (2004). Knapsack problems. Springer Science & Business Media.p.557.

Khatib Z., Chang K., Ou Y. (2001). Impacts of analysis zone structures on modeled statewide trafic. *Journal of Transportation Engineering*. Vol. 127, n°1, pp31-38.

King, G. (1997a). A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data. Princeton University Press. p.337.

King, G. (1997b). Aggregation Problems In G.King. A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data. Princeton University Press,p.337.

Koning M (2012), Essais sur la congestion dans les transports à Paris. Thèse de doctorat.Paris: Sciences économiques. Paris : Université Panthéon-Sorbonne - Paris I. p.247.

Korsu, E., et Massot, M. H. (2006). Rapprocher les ménages de leurs lieux de travail: les

enjeux pour la régulation de l'usage de la voiture en Île-de-France. *Les cahiers scientifiques du transport*. n°50, pp.61-90.

Kousser, J.M. (2001). Ecological Inference from Goodman to King. *Historical Methods*. Vol. 34. n°3.pp.101-126.

Krugman, P. (1995). Development, Geography, and Economic Theory. Cambridge (Mass.). Massachusetts Institute of Technology Press.p.120.

Lakshmanan, T.R. et Anderson, W.P (2004). Contextual Determinants of Transport Infrastructure Productivity: the Case for a New Modeling Strategy. International Conference on Measurement and Management of Infrastructure.Jönköping International Business School, 29-31 octobre 2004. Sweden : Jönköping.p.30.

Le Corbusier (Charles-Édouard Jeanneret-Gris) (1957). La Charte d'Athènes. Paris : éd. De Minuit. coll. « Points »p.190.

Le Gléau, J-P., Royer, J-F.(2011). Le centre d'accès sécurisé aux données de la statistique publique française : un nouvel outil pour les chercheurs. *Courrier des statistiques*. n°130.p.5.

Leroux, P. (1969). L'équipement des français en biens durables fin 1968. *Economie et statistique*.Vol.3. n°3, pp.65-68.

Létinois, L. (2012). Méthodologie de répartition spatiale de la population. Laboratoire Central de Surveillance de la qualité de l'air.p.87 [en ligne] (Consultée le 14/06/2016) <<http://www.lcsqa.org/rapport/2012/ineris/methodologie-repartition-spatiale-population>>

Leurent F (2012). Modelling a vehicle-sharing station as a dual waiting system: stochastic framework and stationary analysis .*Hal*.p-19.

Lipatz, J-L. (2011). Les carroyages : des zonages pour tuer les zonages. Séminaire Belgrand, 23 juin, Marne-la-Vallée,ENSG.

Litman, T.A., (2009) Transportation Cost and Benefit Analysis : techniques, estimates and implications. VictoriaTransport Policy Institute.p.20.

Loonis, V. (2009). L'échantillonnage de la théorie à la pratique. *Courrier des statistiques*. N°126. p.8 [en ligne] (Consultée le 15/06/2016) <http://www.insee.fr/fr/ffc/docs_ffc/cs126e.pdf>

Lopez-Ruiz H.G (2009). « Environnement et mobilité 2050 : des scénarios sous contrainte du facteur 4 (-75 % de CO_2 en 2050) ». Thèse de doctorat : sciences économiques. Lyon : Université de Lyon,p.213.

Marceau, D. J.(1999), The scale issue in social and natural sciences. *Canadian Journal of Remote Sensing*. Vol. 25.n° 4, p. 347-356. [en ligne] <http://www.developpement-durable.gouv.fr/IMG/pdf/III_-_Aires_d_etude_V_01_10_2014.pdf>

Markoff, J., et Shapiro, G. (1973). The Linkage of Data Describing Over-lapping Geographical Units. *Historical Methods Newsletter*. Vol°7. n°1. pp 34–46.

Martínez, L. M., Viegas, J. M., and Silva, E. A. (2007). Zoning decisions in transportation planning and their impact on the precision of results. *Transportation Research Record: Journal of the Transportation Research Board*. n°1994.pp. 58-65.

McCann, P. (2001) *Urban and Regional Economics*. Oxford: Oxford University Press.

Meadows, D. et al. (1972). *The Limits to Growth*. New York : Universe Books. p.212.

Milion C. (2015). Méthodes et modèles pour l'étude de la mobilité des personnes par l'exploitation de données de radiotéléphonie. thèse de doctorat. Univ. Paris Est. p.356.

Morris, J. N., et Hardman, A. E. (1997). Walking to health. *Sports medicine*. Vol. 23 n°5. pp.306-332.

Motte, B. (2006). La dépendance automobile pour l'accès aux services aux ménages en grande couronne francilienne. Thèse de doctorat en sciences économiques, Université Paris 1 - Panthéon Sorbonne. p.383.

Munoz, D. Bouchereau, F. Vargas, C. et al. (2009). *Position Location Techniques and Applications*. Academic Press. Burlington. p.275.

Nguyen-Luong D., Allilo R. (2014), SMOOTH : suivi de la mobilité par GPS pour évaluer des offres de transport nouvelles dans les espaces d'habitat peu denses. Rapport du PREDIT. p.136 [en ligne] (Consultée le 16/06/2016) <www.predit.prd.fr/predit4/documentFo.fo?cmd=visualize&inCde=46500>

Nicolas , J-P. Vanco, F. ,V erry, D. (2012). Mobilité quotidienne et vulnérabilité des ménages. *Revue d'Économie Régionale & Urbaine*. Vol.2012. n°1. pp. 19-44.

- Nicolas J-P, Pochet P, Poimboeuf H et al. (2001).** Indicateurs de mobilité durable. Application à l'agglomération de Lyon. Lyon : Coll. Etudes & Recherches n°16, Laboratoire d'Economie des Transports, p.130.
- Nix, S.(2009).** Quantifying the impacts of the modifiable areal unit problem in travel demand models. Master of Engineering, Toronto : Ryerson University,p.61.
- Ntziachristos, L. et Samaras, Z. (2000).** COPERT III, Computer programme to calculate emissions from road transport, Methodology and emission factors. Technical report. No 49, European Environment Agency.p.86.
- O'Neil, R et King, A. W. (1998).** Homage to St. Michael; or, why are there so many books on scale. *Ecological Scale: Theory and Applications*, Columbia University Press, New York, 1998, Vol. 316.p.3-15.
- O'sullivan, A. (2007).** Urban economics. McGraw-Hill/Irwin.
- OCDE (2014).** La demande de transport à long terme. In : OCDE/FIT. Perspectives des transports FIT 2013 :financer les transports.p.62.
- Openshaw, S. (1977).** Optimal zoning systems for spatial interaction models. *Environment and Planning A*. Vol.9. n°2. pp. 169-184.
- Openshaw, S. (1984).** Modifiable areal unit problem (concepts and techniques in modern geography), Geo Books,Norwich.
- Openshaw, S. and Taylor, P. (1979).** A million or so correlation coefficients: Three experiments on the modifiable areal unit problem. *Statistical Applications in the Spatial Sciences*. Vol.21. pp. 127-144.
- Orfeuill, J.P.(2008).** Une approche laïque de la mobilité. Paris :Descartes and Cie.p.173.
- Ortúzar, J. de D., Willumsen, L.G. (2011).** Modelling Transport, fourth ed. John Wiley & Sons Ltd, Chichester.p.607.
- Parry I, Harrington W, et Walls M. (2007).** Automobile externalities and policies. *Journal of Economic Literature*.Vol. 65. pp.373–399.
- Polèse, M. Shearmur, R. Terral, L. (2015).** Economie urbaine et régionale : Géographie économique et dynamique des territoires. Paris : Economica.p.413.

Quetelard, B. (2010). Se rendre au travail ou faire ses courses motive toujours un déplacement quotidien sur deux. Le recours à la voiture se stabilise. In : Le Jeannic Th., Roussel Ph., François D (eds), La mobilité des Français. Panorama issu de l'enquête nationale transports et déplacements 2008, La revue du CGDD, p. 25-48.

Quételet, A. (1835). Sur l'homme et le développement de ses facultés ou Essai de physique sociale. Paris, Bachelier.p.327.

Raux, C., Godinot, C., et Masson, S. (1996). Développement d'un modèle stratégique de simulation des déplacements, vingt ans de rétrospective à travers les enquêtes déplacements de l'agglomération lyonnaise (1976-1985-1995), rapport d'étape. Lyon, LET, SEMALY.

Robinson, W. S.(1950) Ecological Correlations and the Behavior of Individuals. *American Sociological Review*.n°15. p. 351-357.

Rocci, A. (2007). De l'automobilité à la multimodalité ? Analyses sociologique des freins et leviers au changement de comportements vers une réduction de l'usage de la voiture : le cas de la région parisienne et perspective internationale. Thèse de doctorat en sciences humaines et sociales. Paris. Université de Paris 5. p.544.

Sorokin, P.A. (1927). Social Mobility, New York: Harper and Brothers.

Steg, L. (2003). Can public transport compete with the private car?. *Iatss Research*. Vol 27. n°2. p.27-35.

Tamblay, S., Galilea, P., Iglesias, P., Raveau, S., et Muñoz, J. C. (2015). A zonal inference model based on observed smart-card transactions for Santiago de Chile. *Transportation Research Part A: Policy and Practice*. Vol.84. pp.44-54.

Tao Pham, T.H. (2016). Apports et difficultés d'une collecte des données à l'aide du récepteur GPS pour réaliser une enquête sur la mobilité. Thèse de doctorat en transports, Paris : Paris-Est.

Tobler W. R. (1987). Experiments in Migration Mapping by Computer. *The American Cartographer*. Vol.14. n° 2. pp.155–163.

Tobler, W. (1979). Smooth Pycnophylactic Interpolation for Geographical Regions. *Journal of the American Statistical Association*. Vol 74. n° 367. p.519-536.

Tuot T. (2007). Le Grenelle Environnement, rapport du rapporteur général. Paris : La

documentation française. p.39.

Vallée, J., Commenges, H., Perchoux, C. et al. (2015). Mobilité quotidienne et accès potentiel aux ressources urbaines en Île-de-France : quelles inégalités sociales?. In Matel. F.F., Pumain, D. (dir.). Données urbaines 7. Paris : Economica/Anthropos, p.87-95.

Visser, M. (1994). Beyond the ecological fallacy. *Quality and Quantity*. n°28.pp.435-444.

Von Thünen, J.H. (1826), l'État isolé en relation avec l'agriculture et l'économie nationale.

Wenglenski, S. (2003). Une mesure des disparités sociales d'accessibilité au marché de l'emploi en Ile-de-France. Thèse de doctorat : nouveau régime, en urbanisme, aménagement et politiques urbaines . Paris : Université de PARIS XII. p.373.

Wiel, M. (1999). La Transition urbaine: ou le passage de la ville pédestre à la ville motorisée. Sprimon : Editions Mardaga. p.162.

Zähavi, Y. et Talvitie A. (1980). A Regularities in travel time and money expenditures. Transportation Research Record. n° 750.

Zhang, M. et Kukadia, N. (2005). Metrics of urban form and the modifiable areal unit problem. *Transportation Research Record: Journal of the Transportation Research Board*. Vol.1902, pp. 71-79.

Références internet

ADEME, (2014). Les chiffres clefs. [en ligne] (consultée le 01/02/2015) <www.ademe.fr/chiffres-cles-climat-air-energie-2014>

ADEME. (2014). Les chiffres clefs. [en ligne] (consulté le 01/02/2015) <www.ademe.fr/chiffres-cles-climat-air-energie-2014>

ADEME. Blog Transport du futur [en ligne] (Consultée le 04/04/2016) <<http://transportsdufutur.ademe.fr/2013/06/appel-a-projet-mobilites-2013-connaître-pour-mieux-agir.html>>

ADEME. Comparateur Eco-déplacements [en ligne] consultée le 22/05/2016. <<http://quizz.ademe.fr/eco-deplacements/comparateur/>>

ADEME. Le changement climatique : comprendre ses causes et ses conséquences pour mieux agir [en ligne] (Consultée le 25/05/2016) <<https://ile-de-france.ademe.fr/sites/default/files/files/DI/Changement-climatique/guide-pratique-changement-climatique.pdf>>

Bulletin officiel. [En ligne] (Consultée le 28/03/2016) <http://www.bulletin-officiel.developpement-durable.gouv.fr/fiches/BO201412/met_20140012_0100_0047.pdf>

CASD. Présentation [en ligne] (Consultée le 14/06/2016) <<https://casd.eu/>>

CEREMA. (2013). Les enquêtes déplacements « standard CERTU ». Certu.p.15 [en ligne] (Consultée le 25/03/2016) <http://www.territoires-ville.cerema.fr/IMG/pdf/Les_enquetes_Deplacements_site_web_cle5964c6.pdf>

CEREMA. EMD, EDVM et EDGT : méthodes et guides [en ligne] (Consultée le 26/03/2016) <<http://www.territoires-ville.cerema.fr/emd-edvm-et-edgt-methodes-et-guides-a679.html>>

CEREMA. Portail billettique [en ligne] (Consulté le 24/03/2016) <<http://www.billettique.fr/spip.php?article205>>

Certu. Note de synthèse 02 du CERTU « Mobilité et transports » - Février 2008 – [en ligne] (Consultée le 06/06/2016) <http://www.bv.transports.gouv.qc.ca/mono/0968315/02_Note_de_synthese_02.pdf>

CNIS. Fiches de présentation d'enquêtes statistiques pour examen d'opportunité [en ligne] (Consultée le 16/06/2016) <http://www.cnis.fr/files/content/sites/Cnis/files/Fichiers/commissions/territoires/2013/documents_preparatoires/DPR_2013_2e_reunion_COM_territoires_enquetes_opportuni te.PDF>

DRIEA. Direction Regionale et interdépartementale de l'Equipement et de l'Aménagement d'Ile-de-France [en ligne] (Consultée le 28/03/2016) <http://www.driea.ile-de-france.developpement durable.gouv.fr/presentation-du-sdaurp-de-1965-a970.html#sommaire_3>

DRIEA, Le SDRI de 1194 quel bilan ? [en ligne] (Consultée le 15/06/2016) <http://www.driea.ile-de-france.developpement-durable.gouv.fr/IMG/pdf/151004-region-etat-cesr-pour-colloque-211004_cle53bb51.pdf>

Finance.gouv.fr, open data [en ligne] (Consultée le 11/06/2016) <<http://proxypubminefi.diffusion.finances.gouv.fr/pub/document/18/20339.pdf>>

Géolocalisation : votre portable vous suit à la trace [en ligne] (Consultée le 15/05/2016) <<http://reseau.echelon.free.fr/reseau.echelon/geolocalisation.htm>>

Ifsttar. Mobilletic [en ligne] (Consultée le 29/03/2016) <<http://mobilletic.ifsttar.fr/>>

IGN. Projection INSPIRE. [en ligne] (Consultée le 06/06/2016) <http://geodesie.ign.fr/contenu/fichiers/documentation/rgf93/Projections_INSPIRE_ETRS89.pdf>

Insee. Bulletin du recensement 2016 [en ligne] (Consultée le 15/06/2016) <<http://www.le-recensement-et-moi.fr/rpetmoi/pdf/Bulletin-individuel.pdf>>

Insee. Connaître et comprendre le répertoire SIRENE [en ligne] (Consultée le 04/04/2014) <<http://www.insee.fr/fr/bases-de-donnees/default.asp?page=sirene/institutionnel.htm>>

Insee. Définition de SIRENE [en ligne] (Consultée le 14/05/2016) <<http://www.insee.fr/fr/methodes/default.asp?page=definitions/sys-inf-rep-nat-ent-etab.htm>>

Insee. Documentation générale sur le carroyage [en ligne] (Consultée le 12/04/2014) <http://www.insee.fr/fr/themes/detail.asp?reg_id=0&ref_id=donnees-carroyees&page=donnees-detaillees/donnees-carroyees/donnees_carroyees_doc.htm>

Insee. Dossier complet de l'Île-de-France. [en ligne] (Consultée le 01/06/2015) <http://www.insee.fr/fr/themes/dossier_complet.asp?codgeo=REG-11>

Insee. Fiche pondération [en ligne] (Consultée le 05/05/2016) <<http://www.insee.fr/fr/ppp/bases-de-donnees/recensement/resultats/doc/pdf/fiche-ponderation.pdf>>

Insee. La précision des résultats du recensement [en ligne] (Consultée le 14/06/2016) <<http://www.insee.fr/fr/ppp/bases-de-donnees/recensement/resultats/doc/pdf/fiche-precision.pdf>>

Insee. Le nouveau mode de recensement:élaboration à travers les tests [en ligne] (Consultée le 14/06/2016) <<http://www.insee.fr/fr/ppp/sommaire/imeths01e.pdf>>

Insee, Les exploitations principales et complémentaires [en ligne] (Consultée le 16/06/2016) <http://www.insee.fr/fr/methodes/sources/pdf/RP_exploitations_principales_et_complementaires.pdf>

Insee. Pour comprendre le recensement Annexe B2 : le plan de sondage dans les communes de 10 000 habitants ou plus [en ligne] (Consultée le 06/05/2016) <<http://www.insee.fr/fr/ppp/sommaire/imeths01g.pdf>>

Insee. Documentation du fichier détail des mobilités professionnelles [en ligne] (Consultée le 16/06/2016) <http://www.insee.fr/fr/ppp/bases-de-donnees/fichiers_detail/RP2008/doc/contenu_RP2008_mobpro.pdf>

Insee. Données carroyées- Pourquoi des carroyages [en ligne] (Consultée le 05/04/2016) <http://www.insee.fr/fr/themes/detail.asp?reg_id=0&ref_id=donnees-carroyees&page=donnees-detaillees/donnees-carroyees/donnees_carroyees_carroyage.htm>

Inspire (2007). Directive 2007/2/CE du parlement européen et du conseil du 14 mars 2007. Journal officiel n°108 du 25 avril 2007.p.14 [en ligne] (consultée le 13/05/2016) <<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:FR:PDF>>

Inspire. Data Specification [en ligne] (Consultée le 13/06/2016)) <<http://inspire.ec.europa.eu/index.cfm/pageid/2/list/7>>

Journal officiel de l'Union européenne. Protection des données dans le secteur des communications électroniques [en ligne] (Consultée le 24/03/2016)

<http://europa.eu/legislation_summaries/information_society/legislative_framework/124120_fr.htm>

Journal Officiel, Arrêté du 1er Février 2016 [en ligne] (Consultée le 16/06/2016)
<https://www.legifrance.gouv.fr/affichTexte.do;jsessionid=8EBD409457CDC7FA261A348D1DE36AA6.tpdila12v_2cidTexte=JORFTEXT000032000138&dateTexte=&oldAction=rechJO&categorieLien=id&idJO=JORFCONT000031999976>

Légifrance [en ligne] (Consulté le 12/02/2016)
<https://www.legifrance.gouv.fr/affichTexte.do;jsessionid=5CC613290E3003C2E829BB6E02DEE09A.tpdjo10v_3?cidTexte=JORFTEXT000019109412&dateTexte= >

Les Ateliers du Grand Paris. La métropole polycentrique [en ligne] (Consultée le 21/02/2016)
<<http://www.ateliergrandparis.fr/aigp/conseil/LC1/PresentationRSH.pdf>>

MEDDE, Le CO₂ et le transport [en ligne] (consultée le 29/06/2016)
<http://www.developpement-durable.gouv.fr/IMG/pdf/Guide_Information_CO2-2.pdf>

MEDDE. Chiffres clefs du climat en France et dans le monde [en ligne] (Consultée le 16/06/2016)
<http://www.developpement-durable.gouv.fr/IMG/pdf/Chiffres_cles_du_climat_en_France_et_dans_le_monde_2015.pdf>

MEDDE. Emissions de la France [en ligne] (Consultée le 15/06/2016) <<http://www.developpement-durable.gouv.fr/Emissions-de-la-France,33791.html>>

Ministère de l'environnement, de l'énergie et de la mer [En ligne] (Consultée le 28/03/2016)

Ministère de l'environnement, de l'énergie et de la mer. Les fiches outils [En ligne] (Consultée le 28/03/2016) <<http://www.developpement-durable.gouv.fr/Les-fiches-outils-du-referentiel-d.html>>

Omnil, méthodologie EGT [en ligne] (Consultée le 15/06/2016)
<<http://www.omnil.fr/spip.php?article122>>

Omnil. Enquête Globale Transport [en ligne] (Consultée le 31/06/2026)
<http://www.omnil.fr/IMG/pdf/egt2010_enquete_globale_transports_-_2010.pdf> cité p.3

PREDIT. Bilan thématique du PREDIT Groupe 6 « Politiques de transport » [en ligne] (Consultée le 29/03/2016) <<http://www.predit.prd.fr/predit4/bilan-predit4-go6>>

PREDIT. Connaître pour mieux agir [en ligne] (Consultée le 03/03/2016)

<<http://www.predit.prd.fr/predit4/appel-a-proposition/44471>>

Support Google [en ligne] (consultée le 28/03/2016)

<<https://support.google.com/accounts/answer/3118687?hl=fr>>