Transfert de support de données de déplacements : quelles données de contrôle, et jusqu'à quelles échelles ?

Cyril Pivano, Olivier Bonin, Jean-Paul Hubert

▶ **To cite this version:**

Cyril Pivano, Olivier Bonin, Jean-Paul Hubert. Transfert de support de données de déplacements : quelles données de contrôle, et jusqu'à quelles échelles ?. Journées de méthodologie statistiques, Mar 2015, PARIS, France. 26p. hal-01217297

HAL Id: hal-01217297
https://hal.science/hal-01217297

Submitted on 22 Oct 2015

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.

TRANSFERT DE SUPPORT DE DONNÉES DE DÉPLACEMENTS: QUELLES DONNÉES DE CONTRÔLE, ET JUSQU'À QUELLES ÉCHELLES ?

Cyril PIVANO(*), Olivier BONIN (**), Jean-Paul HUBERT (***)

*(*)*Université Paris Est – *IFSTTAR – LVMT UMR T 9403*
cyril.pivano@enpc.fr
(**) Université Paris Est – *IFSTTAR – LVMT UMR T 9403*
olivier.bonin@ifsttar.fr
(***) Université Paris Est – IFSTTAR – *DEST*
jean-paul.hubert@ifsttar.fr

Résumé

Nous présentons dans ce papier plusieurs méthodes de transfert de support de données de déplacement. Le support spatial est l'objet théorique sur lequel est localisée géographiquement une donnée spatialisée. Les analyses pouvant être effectuées dépendent de sa finesse. Pour se dispenser de faire de nouvelles enquêtes lorsque le maillage est inadapté ou insuffisant nous transférons des données de déplacement sous la forme de matrice origine/destination d'un support vers un autre, l'objectif étant de préserver la qualité de l'information. Pour ce faire nous utilisons des données auxiliaires permettant de plus ou moins bien contrôler la répartition socio-spatiale des origines/destinations. Nous traitons ici du fichier détail des mobilités professionnelles du recensement qui apporte une connaissance sur les déplacements domicile-travail entre communes. Nous proposons plusieurs méthodes pour transférer ces données sur le carroyage de population de 1 km.

Mots-clés

Transfert, Désagrégation, Support spatiaux, Maillage, Mobilité

Introduction de la problématique de transfert de données de déplacements

En sciences humaines et sociales nous traitons très généralement d'objets d'étude localisés dans l'espace géographique supposé continu, et donc décrit par des coordonnées dans un système de références spatiales. Toutefois, lorsque nous dénombrons un effectif quel qu'il soit il est toujours compris dans un ensemble territorial, une maille, dont l'ensemble forme le support spatial de l'information. Le choix de ce support ne va pas de soi. Le support géographique a un effet sur l'observation du phénomène, ce problème est identifié sous le nom de *Modifiable Areal Unit Problem* (Openshaw,S. 1977, 1978, 1979,1981, 1984,1984a,1984b, Arbia G., 1989).

Lorsque l'objet étudié est mobile, les données de mobilité origine/destination apportent une information multi-localisée pour décrire des déplacements. Le choix du support est contraint par ces deux semis de points: celui des origines et celui des destinations. Un support unique apporte une contrainte à la fois sur les origines et sur les destinations. Un support dont on maîtriserait l'effet de MAUP des origines peut tout aussi bien être moins pertinent pour les destinations.

Les questions de connaissance de la mobilité des individus ont évolué avec l'apparition des nouvelles problématiques liées aux enjeux environnementaux et à la pénurie des ressources fossiles avec des impacts sur les besoins en précision géographique. L'analyse des aires de rabattement potentielles des stations de transports en commun demandera de disposer d'un support centré autour des stations d'entrées et de sorties. Si on s'intéresse à la qualité de l'air, nous savons que les émissions de gaz polluants sont maximales lorsque le moteur à explosion des véhicules est froid. Il convient alors de distinguer les premiers kilomètres des sorties journalières. Un maillage fin sur les zones résidentielles est judicieux, les mailles pouvant être plus grandes ailleurs. Nous observons bien que le choix du support est non seulement différent selon les thématiques étudiées mais aussi qu'il est intéressant de disposer de plusieurs support au sein d'une même analyse.

Pour se dispenser de faire de nouvelle enquête lorsque le maillage est inadapté ou insuffisant nous proposons dans ce papier de transférer les données de mobilité vers un autre support, l'objectif étant de préserver la qualité de l'information. Pour ce faire nous utilisons des données auxiliaires permettant de plus ou moins bien contrôler la répartition spatiale des origines/destinations.

Le processus de transfert consiste soit à transférer directement l'information d'un support à un autre, soit à désagréger les flux à l'aide d'un ou plusieurs supports auxiliaires pour construire par agrégation le maillage souhaité. Ces supports auxiliaires servent de données de contrôle construisant la répartition spatiale voulue.

Derrière notre objectif pratique de transfert de données de mobilités se pose la question de la sélection des données de contrôle. Ensuite, il faut réussir à estimer la qualité de la désagrégation. Jusqu'à quelle échelles pouvons-nous descendre ?

L'objectif est de mettre à disposition des outils permettant de changer de support spatial une fois la donnée de mobilité construite sur l'un d'entre eux. Nous nous attachons également à estimer de la qualité des résultats obtenus.

Nous traitons ici du fichier détail des mobilités professionnelles du recensement qui apportent une connaissances sur les déplacements domicile-travail à l'échelle des communes. Nous proposons plusieurs méthodes pour transférer ces données sur le carroyage de population de 1 km. Le choix de ce second support répond aux incitations de la directive européenne INSPIRE.

Comme nous l'avons dit, certaines opérations de transfert proposées utilisent des supports spatiaux auxiliaires plus fins pour ensuite ré-agréger l'information. Nous cherchons alors à estimer la qualité de plusieurs méthodes de désagrégation en sachant que les résultats d'une désagrégation sophistiquée seront par la suite lissés par le processus de ré-agrégation. Il est donc inutile d'effectuer trop d'efforts si une méthode plus simple apporte la même qualité statistique. Les résultats proposés sont confrontés aux observations des zones de sondage de l'Enquête Globale d'Île de France en Seine-et-Marne.

I Définitions, méthodes et enjeux généraux du processus de transfert de données de déplacements

1.1 Définitions et méthodes

Lors d'un processus de transfert, nous souhaitons transposer des données d'un support à un autre. Le support spatial de l'information à transférer est nommé support source. C'est celui qui porte l'information que l'on souhaite transférer. Le support sur lequel nous souhaitons disposer des informations est quant à lui nommé support cible.

Un support polygonal peut être irrégulier ou régulier. Dans le premier cas, il est composé de cellules de formes et de tailles différentes ; c'est ce type qui est généralement nommé maillage, tandis que dans le

second cas, c'est une répétition de forme régulière. Si les cellules sont de forme carrée, nous pouvons appelé ce support un carroyage.

La relation entre les supports source et cible est un élément important à prendre en compte lors de la méthode de transfert. Lorsque les mailles d'un support sont emboîtées dans un autre support nous disons que ce sont des maillages hiérarchiques, emboîtés ou encore alignés. Les communes sont emboîtées dans les départements par exemple. Au contraire, quand les mailles se superposent, comme dans le cas des départements et des bassins versants, nous parlons de maillages non-alignés. Lorsque nous voulons effectuer un transfert entre deux supports non-alignés nous devons passer par un support auxiliaire dont la granulométrie permet par agrégation de reconstituer le support cible souhaité. Ce support auxiliaire doit également servir de données de contrôle. L'opération de transfert consiste dans un première temps à désagréger l'information du support source au support auxiliaire. Puis nous agrégeons l'information du support auxiliaire pour reconstituer le maillage cible. L'opération d'agrégation n'est pas difficile méthodologiquement, il faut veiller toutefois au phénomène de lissage de l'information. En revanche, la désagrégation est une opération plus complexe. L'objectif est de reproduire la structure interne de la maille source en répartissant l'information sur le support auxiliaire. Le problème étant que nous ignorons *a priori* cette structure.

Pour construire la structure interne d'une maille du support source il faut faire l'hypothèse d'une distribution spatiale. L'hypothèse d'équi-répartition dans la maille est certainement le cas le plus simple. Dans ce cas nous pouvons utiliser la pondération surfacique simple (Simple Areal Weighting). La ventilation de la variable se fait proportionnellement à la surface d'intersection d'une maille source avec le maillage cible ou auxiliaire. (Goodchild M., Lam N., 1980)
D'autres distribution peuvent être utilisée. Bonin (Bonin O., 2012) propose des formules mathématiques (intégrales doubles de Poisson) prêtes à être paramétrées (moyenne, écart-type) selon l'hypothèse de distribution voulue. Nous comprenons que le choix des paramètres doivent être savamment choisi. Et ceux pour chacune des mailles du support source.
Pour ne pas devoir constituer arbitrairement la structure interne des mailles sources à reproduire, nous devons nous servir d'informations auxiliaires apportant une structure spatiale. Nous utilisons une information auxiliaire entretenant une relation de corrélation avec la variable à désagréger afin de prédire sa présence ou son degré de concentration à l'intérieur des mailles sources. La variable à désagréger est ventilée proportionnellement à la variable de contrôle contenue par le support auxiliaire. Par exemple, afin de transférer une population d'un maillage à un autre, il est possible de ventiler l'effectif de population proportionnellement à la surface occupée par le bâti dans une maille cible (Plumejeaud C., 2009). Nous nous servons des associations spatiales pour construire la sous-structure.
Les associations spatiales peuvent également s'effectuer par strate si l'on dispose de suffisamment d'information caractérisant les mobiles étudiés. La spécificité de l'étude des méthodes de transfert appliquées à des données mobiles est l'apport d'informations supplémentaires. Les structures à reproduire peuvent dépendre de l'éventuel interdépendance entre l'origine et la destination des déplacements.

1.2 Données sources et données auxiliaires

1.2.1 Présentation des données sources

Les données décrivant un déplacement se caractérisent au minimum par une zone d'origine et une zone de destination. Les déplacements enquêtés sont construits et diffusés en les agrégeant en flux. Il est créé une matrice dénombrant les déplacements pour chaque couple de maille. Ce sont ces matrices qui contiennent l'information source que nous souhaitons transférer sur une autre matrice. La méthodologie de transfert est identique selon le type de déplacement. Qu'il s'agisse de déplacements domicile-travail, de migrations résidentielles, de migrations touristiques, de voyages d'affaires ou même de flux monétaires, téléphoniques et internet il conviendra de disposer d'éléments de connaissance corrélés spatialement à la donnée source pour inférer des zones d'émissions et de destinations. De plus, la connaissance du système de transport utilisable par le mobile étudié apporte des contraintes à respecter améliorant la précision du processus. Nous pouvons nous demander si la position dans la zone d'origine explique la position dans la zone de destination, l'inverse étant aussi possible. Par exemple lorsque nous étudions un marché locatif et que nous considérons la demande comme mobile, c'est-à-dire comme des individus provenant d'un lieu de résidence antérieur nous pouvons supposer que les processus de choix résidentiels diffèrent selon leur origine. Les déterminants des migrations résidentielles ne sont pas les mêmes pour les courtes et grandes distances (Debrand T., Taffin, C.,2005).

Enfin, les caractéristiques intrinsèques du mobile peuvent également être d'une aide. Nous présentons ces caractéristiques comme des données auxiliaires même si la discussion peut être ouverte.

1.2.2 Données de contrôle.

Les données de déplacements peuvent être caractérisées par un ensemble d'attributs, l'itinéraire des déplacements, les motifs, le mode de transport, l'heure de départ et d'arrivée, le profil socio-économique. Il est alors possible de construire des matrices en distinguant chacune de ces caractéristiques et de les traiter indépendamment les unes des autres avec des variables de contrôles adéquates.
Par exemple, si nous prenons le cas d'une enquête sur les migrations touristiques nous pouvons construire les lieux de destination à l'aide d'informations sur le parc hôtelier. Ainsi dans une zone ou l'on sait qu'il y a tant d'individu venant y passer un séjour, nous pouvons estimer qu'ils vont se rendre dans un des hôtel de la zone. Ainsi, nous pouvons utiliser le semis de points du parc hôtelier de la zone comme support cible et nous ne plaçons donc pas totalement au hasard les lieux de destinations. Nous pouvons aussi choisir des variables de contrôle nous permettant d'associer les données du support source au support cible. En disposant des caractéristiques socio-économiques et du type de l'offre nous pouvons inférer des comportements du type: les plus riches iront dans le parc hôtelier de luxe et les moins riches dans les auberges de jeunesse.

Aussi, nous pouvons utiliser différents supports auxiliaires simultanément ou successivement pour contrôler le processus. Pour reprendre notre exemple, si notre zone source est un département de montagne (support source) et que nous étudions le tourisme de sport d'hiver, il est intéressant d'effectuer une première désagrégation du département aux communes (support auxiliaire) ayant une offre de station de sport d'hiver. Puis nous pouvons désagréger au sein des communes sur les hôtels (support cible). Le praticien aura le loisir de choisir à quel moment il fait intervenir les caractéristiques sociales des touristes.

Au demeurant, l'erreur est à mettre en relation avec la distance entre les attributs de la variable auxiliaire. Si nos deux centres d'activité sont à proximité les distances parcourues seront bonnes quelle que soit l'affectation. Si les deux centres ne sont pas à proximité l'erreur aura une importance potentielle plus grande. L'erreur est relative à l'espacement moyen des attributs de la variable auxiliaire. Enfin, il faut encore relativiser cette erreur en fonction des phénomènes de compensation entre les erreurs qui peuvent se produire. Pour illustrer ce phénomène de compensation nous présentons d'abord notre cas pratique de transfert de données de mobilité.

II Cas pratique: le transfert de données domicile-travail du recensement de population sur une grille régulière de 1 km de côté

2.1 Présentation des données sources, cibles, auxiliaires et témoins

2.1.1 Le fichier des mobilités professionnelles comme données sources

Les méthodes de transfert de données de déplacement que nous présentons consistent à transférer les données de mobilité du fichier détail anonymisé des mobilités professionnelles issu des enquêtes annuelles du recensement. Comme nous l'avons dit les caractéristiques intrinsèques des mobiles servent de données de contrôle. Dans ce fichier nous disposons d'un certain nombre de variable décrivant les actifs occupés de plus de 15 ans. Ces dernières apportent des éléments de connaissances sur les profils socio-économiques des actifs. Nous connaissons aussi quelques informations sur les ménages. En revanche, nous ne connaissons pas les couples (triplets, etc) d'actifs occupés composant les ménages.
Le support spatial de diffusion du fichier des « mobilités professionnelles » (MOBPRO) est le maillage communal. Nous connaissons donc les communes de résidence et d'emploi.

2.1.2 Une grille régulière de 1 km de côté comme support cible et une grille régulière de 200 m comme support auxiliaire

Nous souhaitons transférer l'information du fichier MOBPRO sur une grille régulière de 1 km de côté.
Ce carroyage est diffusé par l'INSEE ainsi qu'un carroyage de 200 m. L'INSEE diffuse ces carreaux comme support d'estimations de population, réalisée par ses services. Les carreaux ne contenant pas de population ne sont pas présents, le carroyage diffusé est discontinu. Les données permettant l'affectation de la population aux carreaux sont issues de la taxe d'habitation et des revenus fiscaux localisés.

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

Nous allons donc présenter plusieurs méthodes de transfert entre ces deux supports. Pour certaines d'entre-elles, nous utiliserons le carroyage de population de 200 m comme support auxiliaire. Les informations mobilisées en tant que variables de contrôles sont : l'effectif de ménage, la composition du ménage, le statut d'occupation et le type de logement. Toutes ces variables sont utilisées pour ventiler spatialement les lieux de résidences des ménages.

2.1.3 Les données auxiliaires relatives aux mobilités professionnelles

Nous souhaitons transférer spatialement une information décrivant des individus. Soit nous transférons d'un maillage à un autre cette information individuelle, soit nous reconstituons les ménages pour obtenir plus de degré de contrôle.

Nous allons effectuer une première méthode sans considérer la mise en ménage. Les lieux de résidence peuvent être déduis à l'aide de la couche IGN « Bâti indifférencié ».

D'autres méthodes présentées (qui seront décrites formellement par la suite) vont s'appuyer sur un profilage plus complet des actifs de MOBPRO. En effet, le fichier des mobilités professionnelles fait partie d'un ensemble de fichiers dits détails anonymisés. Parmi eux, le fichier INDIVIDUS décrit les individus de l'exploitation complémentaire des enquêtes annuelles du recensement. Ainsi, les observations de MOBPRO sont une partie des observations du fichier INDIVIDUS. De plus, le fichier INDIVIDUS est lui même une troncature du fichier *LOGEMENTS* décrivant les logements/ménages de l'exploitation principale. Selon les fichiers les variables, les modalités, la connaissance sur les couples (n-uplets) d'individus en ménages et les supports spatiaux diffèrent. Nous allons apparier ces trois fichiers pour disposer de l'ensemble de ces informations.

Afin de déterminer les lieux d'emplois des individus nous allons nous servir du répertoire SIRENE. Nous disposons en France du fichier SIRENE qui répertorie l'ensemble des établissements économiques avec la précision géographique à l'adresse postale.
-« *Conformément aux articles R.123-220 à R.123-234 du code du commerce, le répertoire Sirene enregistre l'état civil de toutes les entreprises et leurs établissements :*

- *quelle que soit leur forme juridique*
- *quel que soit leur secteur d'activité (industriels, commerçants, artisans, professions libérales, agriculteurs, collectivités territoriales, banques, assurances, associations...),*
- *situés en France métropolitaine (y compris la Corse), ainsi qu'en Guadeloupe, Martinique, La Réunion, Mayotte, Saint-Barthélémy, Saint-Martin et Saint-Pierre et Miquelon.*

L'inscription au répertoire SIRENE (Système Informatique pour le Répertoire des Entreprises et de leurs Établissements), suite aux démarches administratives correspondantes, est réalisée gratuitement par l'INSEE.» (INSEE, Connaître et comprendre le répertoire SIRENE).

Ce répertoire contient de nombreuses informations sur les établissements économiques dont l'adresse postale, les classes d'effectifs de salariés et le secteur d'activité en 732 postes. Nous allons pouvoir mettre en correspondance le secteur d'activité des actifs avec ceux du répertoire SIRENE.

2.1.4 L'enquête globale transport comme données témoins

La vérification de la qualité des méthodes de transfert s'appuient sur une comparaison avec une autre source de données. Nous utilisons l'enquête globale transport d'Île-de-France. Outre, être la seule enquête ménage-déplacement ayant une appellation différente, elle se démarque par la précision du maillage utilisé pour localiser les origines et destinations. Les plus anciennes diffusent les flux sur un carroyage de 300 m et la dernière version géolocalise les flux sur un carroyage de 100 m.
Cette précision nous permet d'utiliser cette enquête pour valider nos processus de transfert. Toutefois, cette enquête a un taux d'échantillonnage faible ce qui empêche de valider les flux origine-destination par comparaison de deux zones géographiques petites.

2.1.5 Terrain d'étude

Nous sélectionnons les zones de tirage de l'EGT couvrant la Seine-et-Marne. Nous gardons dans la phase d'analyse des distances les actifs résidant et travaillant dans le département. Ainsi, nous excluons les observations MOBPRO parcourant quotidiennement de très grandes distances dont on ignore s'il s'agit de valeurs aberrantes ou non, tout en espérant préserver des distances assez longues pour analyser les éventuelles structures socio-spatiales des déplacements.

2.2 Les déterminants théoriques du processus de choix des lieux de résidence

Le processus de transfert de support et particulièrement la désagrégation doit dans l'idéal retranscrire la structure géographique des déplacements domicile-travail résultante des processus de choix individuels, dont la combinaison est l'organisation socio-spatiale.
Le terme d'organisation désigne la manière dont sont structurés les attributs de l'espace. Il y a donc à la fois l'idée de connaissance de l'agencement, de la disposition spatiale des éléments à un temps donné. Et, l'idée d'un certains savoirs sur les processus, les dynamiques qui expliquent la structure spatiale à un moment passé, présent et pourquoi pas futur.
Nous employons le terme d'organisation socio-spatiale afin de nommer cette inconnue dont nous voulons retranscrire quelques propriétés. Comme nous étudions les déplacements domicile-travail il nous faut connaître la structure spatiale des lieux de résidences et celle des emplois. La connaissance des processus de choix du lieu de résidence et d'emploi à l'échelle des individus permettrait de retrouver le chaînage entre ces deux semis de points.
Nous posons le problème comme un triptyque: le système de production des lieux de résidences, le systèmes des activités économiques et le système de transport. L'objet de ce papier n'est pas de décrire ce triptyque. Le lecteur avisé connaît la grandeur de la complexité de ces systèmes et des liens les unissant. Rappelons simplement les forces d'attraction et de répulsion principales. Le phénomène urbain est le fruit de la nécessité de l'homme à vivre en société, la ville est le lieu de concentration des hommes et des activités (Bairoch P., 1985). Le niveau de performance des moyens de transports relativise la notion de concentration en affaiblissant la rugosité de l'espace. Le train puis l'avènement de l'automobile a considérablement bouleversé la morphologie de la ville. L'étalement urbain est le résultat de l'accroissement d'une dichotomie entre les logiques résidentielles et celle des emplois. Nous assistons à un allongement des distances domicile-travail. L'automobile augmentant le champ des possibles (ubiquité, instantanéité, immédiateté (Dupuy G.,2000)) les individus peuvent multiplier des desiderata dans leur processus de choix résidentiel, ou bien répondre à des contraintes liées à la pression foncière (Dupuy G.,1999). De plus, nous sommes en face de processus de choix individués puisque ces choix sont réalisés en fonction de la connaissance et des expériences individuelles. Aussi, il faut remettre le principe de rationalité des individus en cause. Nous savons que les choix sont remis en cause de manière ponctuels en fonction du cycle de vie (Bourgin C.,1978). L'organisation socio-spatiale des lieux de résidence doit aussi être étudiée à l'échelle des ménages. Il faut alors ajouter à la liste des déterminants de la mobilité professionnelles les effets de la mutualisation des contraintes des individus d'un même ménage.
Les déterminants de la mobilités sont donc nombreux et ne sont pas identiques selon les individus. La catégorie socio-professionnelle, le statut d'occupation, la position dans le ménage, la position dans le cycle de vie, l'environnement de proximité, les moyens de transport à disposition sont des facteurs influant peu ou prou sur le comportement de mobilités des individus. La structure globale résultante de la combinaison des choix individuels apparaît comme un objet théorique difficilement déterminable.

Certains de ces déterminants sont captés par le recensement de population car les écarts de distance des actifs sont suffisamment grand pour ne pas être lissé par la taille du maillage communale. Les professions intermédiaires, puis les ouvriers ensuite les employés parcourent de plus grandes distances pour aller

travailler (planche d'histogramme n°1). Il faut comprendre qu'ils traversent bien plus de limites communales que d'autres pour aller travailler.

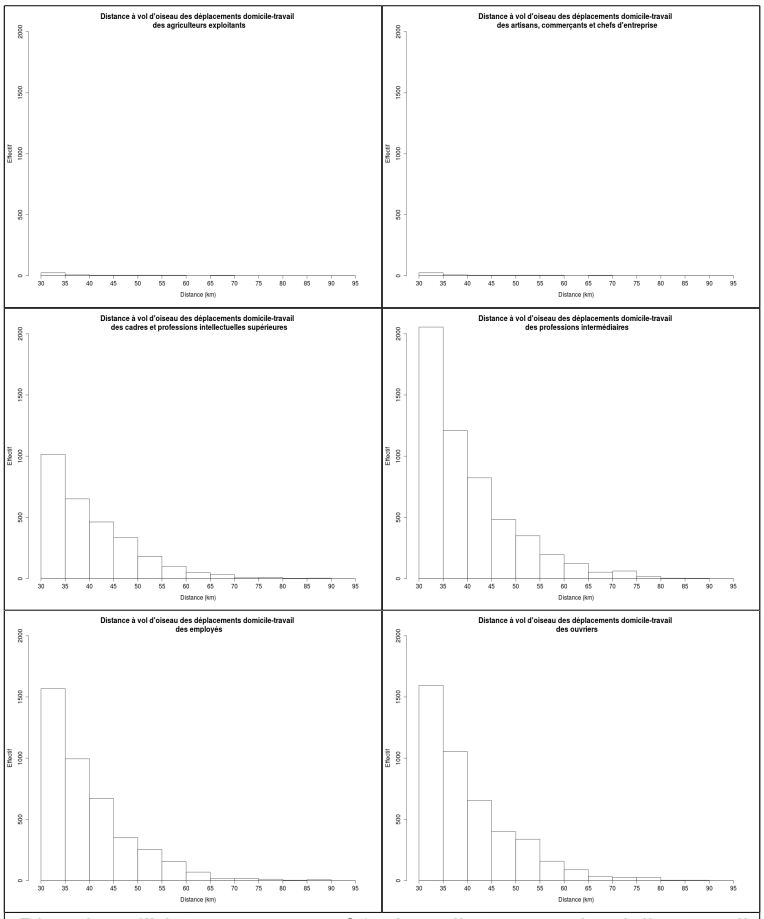


Planche d'histogrammes n°1 des distances domicile-travail de plus de 30 km selon la catégorie socio-professionnelle

En revanche, il ne faut pas extrapoler cette observation dans l'infra-communale, rien nous permet de dire que les cadres parcourent plus de distances que les agriculteurs lorsqu'ils habitent et travaillent dans la même commune. Le problème est que nous ignorons cette hypothétique structure à respecter à l'échelle infra-communal (Pivano C.,2014). Tant bien même elle existerait dans une commune elle prendrait une forme différente dans une autre. Nous pouvons toutefois approcher les structures résidentielles et celle des emplois à l'aide de données ponctuelles les décrivant. En revanche, nous ne connaissons pas le chaînage entre les deux. Nous touchons ici une des spécificités de ce travail de désagrégation de flux. Nous traitons d'une donnée multi-localisée, il convient de désagréger deux informations ponctuelles (origine et destination) et d'apparier ces deux points. C'est de cette appariement que dépend la structure spatiale reproduite par les méthodes de transfert.

Prenons l'exemple d'une commune avec un gros centre d'emploi comme un hôpital, nous savons qu'il y a tant de salariés de l'hôpital résidant dans la commune. Nous souhaitions affecter pour chaque salarié de l'hôpital un bâtiment de résidence. Nous pouvons premièrement supposer que les salariés habitent à proximité directe du centre hospitalier. Deuxièmement il est sûrement utile de distinguer les comportements selon les revenus des salariés et l'offre de logement. L'aire de rabattement résidentielle sera différente entre celle d'un hôpital d'une banlieue pauvre et celle d'une clinique privée de la côte d'azur en raison du voisinage géographique.

En conclusion, nous posons le problème de l'étude de l'organisation socio-spatiale des déplacements comme un système dynamique complexe sensible aux conditions «initiales». Nous sommes donc confrontés aux problèmes classiques d'un contexte chaotique. Une action microscopique peut avoir des conséquences sur la structure globale mais de fait de la somme des facteurs explicatifs à prendre en compte la récurrence d'un même phénomène conduit à des conséquences non similaires. La même cause conduit à des effets différents rendant le système non prédictible. Face à cela, nous comprenons que nous ne pouvons pas produire des méthodes de désagrégation génériques qui tentent de reproduire toutes les innombrables

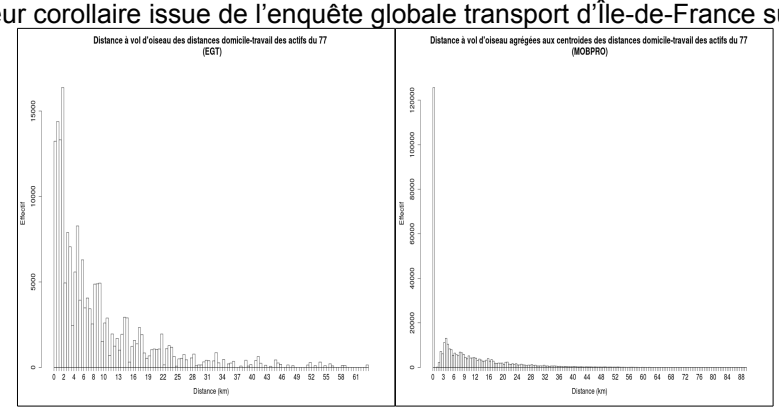
spécificités locales. L'idée est de comparer des chaînages de lieu de résidence et d'emploi selon différentes méthodes stochastiques, et de mettre à disposition des outils permettant de juger de la qualité des méthodes de désagrégation. Nous comprenons qu'il est très difficile voire impossible d'appréhender toute la réalité dans le processus de désagrégation. La désagrégation spatiale est un processus méthodologique ne faisant que produire une structure à partir des variables de contrôle fournies en entrée.

III L'hypothèse de l'équi-répartition comme structure spatiale

3.1 Conséquences de l'hypothèse de l'équi-répartition sur le processus de désagrégation

En désagrégeant à l'aide de l'hypothèse d'équi-répartition pour reconstruire la structure interne des mailles sources la marge d'erreur est bornée par une limite qui dépend de la taille des mailles.

L'étude empiriques des erreurs nous montre qu'il existe un intervalle d'erreur qui dépend de la superficie des communes *i.e la marge d'erreur de la désagrégation est bornée* (Terrier C., 2006). Puisque nous désagrégeons des flux inter-communaux la marge d'erreur de localisation des lieux de résidence dépend alors de la superficie des communes. Plus la commune est petite moins nous avons de possibilités pour localiser les ménages et par conséquent nous avons moins de chances de nous tromper. En ce qui concerne les flux, il faut doubler le raisonnement, comme nous connaissons le lieu de résidence et de travail à l'échelle communale la source d'erreur produite dépend de la superficie des deux communes. L'erreur se produit aux extrémités d'un déplacement. Par conséquent, l'erreur est relativement plus faible pour une navette de grande distance que pour les plus petites. Ces derniers ont proportionnellement à leur taille une plus grande distance parcourue au sein des communes d'arrivée et de départ donc au sein des zones d'erreurs potentielles. Le cas le plus spectaculaire est la disparition de la possibilité de mesurer les déplacements intra-zones. L'origine et la destination sont ramenées au même barycentre alors la distance devient nulle.
De fait, le processus de désagrégation préservera mieux les caractéristiques des mobilités des grandes distances. Une personne effectuant de longues distances le fera également à l'issue de la désagrégation puisque nous gardons la commune de départ et d'arrivée. Et, plus la distance sera grande plus l'erreur aux extrémités sera insignifiante. C'est pourquoi nous arrivons à observer des déterminants de mobilité à l'échelle communale. En revanche, en ce qui concerne les déplacements entre des communes limitrophes l'erreur relative commise peut être plus importante comme dans le cas de conurbation urbaine. Enfin l'erreur aura une importance maximale dans le traitement des déplacements internes aux mailles sources. Il nous faut observer ces variations. Nous construisons les histogrammes de distances des navettes du recensement et leur corollaire issue de l'enquête globale transport d'Île-de-France sur la Seine-et-Marne.



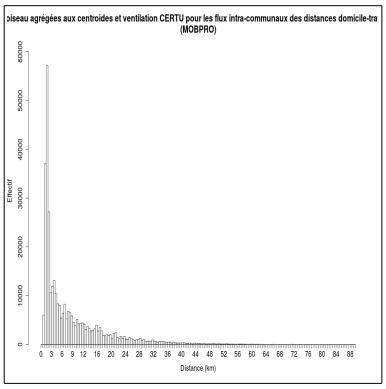
L'histogramme de droite est fabriqué à l'aide du recensement de population en agrégeant les origines/destinations aux centroïdes des communes: ce qui revient à avoir distribué les O/D de manière aléatoire dans la commune.

Nous observons deux choses: premièrement les grandes distances sont similaires (en faisant attention à l'échelle des ordonnées). L'hypothèse d'équi-répartition semble reproduire comme attendu des grandes distances cohérentes. De plus comme nous l'avons vu dans la partie présentant les déterminants socio-économiques il n'y a pas lieu de segmenter la population selon les profils socio-économiques. Nous retrouvons une structure des distances plausibles en raison d'un phénomène de compensation.

Deuxièmement, l'hypothèse d'équi-répartition sur-estime très fortement les distances de 0 à 500 m et sous-estime fortement les distances de 500 à 1500 m qui ont totalement disparu. La disparition de ces derniers s'expliquent de part l'agrégation aux mêmes centroïdes de déplacements intra-communaux.

Pour traiter le cas des déplacements intra-zones nous pouvons effectuer une imputation a posteriori proposée par le Certu (Gascon, 2009) qui préconise de construire les distances intra-zones en fonction de la superficie des zones, soit :

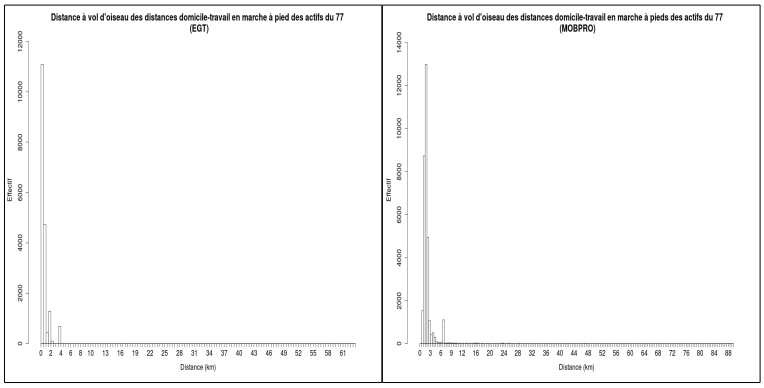
$$Distance = \frac{1}{2}\sqrt{Surface\,de\,la\,zone}$$



En effectuant ce traitement, les déplacements de 0 à 500 m ont désormais disparu. Il existe pourtant dans la réalité des déplacements de cette classe. Plus généralement, nous remarquons, que les petites distances sont de manière générale plus grandes que celle de l'EGT. Il y a très peu de distances inférieures à 1 km (deux premières barres) alors que les déplacements inférieures à 1 km sont très nombreux dans l'EGT. La méthode surfacique du Certu a tendance à sur-estimer les déplacements courts en supposant qu'ils soient égaux à la moitié du rayon de la zone, en considérant celle-ci comme un carré.

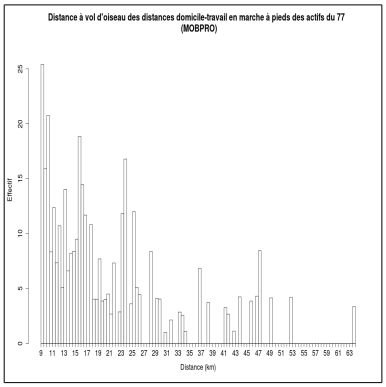
Cette analyse des petites et grandes distances démontre que la variation des extrémités des déplacements se compensent entre elles de manière différente selon un gradient de distance. Si pour les grandes distances nous savons que les erreurs se compensent en retranscrivant des indicateurs plausibles de la réalité, les petites distances sont bien plus problématiques.

Pour s'affranchir des contraintes de compensation des grandes distances nous devons isoler les petites distances pour les examiner de plus près dans l'étape de validation. Afin de définir ce qu'est une petite distance nous pouvons isoler les déplacements en marche à pied. Observons les histogrammes de la marche à pied.



Nous retrouvons le décalage des courtes distances exposé plus haut. Nous postulons alors qu'une méthode de désagrégation pertinente aura une distribution des distances en marche à pied plus proche de celle de l'EGT et donc éloignée de la méthode du CERTU.

Toutefois, les enquêtes du recensement sont entachées d'erreur pour la marche à pied, tout comme l'EGT. Si l'on s'intéresse aux déplacements déclarés effectués en marche à pied nous remarquons que certains parcouraient plusieurs dizaines de kilomètres à pied pour aller travailler. Dans l'histogramme suivant nous avons enlevé les déplacement de moins de 9 km.



D'autres indicateurs pour analyser les petites distances peuvent être utilisé. Nous proposons les histogrammes suivants:
- histogrammes des déplacements intra-zones
- histogrammes des déplacements de communes limitrophes

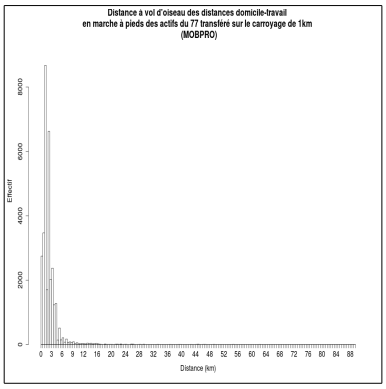
Aussi comme nous disposons du mode de transport nous pouvons éventuellement mesurer les distances d'accès aux stations d'entrée et de sortie des utilisateurs des transports publiques afin de vérifier si elles sont cohérentes. Dans le fichier de données une observation ayant comme caractéristique d'utiliser les transports en commun alors qu'elle est située par le processus de désagrégation dans un hameau isolé du centre sera suspecte. Seulement cette incohérence peut être aussi due aux techniques d'enquêtes. L'individu que décrit l'observation statistique peut tout aussi bien être conduit par son conjoint à une station de transport en commun à plusieurs kilomètres. L'enquête étant construite pour retenir un seul mode de transport, la multi-modalité du déplacement est lissée sur le mode de transport le plus lourd. Nous sommes donc jamais totalement sûrs de savoir si l' l'on se trompe ou pas.
Il nous faut utiliser des méthodes d'affectation de trafic pour construire ces distances d'accès. Nous pourrons alors fabriquer les histogrammes :
- des distances d'accès aux stations d'entrée de transport en commun pour évaluer la qualité de la répartition résidentielle.
- des distances d'accès aux stations de sortie de transport en commun pour évaluer la qualité de la répartition des emplois. Nous regarderons cela dans la thèse de C.Pivano en cours de réalisation.

Nous ajoutons dans les pistes méthodologiques de validation que le processus de désagrégation peut être effectué avec une succession de désagrégations sur des supports de plus en plus fin. Nous pouvons alors observer la répartition spatiale des origines et destinations en fonction de l'utilisation ou non de l'information de ces supports auxiliaires.
Mais l'utilisation d'une succession de support a également son influence sur les résultats. Lorsque nous utilisons un support intermédiaire dans le processus de désagrégation nous utilisons un maillage auxiliaire, et nous sommes sujets au MAUP. Si nous saisissons le biais du MAUP alors il peut être utilisé comme une certaine forme de contrôle. L'emploi des deux termes est donc possible dans tout les cas puisque dès que nous mobilisons l'emploi d'un maillage il structure les résultats. Il convient bien plus de savoir comment l'emploi de supports auxiliaires structure (contrôle) les résultats que ce soit volontairement ou involontairement, comme dans le cas d'un processus d'agrégation.

3.2 Lissage par ré-agrégation

Une fois le processus de désagrégation effectué nous allons ré-agréger l'information sur le carroyage de 1 km. La question est de savoir si l'utilisation de méthodes sophistiquées n'est pas superflue lorsque nous ré-agrégeons le résultat de désagrégation à une échelle plus macroscopique. Autrement dit est-ce que le lissage de la ré-agréagtion produit une perte d'information trop importante ? Ou encore est-ce que des méthodes moins sophistiquées donnent des résultats similaires par le biais de phénomène de compensation ?

Nous effectuons la ventilation de MOBPRO sur les carreaux de 1 km. Nous effectuons un transfert des communes aux carreaux par une pondération surfacique sans passer par un support auxiliaire. Les probabilités d'affectation dépendent de la superficie d'intersection de la maille cible dans la maille source. Ensuite les distances sont calculés aux centroïdes des carreaux. Nous observons la figure suivante pour les distances en marche à pied.



Nous observons que les distances sont plus grandes que celles mesurées à l'aide de l'EGT. Toutefois, des distances de 0 à 500 m réapparaissent. Elles avaient totalement disparu dans l'approche du CERTU. Les déplacements en marche à pied sont courts et donc bien souvent intra-communaux. En fonction de la taille de la commune, les origines et destinations ont pu être réparties dans des carreaux différents. Certains déplacements ont eu comme affectation le même carreau. La valeur de leur distance est nulle. Le premier intervalle de 0 à 500 m ne contient formellement que des déplacements nuls. Nous ignorons en fait leur longueur par contre la méthode arrive à nous donner une estimation de leur volume.
Nous observons bien une déformation de l'information dans le processus de ré-agrégation, il apparaît même qu'elle a tendance à produire des distances nulles. Dans ce cas, elle améliore le procédé de transfert puisque nous avions fortement sous-estimé ces déplacements; il n'y en avait même plus aucun.

IV Présentation des méthodes de désagrégation contrôlées

Nous présentons maintenant trois méthodes de désagrégation, chacune d'elles ayant un niveau d'exigence de représentativité de la structure socio-spatiale différent. Premièrement nous proposons une méthode ayant pour objectif de reproduire la distribution des distances des individus sans considérer les profils socio-économiques des mobiles, et la mutualisation des contraintes des ménages (méthode 1). Ensuite, nous proposons une méthode considérant la mise en ménage. Nous verrons que la méthode utilisée nous permet de descendre de la commune à l'IRIS en considérant les profils socio-économiques (méthode 2). En revanche, le passage des IRIS aux carreaux de 200m se fait dans cette méthode sans considérer le profil des ménages. La dernière méthode propose une solution empirique pour préserver cette information à l'échelle des carreaux de 200m (méthode 3). Le tableau suivant résume les données de contrôle utilisées. Pour chacune des méthodes on fera l'analyse des indicateurs retenus plus hauts avec et sans le processus de ré-agrégation.

Commune (Données sources)		IRIS	Carreaux 200m	Coordonnées X/Y
Méthode 1				
Origine	Fichier des mobilités professionnelles	→		Bâtiment (IGN)
Destination	Fichier des mobilités professionnelles	→		SIRENE(NA5)
Méthode 2				
Origine	Fichier des mobilités professionnelles	→*Fichier détails Individus CANTVILLE*	→ Carreaux 200m simple	
Destination	Fichier des mobilités professionnelles	→		SIRENE (NA17)
Méthode 3				
Origine	Fichier des mobilités professionnelles	→ *Fichier détails Individus CANTVILLE* et Logements	→ Carreaux 200m multi-critère	
Destination	Fichier des mobilités professionnelles	→		SIRENE (NA17)

4.1 Méthode 1 : couche bâti de l'IGN pour le lieu de résidence et utilisation de SIRENE pour les lieux d'emploi.

Cette première méthode s'appuie sur la couche d'information géographique « bâtiment indifférencié » de l'IGN pour construire les lieux de résidence, à l'instar du projet MobiSim (Antoni, 2010). Les lieux de résidence et d'emploi sont construits à l'aide d'un tirage de Monte-Carlo pondéré par le volume des bâtiments. Pour chaque actif occupé du fichier MOBPRO nous tirons au sort un bâtiment de la couche d'information géographique. Plus le volume du bâtiment est grand plus il a de chances d'être choisi.

Les lieux d'emploi sont approchés à l'aide du répertoire SIRENE. Nous utilisons l'information du secteur d'activité (NA5). Pour chaque actif du fichier MOBPRO nous sélectionnons les établissements économiques du même secteur. Puis nous procédons à un tirage de Monte-Carlo pondéré par les classes d'effectif que nous renseigne le répertoire des activités. Nous soulignons que cette méthode de désagrégation passe de l'échelle communale aux coordonnées des bâtiments et des établissements économiques sans support auxiliaire intermédiaire.

4.2 Méthode 2: appariement et ventilation sur le carroyage des lieux de résidence et utilisation de SIRENE et pour les lieux d'emplois.

La méthodologie de construction des lieux de résidence passe par plusieurs supports auxiliaires. Nous procédons en deux temps, premièrement nous passons l'information du support communal aux IRIS. Nous allons ensuite désagréger l'information des IRIS aux carreaux de 200 m. Nous signalerons les différentes conséquences de ce processus sur les communes découpées ou non en IRIS.

L'appariement des fichiers détails comme méthode de désagrégation des communes aux IRIS :

Le recensement de population est diffusé sous plusieurs fichiers. Les fichiers détails sont notamment constitués du fichier INDIVIDUS renseignant sur le canton ou la «ville» de résidence. Il offre des informations sur l'ensemble des individus composant l'exploitation complémentaire. L'avantage de ce fichier est qu'il permet de reconstituer statistiquement les individus d'un même ménage. La précision géographique de la localisation résidentielle diffère selon la population communale: les individus habitant les communes de moins de 10 000 habitants sont géolocalisés au canton de résidence pour des raisons de secret statistique,

les individus résidant dans une commune de plus de 10 000 habitants sont renseignés à l'aide de l'IRIS de résidence.
Le fichier MOBPRO décrit les mobilités professionnelles des actifs occupés de plus de 15 ans décrits dans le fichier INDIVIDUS. Il s'agit donc d'une sélection des actifs occupés du fichier INDIVIDUS dont certaines variables ont été tronquées, modifiées, supprimées et ajoutées. Ici, nous connaissons la commune de résidence et d'emploi de toutes les observations. En revanche, nous ne disposons plus de la variable permettant d'identifier les personnes d'un même ménage et de la variable localisant les actifs à l'échelle des IRIS.
L'objectif de l'appariement est donc d'identifier pour chaque actif occupé sa correspondance dans le fichier INDIVIDUS, ainsi nous pourrons disposer des avantages des deux fichiers: une meilleure précision géographique des lieux de résidence, la connaissance de la commune d'emploi, et la connaissance des autres individus du ménage.
La première étape est de sélectionner les actifs occupés de plus de 15 ans dans le fichier INDIVIDUS et de vérifier que l'effectif correspond à celui du fichier Mobpro.
La suite de l'appariement de ces deux fichiers se fait en construisant une clef d'appariement à l'aide de l'ensemble des variables communes aux deux fichiers. Il faut faire attention aux modalités qui diffèrent. Idéalement, nous souhaitons avoir une clef unique pour chaque observation ce qui permettrait de déduire de manière sûre les correspondance de ces deux fichiers. Malheureusement les 23 variables de la clef d'appariement ne suffisent pas à distinguer systématiquement chaque actifs. Nous sommes donc en présence d'une incertitude quant à l'appariement à effectuer. Dans certains cas nous pouvons la résoudre, dans d'autres nous n'avons pas assez d'éléments pour le faire.

MobPro				INDIVIDUS		
Numéro d'identifiant de l'observation	Commune	Clé de couplage		Clé de couplage	Canton	ID_MEN
1	77325	M 11 7701 ZZZZZ 2 4 2 13 16 6 GU 1 2 1 1 22 1 4 1 41 10 30 4	?	M 11 7701 ZZZZZ 2 4 2 13 16 6 GU 1 2 1 1 22 1 4 1 41 10 30 4	7701	77011096
2	77523	M 11 7701 ZZZZZ 2 4 2 13 16 6 GU 1 2 1 1 22 1 4 1 41 10 30 4		M 11 7701 ZZZZZ 2 4 2 13 16 6 GU 1 2 1 1 22 1 4 1 41 10 30 4	7701	77011137

Les variables qui ne sont pas dans la clef de couplage peuvent être librement interverties entre les individus doublons. La lecture doit être effectuée en s'interrogeant sur les conséquences des appariements des doublons. Les variables décrivant les trajectoires résidentielles, les superficies des logements, les compositions des familles au sein des ménages, les âges détaillés en année et les lieux de naissances ne sont pas considérées dans la clef de couplage. Ainsi en cas de doublons et d'interversion des individus au sein des ménages ce sont ces caractéristiques qui ne pourront être considéré dans les analyses des résultats.

Plus gênant pour notre sujet de recherche, en cas de doublon il est possible d'échanger les individus dans des ménages différents. La mutualisation des contraintes de localisation n'est plus respectée. Cette interversion se fait entre deux ménages d'un même IRIS pour les communes de plus de 10 000 habitants. Pour les communes non découpés en IRIS, il est possible d'intervertir entre deux communes d'un même canton.

Premièrement, nous supposons que l'interversion de deux individus d'un ménage de une personne n'a strictement aucune importance. Deuxièmement, il est possible de retrouver les bons appariements à l'aide de la présence d'un ou plusieurs actifs. L'algorithme de couplage compare alors les clefs d'appariement des autres actifs occupés des ménages. Enfin, s'il y a seulement un actif occupé dans le ménage (de plusieurs personnes) nous ne disposons plus de sources d'information suffisantes pour solutionner le problème.

Toutefois nous pouvons supposer que l'interversion n'a pas de conséquence sur la somme des distances parcoures par les individus du ménage.

Le tableau suivant résume les erreurs théoriques de la méthode d'appariement.

Type de commune	Type de cas	Type d'erreur possible
Communes découpées en IRIS	Ménages de plusieurs actifs occupés	Mauvais ménage
	Ménages de plusieurs individus avec un seul actif occupé	Mauvais ménage
	Ménages occupés de 1 personne	Mauvais ménage
	Hors-ménages	Mauvais individu
	Ménages non occupés	Pas dans MOBPRO
Communes non découpées en IRIS	Ménages de plusieurs actifs occupés	Mauvais ménage/commune
	Ménages de plusieurs individus avec un seul actif occupé	Mauvais ménage/commune
	Ménages occupés de 1 personne	Mauvais ménage/commune
	Hors-ménages	Mauvais individu/commune
	Ménages non occupés	Pas dans MOBPRO

Connaissant les erreurs potentielles nous appliquons l'appariement des fichiers du recensement de 2008 du département de Seine-et-Marne en nous attachant à observer les cas problématiques. L'examen s'effectue en utilisant le fichier INDIVIDUS puisque c'est lui qui offre la meilleure connaissance sur les ménages (mais il est plus difficile de compter les ménages bi-actifs sans la variable INPOM).

Type de commune	Type de cas	Type d'erreur possible	Effectif d'individus doublons (sans poids)	Effectif d'individus total(sans poids)
Communes découpées en IRIS	Ménages de plusieurs actifs occupés	Mauvais ménage	4039	86228
	Ménages de plusieurs individus avec un seul actif occupé	Mauvais ménage	1000	21768
	Ménages occupés de 1 personne	Mauvais ménage	2228	16609
	Hors-ménages	Mauvais individu	443	992
	Ménages non occupés	Pas dans MobPro	-	-
Communes non découpées en IRIS	Ménages de plusieurs actifs occupés	Mauvais ménage/commune	1673	47099
	Ménages de plusieurs individus avec un actif	Mauvais ménage/commune	287	8972
	Ménages occupés de 1 personne	Mauvais ménage/commune	544	4803
	Hors-ménages	Mauvais individu/commune	106	414
	Ménages non occupés	Pas dans MobPro	-	-
TOTAUX			10 320 actifs occupés	186 885 actifs occupés

Ce tableau indique que le nombre de doublons est assez faible, ils correspondent à 5,6 % (10320/186885) des observations des individus actifs occupés. Rappelons que, même si nous traitons de l'interversion pour les grandes communes, l'interversion des actifs au sein de ménage d'un même IRIS n'est pas gênante.
Le problème de mauvais appariement pose problème uniquement pour les communes non découpées en IRIS. Les choix doivent être effectués entre plusieurs communes d'un même canton.
Les actifs problématiques sont donc ceux résidant dans les communes de moins de 10 000 habitants et bi-actifs, soit 0,9% (1673/186885) des individus actifs occupés de la Seine-et-Marne en 2008.

Pour ces cas de doublons, l'algorithme de couplage vérifie que les communes de résidence affectée aux individus correspondent à celle des autres individus actifs occupés du ménage. De cette manière nous réussissons à résoudre ce problème pour ces ménages bi-actifs représentant 0,9% de l'effectif total. Toutefois, l'algorithme d'appariement ne peut traiter le cas si tous les individus actifs occupés d'un même ménage ont des clefs d'appariement doublons. Ce cas est très rare, nous ne l'avons pas rencontré dans les fichiers de Seine-et-Marne.

A l'issue de ce travail d'appariement nous disposons de l'IRIS de résidence pour les actifs de MOBPRO pour ceux habitants dans les communes de plus de 10 000 habitants, et nous connaissons désormais la composition des ménages.

Des Communes/IRIS aux carreaux de 200 m:

Il nous reste à désagréger l'information des ménages localisés aux communes/IRIS aux carreaux de 200 m. Nous utilisons une méthode s'appuyant sur les effectifs de ménages. En effet, comme nous avons re-constitué cette information à l'étape précédente nous souhaitons la préserver. En revanche, nous utilisons une méthode ne considérant pas le profil socio-économique des ménages.

Nous ventilons tous les ménages du fichier INDIVIDUS dans le carroyage et pas seulement les actifs de MOBPRO.
Nous construisons des probabilités d'affectation au carroyage de 200 m en fonction de l'effectif de ménage des carreaux au sein de la commune. Le nombre de ménage à placer dans un carreau est donné par la formule suivante:

$$n^i = N \times P(x \in n^i)$$

où :

n^i est le nombre total de ménage du carreau i

N est le nombre total de ménage de la commune/Iris

$P(x \in n^i)$ est la probabilité que le ménage x appartienne au carreau i

Cette probabilité est construite à l'aide de la superficie des mailles, soit:

$$P(x \in n^i) = \frac{C^i}{\sum_1^i C}$$

avec :

C^i L'effectif du carreau i intersectant avec la commune/Iris

$\sum_1^i C$ La somme des effectifs des carreaux intersectant la commune/Iris

Le problème étant de déterminer un effectif de ménages lorsqu'un carreau intersecte plusieurs mailles. En effet dans le cas des mailles non contenues entièrement dans les IRIS les effectifs déclarés dans la maille fine dénombrent des ménages pouvant appartenir à plusieurs communes/IRIS. Il faut alors estimer les effectifs appartenant aux IRIS I*m* soit $\sum C^i \in I^m$. Nous réécrivons donc la formulation de la manière suivante :

$$n^i = N^i \times \frac{C^i \in I^m}{\sum_1 C \in I^m}$$

où I^m est l'IRIS de la commune *m*.

Il est alors possible d'estimer les effectifs des variables de contrôle $\sum C^i \in I^m$ en désagrégeant au préalable les effectifs des variables de contrôle en fonction de leur superficie intersectant avec le maillage source.

$$C^i \in I^m = C^i \times \frac{Surface\, C^i \cap I^m}{Surface\, C^i}$$

Choix du lieux d'emplois:

Enfin, nous allons construire les lieux d'emplois. Nous disposons du secteur d'activité en 5 postes dans le fichier détail des mobilités professionnelles. Avec l'appariement nous connaissons désormais le secteur d'activité en 17 postes. Nous nous servons de cette information pour choisir un établissement de travail de même secteur d'activité pondéré par les classes d'effectif diffusées dans le répertoire.

4.3 Méthode 3 : appariement, ventilation multi-critère sur carroyage pour les lieux de résidence et utilisation de SIRENE pour les lieux d'emplois.

Nous procédons également à une désagrégation à l'aide d'une suite de supports auxiliaires. Nous procédons cette fois-ci à l'appariement des trois fichiers détails LOGEMENTS/INDIVIDUS/MOBPRO. Nous procédons comme précédemment pour apparier les fichiers INDIVIDUS et MOBPRO. Nous procédons également à l'appariement des fichiers INDIVIDUS et LOGEMENTS. Nous ne présentons pas ici la méthode, faute de place. Brièvement, nous utilisons également des clefs d'appariement des variables communes aux deux fichiers. Nous disposons alors pour ces trois fichiers d'un identifiant unique des ménages/logements.

Puis, nous ventilons les lieux de résidence des IRIS sur le carroyage de 200 m en essayant de préserver la structure multidimensionnelle de l'information. Cette fois-ci nous distribuons les observations du fichier

LOGEMENT en préservant certaines caractéristiques des profils socio-économiques des occupants des ménages/logements.

Pour approcher la distribution socio-spatiale des ménages internes aux IRIS ou aux communes de moins de 10 000 habitants le carroyage de 200 m nous informe sur l'effectif de classe de composition de ménages(x), de statut d'occupation (y), et de type de logements (*z*) pour chaque carreau.

Le problème est que nous ne disposons pas des fréquences jointes des modalités de ce triptyque, alors que nous les avons dans les observations à désagréger et que c'est ce croisement qui, nous le supposons, permettra de retrouver la structure spatiale des trois variables. Les modalités sont:
- Composition des ménages : ménages de 1 personnes, de 2 à 4 personnes , de plus de 5 personnes.
- Statut d'occupation : Propriétaires, Locataires.
-Type de logements : Maison individuelle , Logement collectif.

Nous connaissons les fréquences jointes des variables à désagréger. En revanche nous ne connaissons pas celle des variables de contrôle. Le processus de désagrégation multi-critère se heurte aux problèmes de la prise en compte de l'interdépendance des variables.

$$n^i_{xyz} = N^i_{xyz} \times \frac{C^i_{x \cap y \cap z}}{\sum_1 C_{x \cap y \cap z}}$$

Une première stratégie consiste à reconstruire les fréquences jointes des événements sous l'hypothèse d'indépendance:

$$P(X \cap Y) = P(X) \times P(Y)$$

Soit :

$$\frac{C^i_{x_1 \cap y_1 \cap z_1}}{\sum_1^i C^i} = \left(\frac{C^i_{x_1}}{\sum_1^i C^i}\right) \times \left(\frac{C^i_{y_i}}{\sum_1^i C^i}\right) \times \left(\frac{C^i_{z_i}}{\sum_1^i C^i}\right)$$

Dans ce cas, l'avantage est de distribuer totalement les effectifs du recensement dans les carreaux. Cependant, cette estimation des intersections se fait sous l'hypothèse d'indépendance alors que nous souhaitons respecter les lois de probabilité jointes. De plus, nous signalons que l'hypothèse d'uniformité revient à effectuer un tirage de Monte-Carlo en prenant comme variable de contrôle la probabilité mesuré à l'aide d'un effectif total de population.

Si nous souhaitons préserver les lois de probabilités jointes données par les données à désagréger nous proposons de mobiliser l'information contenue dans la variable à désagréger.

$$\frac{C^i_{x \cap y \cap z}}{C^i} = \frac{\sum (x^i_x)}{\sum (x_x)} \times \frac{\sum (x^i_y)}{\sum (x_y)} \times \frac{\sum (x^i_z)}{\sum (x_z)}$$

avec :

x_i Un ménage du carreau i

L'application des lois de probabilité jointes des données à désagréger aux variables de contrôles a pour conséquence de transmettre la structure sociale de la maille source à la maille cible sous contrainte de respect des effectifs. Pour chaque carreau on essaie de reproduire la même structure de profil. S'il y a 10% de ménage de une personne locataire dans le collectif dans l'IRIS on estime qu'il y aura également 10% de ce profil dans chacun de ces carreaux.

Toutefois la méthode d'estimation ne peut pas être effectuée telle quel dans le cas d'un transfert de mailles non-alignées. En effet, dans le cas des mailles non contenues entièrement dans les IRIS les effectifs déclarés dans la maille fine dénombrent des ménages pouvant appartenir à plusieurs IRIS. Il faut alors estimer les $\sum C^i_{x \cap y \cap z}$ appartenant aux IRIS I*m* soit $\sum C^i_{x \cap y \cap z} \in I^m$. Nous réécrivons donc la formulation de la manière suivante:

$$n^i_{xyz} = N^i_{xyz} \times \frac{C^i_{x \cap y \cap z} \in I^m}{\sum_1 C_{x \cap y \cap z} \in I^m}$$

Il est alors possible d'estimer les effectifs des variables de contrôle $\sum C^i \in I^m$ en désagrégeant au préalable les effectifs par modalités des variables de contrôle en fonction de leur superficie intersectant le maillage source.

$$C^i_x \in I^m = C^i_x \times \frac{Surface\, C^i \cap I^m}{Surface\, C^i}$$

$$C^i_y \in I^m = C^i_y \times \frac{Surface\, C^i \cap I^m}{Surface\, C^i}$$

$$C^i_z \in I^m = C^i_z \times \frac{Surface\, C^i \cap I^m}{Surface\, C^i}$$

Cette estimation est une désagrégation surfacique et par construction répartit les profils en considérant la répartition de la population de manière uniforme au sein des carreaux. Il est tout a fait possible d'opérer avec des données de contrôle non-surfaciques pour préserver une sous structure spatiale. Et si nous voulons reconstruire une sous-structure spatiale avec les lois de probabilité jointes nous répétons l'ensemble du processus avec de nouvelles variables de contrôle qui contrôle les variables de contrôle du problème initial. Dans notre cas nous pouvons estimer les effectifs de nos variables de contrôle *x y z* appartenant à l'IRIS Im à l'aide de la couche bâti de l' IGN. Toutefois cette information auxiliaire n'offre pas d'éléments pouvant distinguer les comportements selon le profil des ménages.

Autre problème, dans le cas du recensement nous ne disposons pas d'individus à proprement parler mais d'observations statistiques avec des poids de sondage indiquant l'effectif de personnes que ces observations représentent. Le problème est donc le suivant, nous cherchons à distribuer les ménages dans les carreaux de manière à approcher l'égalité suivante:

$$n^i_{xyz} = N^i_{xyz} \times \frac{C^i_{x \cap y \cap z}}{\sum_1 C_{x \cap y \cap z}} = \sum (x^i_{xyz} \times w^i_{xyz}) \in C = \sum_1^i (x_{xyz} \times w_{xyz}) \in I_m \times \frac{C^i_{x \cap y \cap z} \in I^m}{\sum_1^i C_{x \cap y \cap z} \in I^m}$$

avec :

w_i Le poids de sondage de l'observation x_i

L'ajout des poids de sondage est problématique pour le respect de la contrainte d'égalité $\sum x^i \times w^i \in I^m = \sum c \in I^m$ car nous ne connaissons pas les sommes $\sum c \in I_m$ qu'il nous faut construire avant d'estimer les probabilités jointes.

Pour approcher cette distribution nous proposons de formaliser ce problème de ventilation comme un problème de satisfaction sous contraintes à l'image du problème du sas à dos (Kellerer, 2004) . Toutefois dans notre problème nous ne connaissons pas de manière sûre les marges de certains carreaux. Nous ne connaissons pas le volume de certains sac à dos. Nous cherchons donc les ménages appartenant aux carreaux ($x^i \in C^i$) avec les deux contraintes:

- $x^i_{xyz} \in C^i = C^i_{xyz}$

  L'effectif des modalités de chaque variable du recensement doit correspondre à ceux des carreaux
- $\sum x^i \times w^i \in I^m = \sum C^i \in I^m$

  Les effectifs de ménages distribués dans le carreau doivent lorsque on les ré-agrège à la même échelle doivent être identiques. Ce qui signifie que nous avons bien distribué les ménages dans le cas de maille cible superposant plusieurs IRIS (où nous ne connaissons pas les effectifs appartenant à la maille source).

Nous allons construire une distribution des x_i approchant au mieux ces contraintes. Les limites de cette approximation sont:

- les deux jeux de données ne sont pas construits avec les mêmes sources et donc les effectifs ne sont pas identiques pour une même zone.
- Nous ne connaissons par les lois de probabilité jointes des variables dans le carroyage.

- Nous ne connaissons pas l'effectif de ménage du carroyage pour les IRIS à cause du problème de non-alignement des maillages. Nous ne connaissons pas les marges des carreaux qui n'intersectent pas totalement avec un seul et unique IRIS.

Nous proposons une solution approximative pour approcher la solution théorique. Notre résolution de ce problème de satisfaction sous contraintes passe par l'algorithme de type glouton suivant.

Pour chaque IRIS nous sélectionnons les logements de l'exploitation principale c'est-à-dire les observations du fichier détail anonymisé LOGEMENT. Pour chacun d'entre eux nous sélectionnons un carreau avec la propriété suivante.

$$\sum x_x^i \times w^i < c_x^i \text{ et } \sum x_y^i \times w^i < c_y^i \text{ et } \sum x_z^i \times w^i < c_z^i$$

Nous sélectionnons un carreau qui n'est pas totalement rempli par des ménages de même modalités. Si nous trouvons un carreau avec cette propriété nous plaçons le logement dans celui-ci et nous ajoutons son poids de sondage dans le tableau comptabilisant les effectifs placés. Si nous sommes en présence de plusieurs carreaux possibles nous choisissons celui avec la plus grande surface intersectant l'IRIS en cours de traitement.

A l'issu de cet algorithme nous avons distribué les logements au sein du carroyage. Avec l'appariement LOGEMENT/INDIVIDUS nous pouvons retrouver les ménages de l'exploitation complémentaire. Ayant effectué l'appariement INDIVIDU/MOBPRO nous retrouvons les actifs occupés. Nous précisons que nous utilisons l'exploitation principale car il y a plus d'observations et les poids de sondage sont donc plus faibles, ce qui augmente l'espace des possibilités et donne de meilleure résultats que si nous procédons à la ventilation multi-critère des observations du fichier INDIVIDUS.

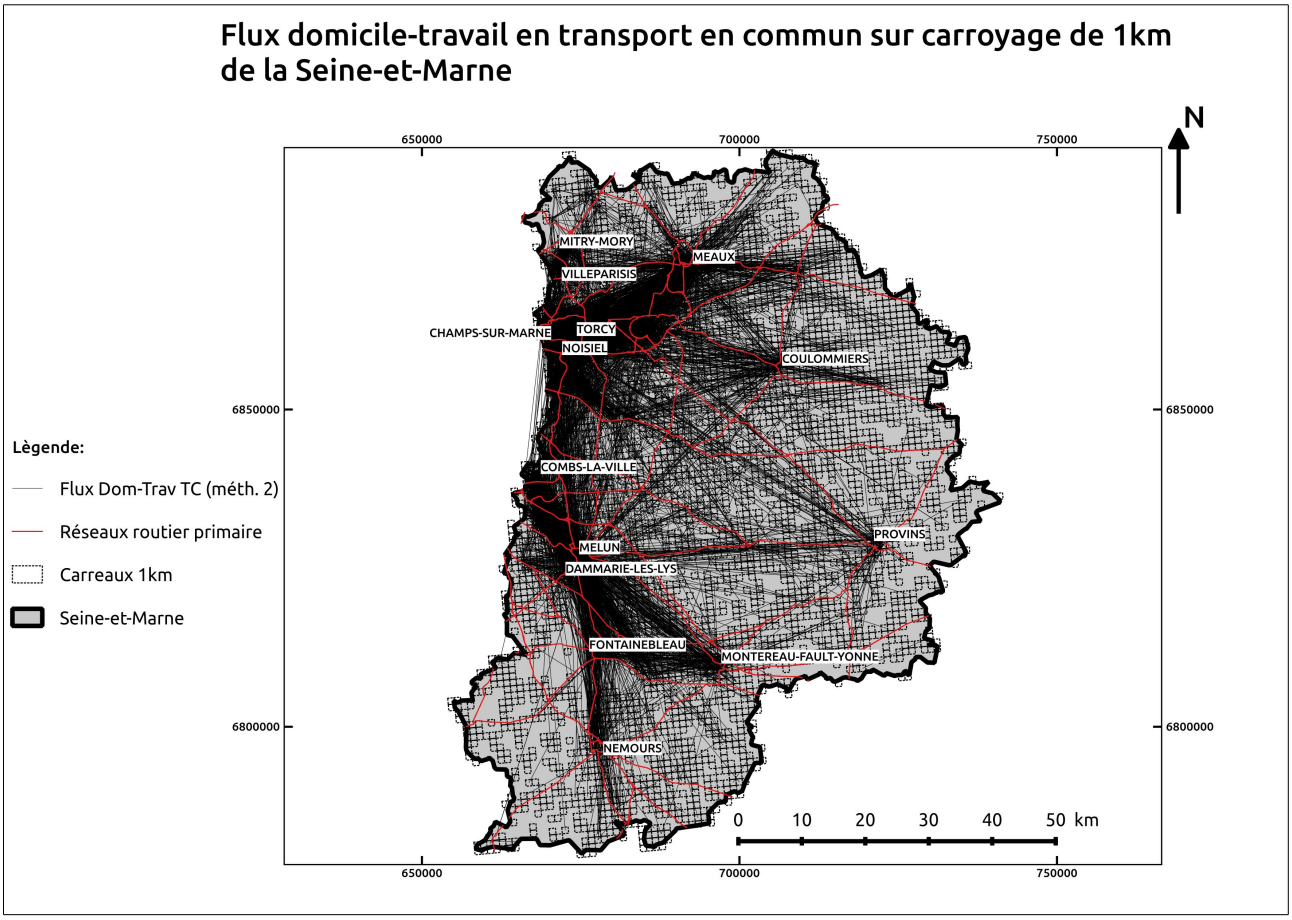
4.4 Commentaires sur les méthodes 2 et 3

Nous remarquons qu'aucune des méthodes proposées ne mobilise l'information sur le mode de transport utilisé, alors qu'elle peut être utile (partie 1.2.1). Nous avons laissé les contraintes liées aux systèmes transport en dehors du champ. C'est donc notre degré de liberté et nous observons son évolution dans l'étape de validation.
Nos méthodes s'appuient sur l'interdépendance entre lieu de résidence, lieu de travail et profil socio-économique pour reconstituer des distributions probables de distances. Nous ne tenons pas compte des contraintes sur les modes de transports pour laisser un degré de liberté et observer comment cela se comporte. Par exemple, nous savons que la grande partie des déplacements en marche à pied sont effectués en centre-ville. La méthode va distribuer des individus se déplaçant à pied dans le centre-ville par l'intermédiaire du lien entre la composition du ménage, le statut d'occupation, et le type de logements et la localisation en centre-ville. Toutefois nous ne quantifions jamais l'intensité de relation de ces variables. Nous examinons la cohérence de la désagrégation à l'aide d'indicateurs sur les distances.
Nous aurions pu utiliser des méthodes s'appuyant sur des distributions de distances que nous essayons de reproduire (Chacker W. ,2009). Nous ne la faisons pas ici car il nous semble judicieux d'utiliser d'abord des méthodes d'appariement et de ventilation multi-critère pour descendre de plusieurs niveaux, d'observer et comprendre les résultats et comme nous le verrons dans la partie consacrée aux résultats d'affiner le cas échéant à l'aide de ces méthodes.

4.5 Méthodologie de ré-agrégation

Pour toute ces méthodes nous procédons dans un deuxième temps à l'agrégation des lieux de résidence et d'emplois sur le carroyage de 1 km. Cette opération ne comporte pas de problème particulier puisque le maillage de 200 m est construit à partir du maillage de 1 km, et les localisations d'emploi sont des coordonnées (X/Y).

V Résultats et conclusion



La carte suivante expose les flux de déplacement en transport en commun transférés au carroyage de 1 km à l'aide de notre deuxième méthode. Nous nous attachons maintenant à évaluer de la qualité des méthodes selon la longueur des déplacements. La planche d'histogramme n°2 représente les distributions des distances de plus de 15 km.

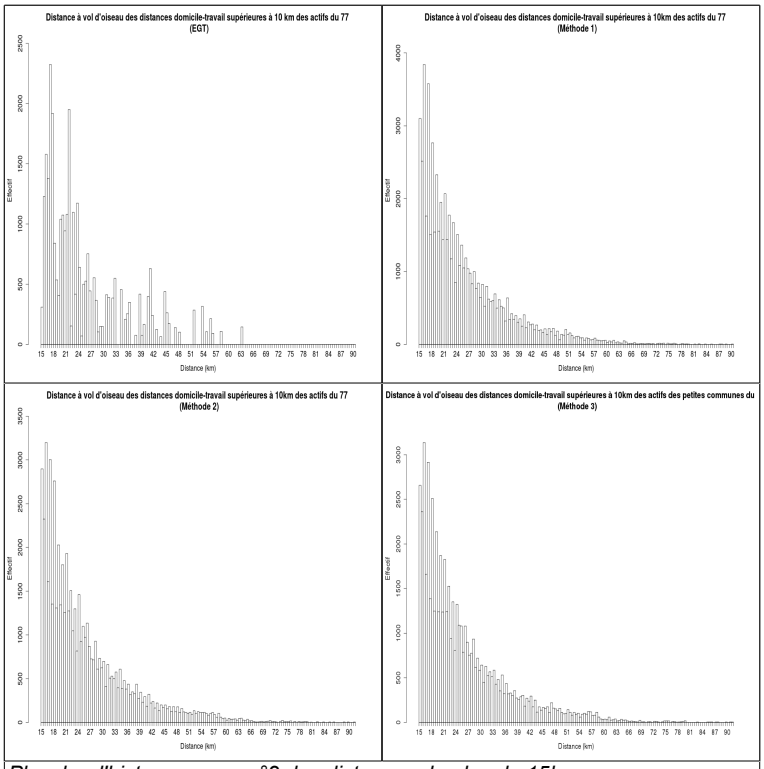


Planche d'histogrammes n°2 des distances de plus de 15km

Nous remarquons que les trois méthodes que nous avons mis au point ont des distributions de distances similaires, et globalement cohérente avec l'EGT. Nous observons une forte pente de 15 à 20 kilomètres puis un ralentissement de forme exponentielle négatif à l'instar de l'enquête globale transport. Nous rappelons que les erreurs potentielles sont comprises dans un intervalle d'erreur proportionnel à la superficie des communes. Nous observons que cet intervalle est suffisamment réduit pour ne pas affecter les distributions des distances de plus de 15 km.

Pour analyser les petites distances nous effectuons les histogrammes des déplacements intra-communaux et ceux des déplacements effectués entre deux communes limitrophes. Aussi, nous discriminons les déplacements en fonction de leur zone d'origine. Nous distinguons ainsi les déplacements ayant pour origine une commune de plus de 10 000 habitants ou non.

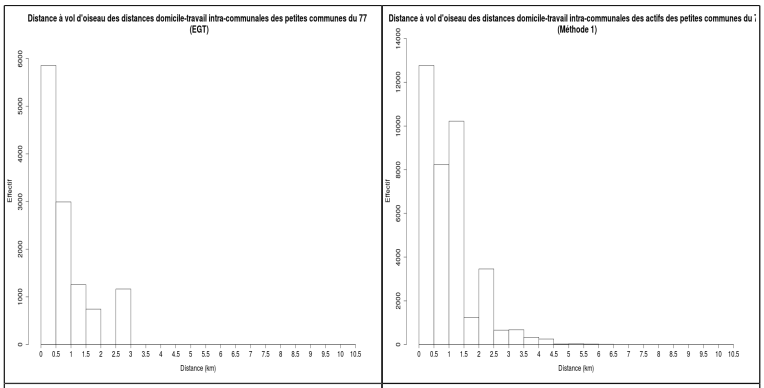


Planche d'histogrammes n°3 des distances des déplacements intra-communaux des petites communes (moins de 10 000 habitants)

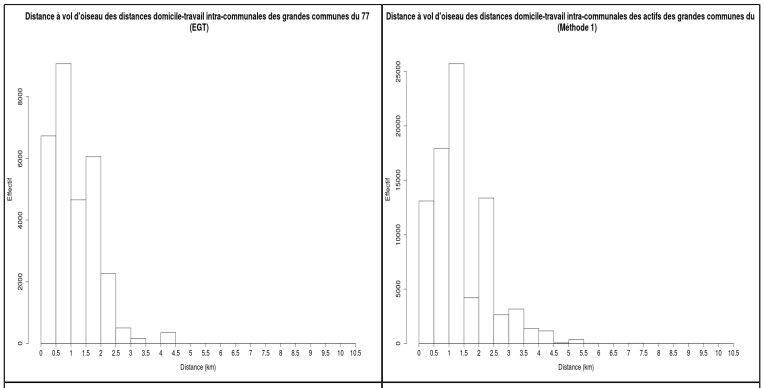


Planche d'histogrammes n°4 des distances des déplacements intra-communales des grandes communes (de plus de 10 000 habitants)

A l'aide de l'EGT, nous observons que les distances intra-communales sont plus grandes dans les grosses communes que dans les petites communes. Pour les communes de petites tailles démographiques la plupart des déplacements intra-communaux sont inférieurs à 1 km (selon l'EGT). Les trois méthodes de transfert reproduisent eux aussi de forts effectifs inférieurs à 1km, cependant la décroissance des effectifs selon

l’abscisse des longueurs est moins rapide que celle de l'EGT. Comparé à l'EGT il y a de trop de nombreux déplacements intra-communaux allant jusqu'à 2,50km. Ensuite, nous observons une queue de distribution plus longue pour les méthodes 2 et 3.

Pour les communes de plus de 10 000 habitants les distances relevées par l'EGT sont plus grandes, elles mesurent jusqu'à 2 km. Les distributions des méthodes de transfert reproduisent eux aussi d'importants effectif de moins de 2 km. La non ressemblance à une courbe de type exponentielle inverse s'explique ici par le pas des abscisse de 500m. En revanche nous observons une queue de distribution plus grande que celle de l'EGT.

L'analyse des déplacements effectuées entre deux communes limitrophes (planche d'histogramme n°5 et n°6) nous montre aussi que les déplacements entre deux communes limitrophes sont moins long pour les petites communes que pour les grandes communes. Il y a de fort effectif jusqu'à 4 kilomètres dans les petites communes, les déplacements sont nombreux dans les grandes communes jusqu'à 6 kilomètres. Nous remarquons également que nos trois méthodes sont similaires selon le type de commune. Comme pour le cas des flux inter-communaux les méthodes de transfert sur-estiment de quelques kilomètres les déplacements limitrophes. Ils ont une pente moins forte.

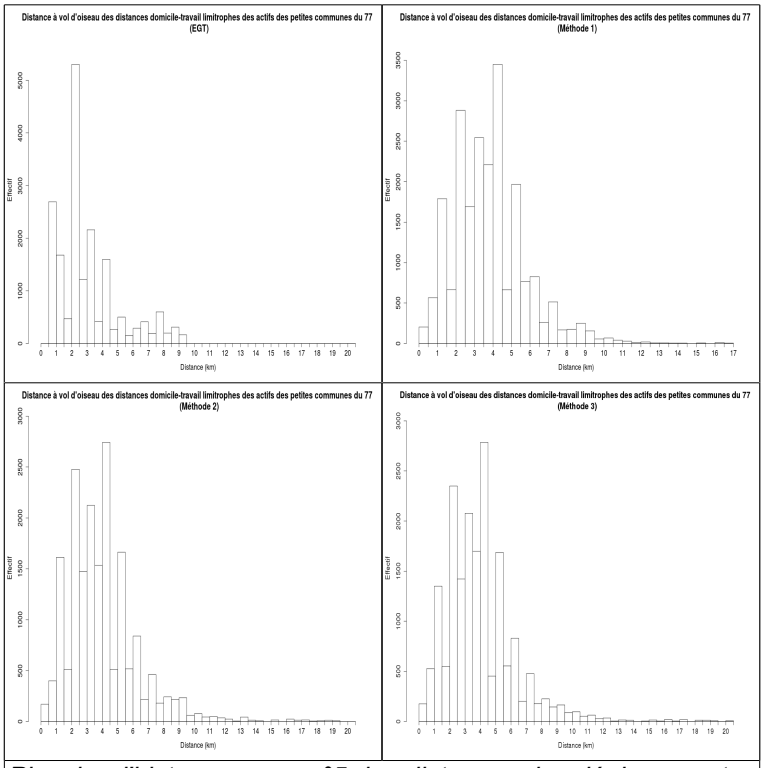


Planche d'histogrammes n°5 des distances des déplacements entre deux communes limitrophes (petites communes)

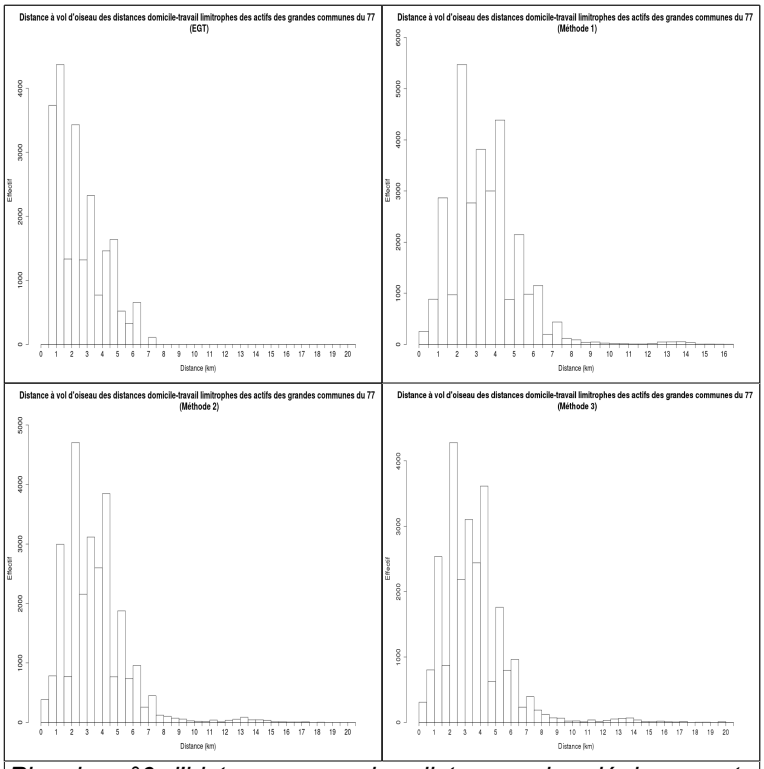
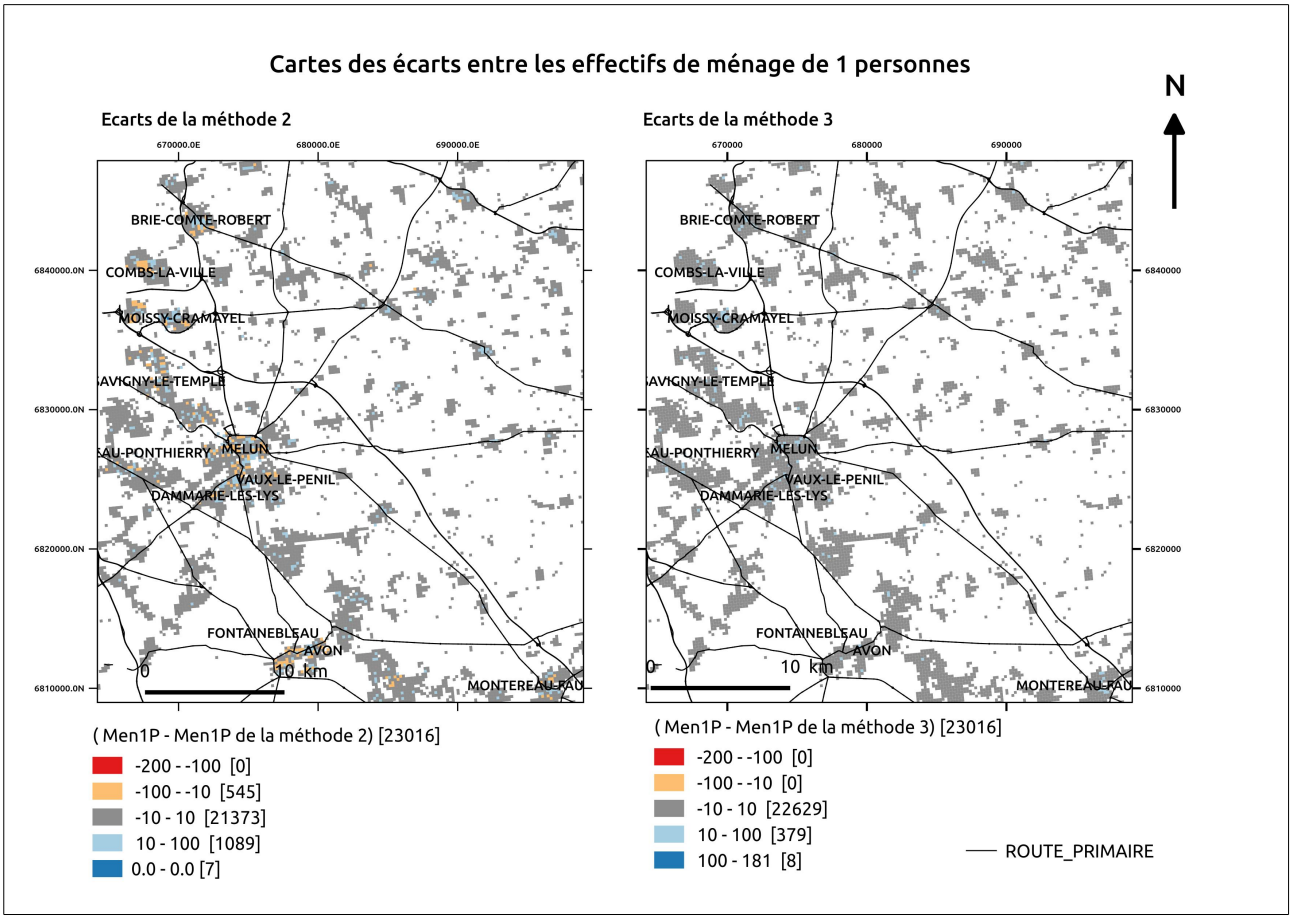


Planche n°6 d'histogrammes des distances des déplacements entre deux communes limitrophes (grandes communes)

Bien que les trois méthodes de transfert aient des distributions de distances semblables elles prennent en compte différemment les profils socio-économiques. La méthode n°1 ne considérant pas les profils sociaux il est inutile d'exposer la répartition résidentielle des individus. Pour comparer les méthodes de transfert des IRIS/communes aux carreaux de 200 m selon les méthodes par effectif de population ou multi-critères nous cartographions les écarts d'effectif des ménages de une personne avec le carroyage de 200 m. Nous avons choisi cette variable car elle est plus sensible.



Nous nous sommes concentrés sur cette carte sur la partie sud du département afin de mieux distinguer les écarts avec le carroyage. Il est figuré en bleu les carreaux dont les méthodes de désagrégation ont sous-estimées les effectifs. Les carreaux sur-évalués par les méthodes de désagrégation sont en rouges. Les écarts de moins de dix ménages de une personne sont en gris. Pour les deux méthodes nous observons que les carreaux en zones peu denses sont bien représentés. Il faut toutefois faire attention aux faibles effectifs. En revanche, la méthode n°2 a plus de mal à respecter les profils des ménages dans les centres plus denses. Quant à la méthode multi-critères, elle sous-estime légèrement les centre-villes.

Nous remarquons que la structure des lieux de résidence diffère bien selon les méthodes de transfert alors que les histogrammes de distributions des distances sont similaires. Nous en concluons que si nous souhaitons transférer les déplacements domicile-travail sur un carroyage de 1 km nous pouvons nous servir de la méthode n°1 qui ne considère pas les profils des ménages. Cette méthode plus simple apporte des résultats similaires aux méthodes plus sophistiquées. Lorsque nous prenons en compte la structure socio-spatiale du carroyage de population nous obtenons les même distribution de distance. Toutefois, nous avons observé que cette méthode sur-estime les déplacements courts à l'instar des autres méthodes. Les histogrammes ont une pente moins forte que celle de l'EGT.

Nous proposons en perspective de ces recherches d'effectuer un travail supplémentaire pour traiter de cette sur-estimation. Nous pouvons rapprocher les distances courtes produites à l'aide de distribution de référence (partie 4.4). Avec les méthodes 2 et 3 nous disposons des profils des ménages, nous sommes en alors en mesure de choisir un lieu d'emploi au lieu de résidence à l'aide de critère de distance spécifique aux catégories socio-professionnelles. Il nous faudra également observer si des critères de distance ne prenant pas en compte les profils socio-économiques peuvent être établis afin de construire une méthode simple.

Références :

Antoni J.P. (2010), Modéliser la ville. Forme urbaine et politiques de transport. Economica, Coll. Méthodes et approches, pp. 50-77

Arbia, G. (1989) *Statistical effect of spatial data transformations: a proposed general framework*, Taylor and Francis.

Bairoch, P. (1985). *De Jericho à Mexico: Villes et économie dans l'histoire*, Paris, Gallimard.

Bonin O., Lemssouguer M.H. (2012). *Approche probabiliste des liens entre distances et maillages, application à l'exploitation d'enquêtes origine destination*. Site des JMS-Journées de Méthodologie Statistiques.

Bourgin C (1978), *Les évolutions dans l'usage des modes de transports - influence des moments de transition dans le cycle de vie* , IRT, rapport de recherche, N°36, Arcueil, 38p.

Chaker, W., Proulx, M., Moulin, B., & Bédard, Y. (2009). Modélisation, Simulation et Analyse d'Environnements Urbains Peuplés. *Revue internationale de Géomatique*.

Debrand T., Taffin C. (2005) , *Les facteurs structurels et conjoncturels de la mobilité résidentielle depuis 20 ans*. *Economie et statistique*, 2005, vol. 381, no 1, p. 125-146.

Dupuy D., (2000) *L'urbanisme des réseaux, théories et méthodes, éclairer les controverses* , Ed, Armand Collin

Dupuy, G. (1999). *La dépendance automobile: symptômes, analyses, diagnostic, traitements*. Anthropos.

Gascon M.-O., Quetelard, B., Patiès, C., (2009) *Calcul a posteriori des distances dans les enquêtes ménages déplacements*, CERTU.

Goodchild M.F., Lam N.S. (1980) *Areal interpolation: a variant of the traditional spatial problem*. London, Ontario, Department of Geography, University of Western Ontario.

Kellerer, H., Pferschy, U., Pisinger, D. (2004). *Knapsack problems*. Springer Science & Business Media.

Openshaw S. (1977) *A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modeling*, Transactions of the institute of British geographers, 1977, p.459-472.

Openshaw S. (1978) *An empirical study of some zone-design criteria*, Environment and planning A, 1978, vol. 10, no 7, p. 781-794.

Openshaw S., Taylor P.J., (1979) *A million or so correlation coefficients: three experiments on the modifiable areal unit problem* , Statistical applications in the spatial sciences, 1979, vol. 21, p. 127 -144.

Openshaw S. (1981) *Le problème de l'agrégation spatiale en géographie* , L'Espace géographique, 1981, vol. 10, no 1, p. 15-24.

Openshaw S. (1984a). *Concepts and techniques in modern geography number 38: the modifiable areal unit problem,* Norwick, Geo Books, 1984.

Openshaw S. (1984b) *Ecological fallacies and the analysis of areal census data*, Environment and Planning A, 1984, vol. 16, no 1, p. 17-31.

Plumejeaud, C. *et al* (2009). Etude de méthodes de transfert d'indicateurs associés à différents découpages du territoire-Application à la ville de Grenoble. *Actes du colloque International de Géomatique et d'Analyse Spatiale-SAGEO*.2009.

Pivano C., Bonin O., Hubert J-P., (2014). *Désagrégation de flux pour le calcul des distances intra-communales des mobilités professionnelles*. *Actes du colloque International de Géomatique et d'Analyse Spatiale-SAGEO*.

Terrier C. (2006). Géomathématique des flux : influence du maillage sur les statistiques de déplacements.

Sites internet

INSEE, Documentation générale sur le carroyage , site de l'INSEE, [en ligne] http://www.insee.fr/fr/themes/detail.asp?reg_id=0&ref_id=donnees-carroyees&page=donnees- detaillees/donnees-carroyees/donnees_carroyees_doc.htm (consultée en Avril 2014)

INSEE, Connaître et comprendre le répertoire SIRENE, site de l'INSEE, [en ligne] http://www.insee.fr/fr/bases-de-donnees/default.asp?page=sirene/institutionnel.htm (consultée en Avril 2014)