

MODÉLISATION DES VARIATIONS AU COURS DE LA JOURNÉE DE LA POPULATION PRÉSENTE

AU NIVEAU INFRACOMMUNAL¹

Cynthia Gaborieau-Faivre (*)

(*) Insee, Département de l'action régionale

Introduction

Tour à tour, différents quartiers de la ville se peuplent et se dépeuplent au cours de la journée au gré de la succession des rythmes d'activité. Ceux d'entre eux qui concentrent le plus d'individus à un moment donné font rapidement l'objet d'attention par les acteurs publics, que ce soit pour les sécuriser (protection civile des populations) ou pour les doter de nouvelles infrastructures ou équipements².

Dès lors, pouvoir cibler ces zones et surtout se doter d'outils pour visualiser la hiérarchie urbaine qu'elles génèrent est essentiel.

On peut certes opérer sur le terrain pour y parvenir (enquêtes cordons, dénombrement par des caméras, ...), mais cela est coûteux en temps et en moyens mis en œuvre.

Le kit « population présente »³ exposé ici innove en exploitant les principaux déterminants de la localisation fine des individus au cours de la journée (sexe, catégorie socio-professionnelle, enfants à charge, vie en couple) pour modéliser leur lieu de présence à différents horaires.

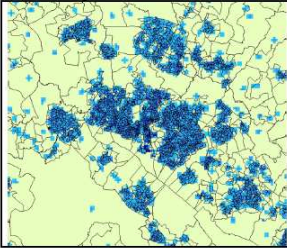
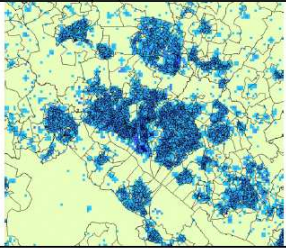
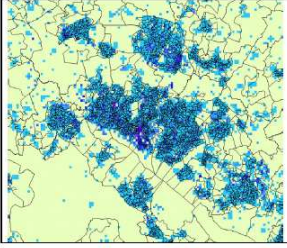
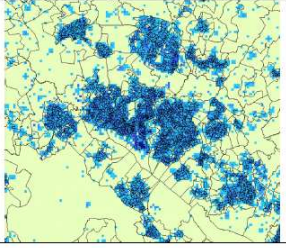

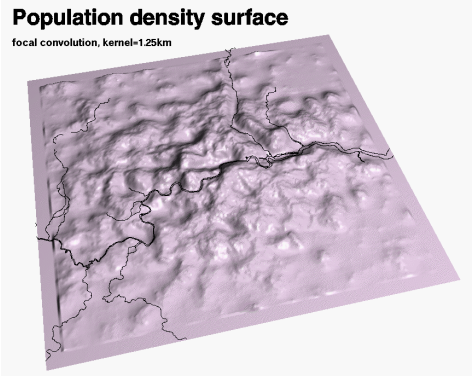
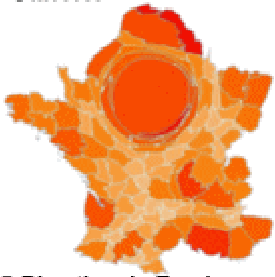
1 - L'état de l'art

À l'étranger, des estimations de populations à leur lieu de résidence et à leur lieu de travail ont pu être réalisées sur la base de registres, permettant d'estimer des populations de nuit et de jour (Canada), ou bien sur la base de modélisations conduisant à des estimations de populations par tranches horaires très fines (travaux de D. Martin au Royaume-Uni).

¹ L'auteur remercie Patrick Corbel, Jean-Michel Floch, Benoît de Lapasse, Jean-Luc Lipatz pour leurs précieux conseils et remarques, ainsi que Isabelle Pertile, Claire Poisson et Stéphane Vigneau pour leurs tests et suggestions.

² Ainsi derrière la prise de conscience récente de la nécessité d'accroître l'offre de services dans les gares parisiennes (coiffeurs, crèches, ...) c'est bien la question de l'aménagement des zones fortement réceptrices de flux qui est sous-jacente, avec pour enjeu la réalisation de travaux d'aménagement importants (implantation de restaurants, développements des espaces verts tels que des parcs ...)

³ La version du kit est à ce stade encore expérimentale et sera finalisée lors de la mise à jour évoquée en fin d'article.

<p>Source : site StatCanada</p>	<p>Source : travaux de D. Martin</p>
<div style="display: flex; flex-wrap: wrap;"> <div style="width: 50%;"> <p>(a) 02:00 night-time residential model</p>  </div> <div style="width: 50%;"> <p>(b) 08:00 early daytime model</p>  </div> <div style="width: 50%;"> <p>(c) 09:00 standard workday model</p>  </div> <div style="width: 50%;"> <p>(d) 18:00 evening model</p>  </div> </div>	<p>Figure 3: Preliminary daytime population model of the Southampton region, rendered as a 3-dimensional kml layer and overlaid in Google Earth.</p> 
<p>Source : travaux de D. Martin</p>	<p>Source : Direction du Tourisme (Carte départementale par anamorphose).</p>
<p>Population density surface focal convolution, kernel=1.25km</p> 	<p>Janvier</p>  <p>© Direction du Tourisme</p>

Récemment, différents travaux méthodologiques ont été réalisés en France sur l'estimation de populations présentes :

- l'estimation de ces dernières à un niveau départemental (travail réalisé par le ssm tourisme), puis la ventilation de ces populations au sein des bassins de vie (cette méthode est utilisée dans l'investissement du psar Analyse Territoriale « estimation d'une population présente à un moment donné sur un territoire donné » sur cette problématique).

- par ailleurs des explorations réalisées en France dans ce domaine à l'infracommunal ont mobilisé (entre autres) des données sur les déplacements.

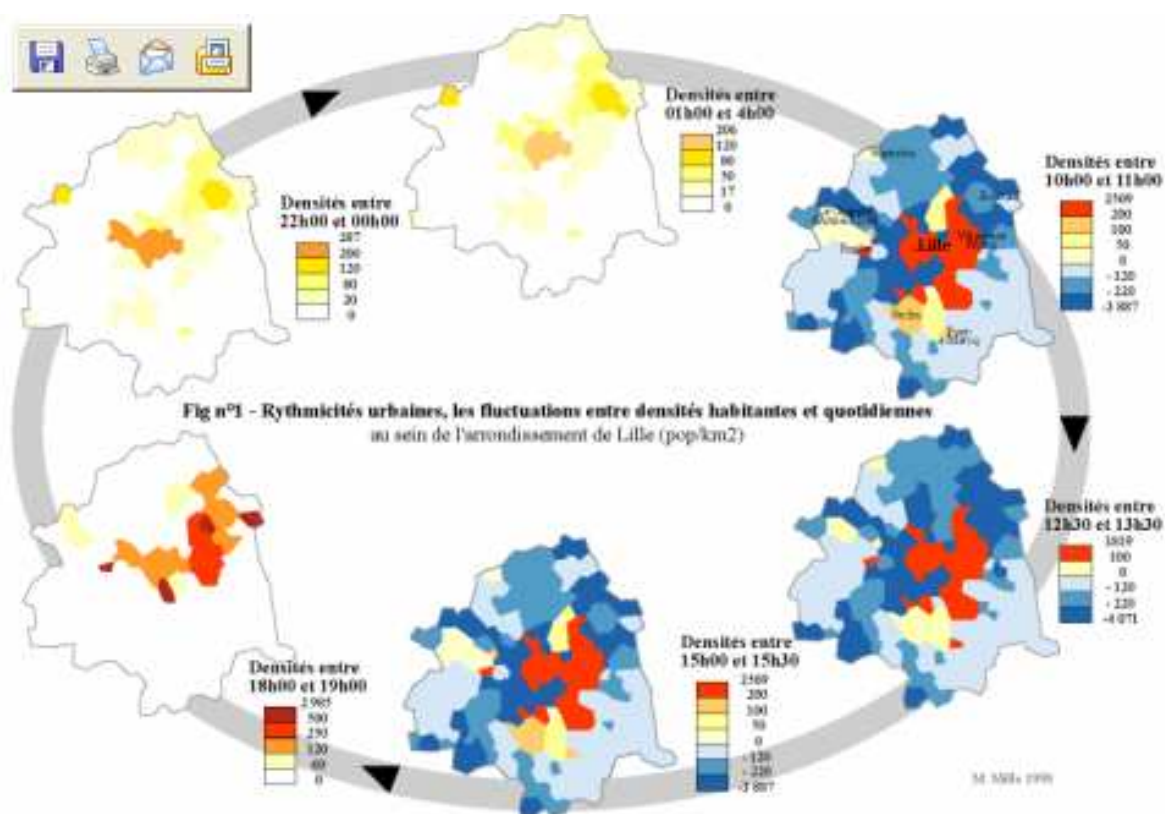


Fig n°1 - Rythmicités urbaines, les fluctuations entre densités habitantes et quotidiennes au sein de l'arrondissement de Lille (pop/km2)

fig1

Figure n° 1 : Rythmicités urbaines, les fluctuations entre densités habitantes et quotidiennes au sein de l'arrondissement de Lille (pop / Km)

Source : Matthieu Mille, « Des densités habitantes aux densités mouvantes l'exemple de la métropole lilloise », *Cybergeog*, Espace, Société, Territoire, article 121, mis en ligne le 25 février 2000, modifié le 03 novembre 2008. URL : <http://cybergeog.revues.org/index1866.html>⁴

La démarche adoptée dans l'article a permis d'aller plus loin que les cartes de population de jour et de nuit élaborées au Canada en proposant un découpage de la journée plus fin, sans toutefois aller jusqu'au degré de finesse des résultats de David Martin faute de disposer de données suffisantes (sur les transports et le tourisme notamment).

Les enquêtes ménages déplacement n'ont pas été mobilisées, car d'une part elles utilisaient comme localisation des origines et destinations des individus des « zones fines »⁵ de taille variable induisant un problème de MAUP⁶, et d'autre part elles n'avaient été réalisées que dans les plus grosses agglomérations et ne permettaient donc pas la production de résultats sur tout le territoire français métropolitain.

Quant aux travaux conduits par le PSAR Analyse Territoriale et le SSM tourisme, ils n'utilisaient pas la même résolution spatio-temporelle (mais considéraient un mois donné et un bassin de vie donné).

⁴ La méthode employée ici a consisté à choisir un découpage horaire de la journée fondé sur l'analyse des courbes de fréquentation des autoroutes (périodes de stabilité), puis à ventiler la population au travers des îlots en s'appuyant sur des courbes de fréquentation horaires des principaux établissements de la commune ainsi que sur des enquêtes commerciales.

⁵ Ces zones, fonction généralement de la densité de population, correspondent à des regroupement d'îlots Insee dans les zones urbaines denses et aux contours communaux dans les zones rurales.

⁶ MAUP ou Modifiable areal unit problem, c'est-à-dire problème de non homogénéité du zonage sur tout le territoire couvert se traduisant sur (et faussant) le message des cartes pouvant être produites.

2 - Méthodologie proposée

Une première méthode si l'on s'intéresse au détail de l'occupation dynamique d'un territoire est une démarche que l'on peut qualifier de « top-down » qui consiste à répartir géographiquement un total connu ou estimé (par exemple un total de personnes à l'extérieur le long des rues résidentielles).

La méthodologie proposée ici s'insère dans cette gamme de méthodes. Dans une première phase les volumes des populations « macros » situées respectivement à leur lieu de résidence, sur leur lieu de travail et à l'extérieur sont estimées.

Puis ces totaux sont ventilés au travers de clés de répartition spatiale ad hoc. Tout l'enjeu du kit va donc être de réaliser l'estimation de ces totaux et de proposer des clés à la fois raisonnables et réalistes.

2.1. Cadre du kit

Le cadre temporel qui va nous intéresser ici est assez restreint, à savoir la journée. Il va s'agir de situer la population présente, au niveau infracommunal, au sein d'un ensemble de communes contiguës de la France métropolitaine (non nécessairement classées en « grandes communes »⁷) à différents créneaux de la journée pour la semaine (respectivement le week-end). Le kit pourra être mobilisé afin de fournir des éléments de cadrage sur les zones recevant habituellement à certaines heures de la journée d'importants flux de personnes, conduisant à des pics de population présente distinguables au niveau communal. Le cas échéant des zonages infra-communaux précis pourront être étudiés en lieu et place du contour pluricommunal, tels par exemple les zones inondables d'une commune ou encore les zones de bruits.

Au sein de la journée, huit créneaux seront identifiés par la suite de sorte à avoir une couverture complète du cycle journalier :

Un découpage a priori intuitif de la journée serait :

Nuit 1H-4H : cœur de la nuit, période de plus grande immobilité
Nuit 4H-6H : petit matin
Jour 6H-8H : « début » des navettes domicile-travail
Jour 8H-12H : matin
Jour 12H-14H : pause méridienne, important « chassé-croisé »
Jour 14H-17H : début de l'après-midi
Jour 17H-20H : fin de l'après-midi
Nuit 20H-1H : soirée

Le créneau de midi constitue alors le créneau a priori le plus difficile à modéliser du fait de l'important « chassé-croisé » qui s'opère entre ceux prenant leur pause méridienne assez tôt et ceux au contraire qui sortent plus tardivement, et du fait de pratiques très diverses sur ce créneau (personnes retournant manger chez elles, personnes mangeant à la cantine de leur établissement, personnes mangeant dans un autre lieu de restauration ...).

Néanmoins, l'investissement devant pouvoir servir pour des préoccupations de protection civile des populations, une couverture totale du cycle journalier s'avère nécessaire, y compris sur ce créneau.

Par ailleurs une séparation entre les jours de semaine qui correspondent globalement à des périodes d'emploi, d'école ou d'université pour tous les membres de la famille et les jours de week-end pour lesquels l'ensemble des membres du ménage est a priori libéré a été effectuée.

D'autres séparations ont été envisagées : distinguer le samedi du dimanche (rappelons que 40% du chiffre d'affaires du commerce est réalisé les samedis) ou le mercredi qui est choisi par certaines mères de famille comme jour de temps partiel. Des tests ont montrés, au moins pour les mercredis à un niveau agrégé, que cela n'était très significatif, et par ailleurs cela aurait conduit à travailler sur un

⁷ Pour y parvenir on fera usage des fichiers issus de la géolocalisation à l'aide du cadastre de l'exploitation principale du RP produits à la Division des Etudes Territoriales de l'Insee.

nombre trop faible d'observations issues de l'enquête emploi du temps, et n'aurait sans doute pas été assez robuste.

Désormais pour plus de simplicité les créneaux seront numérotés de 1 à 8, le créneau 1 désignant la période 1H-4H.

2.2. Nécessité de cadrer le réservoir de population et sa porosité éventuelle

Une typologie simplifiée des entrants dans la zone (venant se rajouter aux résidents de la zone en question) est utilisée : elle comprend les personnes venant y travailler ou y étudier ainsi que les personnes en provenance du bassin de vie venues pour les équipements dont dispose la commune. Quant aux sortants, on y inclut les actifs sortant pour travailler et les étudiants sortant pour étudier ainsi que les personnes sortant à la recherche d'équipements non présents sur le territoire de la commune.

On a omis dans ce décompte les touristes venus au moins une journée ainsi que les excursionnistes venus pendant un laps de temps inférieur à la journée⁸. Ces populations ne seront donc pas prises en compte dans la suite de l'analyse. Néanmoins l'outil présenté ici propose en cas de disponibilité de données sur les touristes et leur localisation fine de les rajouter en « surimpression ».

De manière générale, tout événement ponctuel type « rassemblement pour assister à un spectacle » peut être rajouté par le même procédé. Cette gestion en surimpression résulte du fait que l'on ne sait pas a priori où sont prélevés les personnes affluant vers le spectacle, la seule connaissance raisonnable du phénomène se limitant au volume de places et au périmètre de la salle. Toutefois cela ne gêne en rien pour atteindre le but poursuivi car on va calibrer les équipements ou les plans d'évacuation en fonction de la population potentielle soit la population maximale pouvant être présente compte-tenu des informations disponibles.

Le traitement des « points d'accumulation » avec toute la population en transit dans les gares, dans les aéroports, dans les zones portuaires se heurte à des difficultés matérielles pour récupérer des données sur les visiteurs des gares à différents créneaux de la journée.

Compte-tenu néanmoins de l'importance des flux concernés, une estimation de la population présente dans les principales gares SNCF (celles accueillant plus d'un million de voyageurs par an) est réalisée. Cette population est modélisée en deux composantes : population en migration domicile-travail présumée venir en TER et reste de la population en visite est produite et intégrée aux cartes générées par l'investissement. Cette estimation s'appuie sur le nombre de voyageurs par an des principales gares françaises et sur des modélisations raisonnables de la répartition du trafic dans les gares en fonction des différentes heures de la journée.

2.3. Fondements du kit

2.3.1. Une hypothèse d'une attraction différenciée sur les segments de la population des lieux de résidence, de travail et d'étude ou encore de l'extérieur, et qui évolue au cours de la journée.

Les effectifs des lieux d'emplois, d'étude, de loisir, de shopping, ... fluctuent au cours de la journée. Or les flux, une fois décomposés par profils socio-démographiques des individus, s'avèrent profondément dissymétriques.

⁸ Toutefois, concernant ces derniers, on peut supposer que leur venue est en partie liée à des conditions météorologiques particulières ou encore à la tenue d'événements exceptionnels dans la commune (meeting sportif, conférence, exposition ...) qui ne peuvent être contrôlées dans le cadre de l'investissement. Certaines personnes encore sont simplement « de passage » (cas par exemple où une autoroute traverse une commune). Leur importance revêt sans doute en partie un caractère saisonnier. Quant aux touristes, l'absence de connaissance sur les flux de personnes venues en vacances dans la commune en logeant chez des amis ou de la famille constitue à son tour une difficulté. En outre pour les touristes comme pour les excursionnistes l'absence de données empêche leur localisation fine à différents créneaux de la journée.

D'où l'idée d'introduire une segmentation de l'analyse en fonction de facteurs socio-démographiques tels que l'âge, le genre, la catégorie socio-professionnelle, ...

2.3.2. Une modélisation au niveau individuel désagrégée

On réalise un couplage des données du recensement avec celles de l'enquête emploi du temps afin d'ajouter une dimension « localisation au cours de la journée » à la caractérisation des individus échantillonnés.

L'enquête emploi du temps permet de connaître la teneur habituelle du carnet d'activité journalier d'un individu selon son profil socio-démographique, et pour chacune des activités décrites indique si l'individu est situé chez lui, à son travail ou encore à l'extérieur⁹.

On peut donc dès lors pour chaque individu enquêté au recensement de la population déterminer au vu de ses caractéristiques sa propension à être chez lui, à son travail ou à l'extérieur pendant tel ou tel créneau horaire de la semaine ou du week-end à l'aide de modèles issus de l'enquête emploi du temps (cf. encadré infra).

Plus précisément on va travailler sur chaque créneau horaire et par période (week end/ jour de semaine). Pour chaque individu de la zone d'étude (personne y habitant ou y travaillant ou y étudiant) on calcule une propension en fonction de ses caractéristiques à résider chez lui, à se trouver sur son lieu de travail ou à être à l'extérieur. La somme des propensions sur un individu vaut 1 par construction.

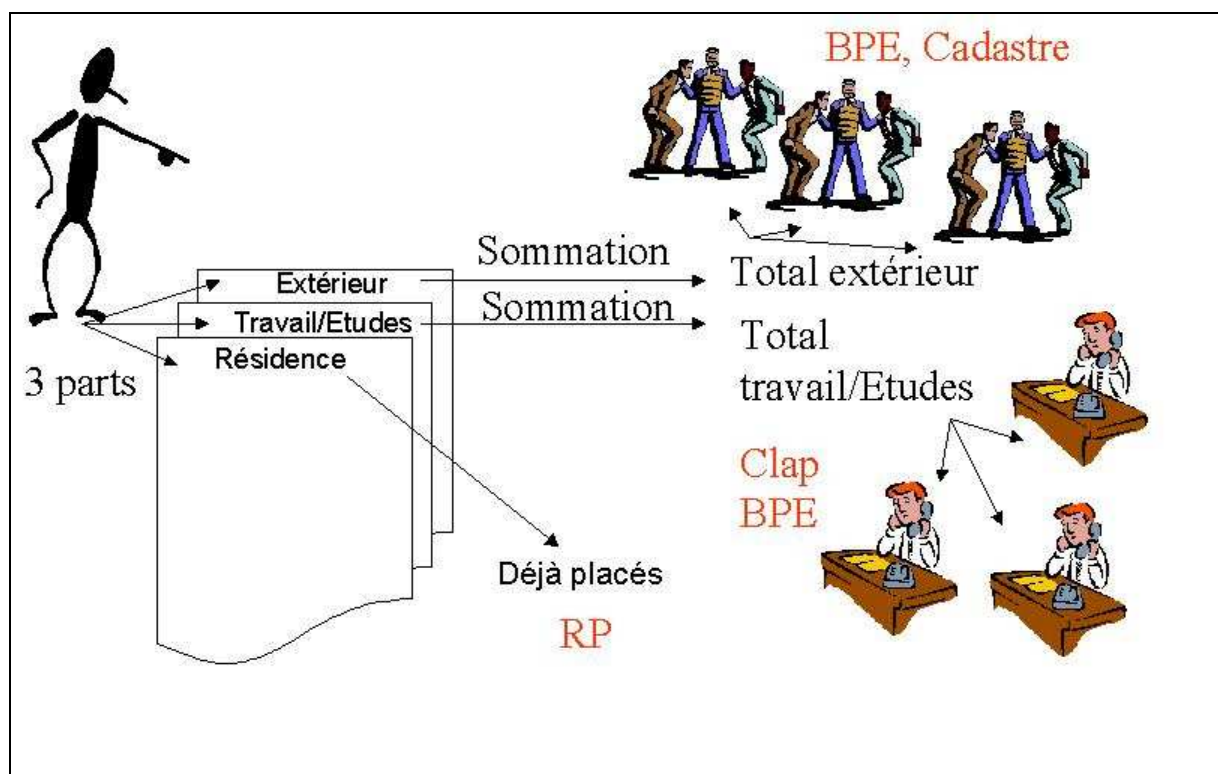
2.3.3. Une agrégation des résultats individuels pour constituer des populations « macros » à leur lieu de travail ou d'étude et à l'extérieur

On calcule alors simplement le nombre total de personnes se trouvant à leur lieu de travail et à l'extérieur, en sommant toutes les propensions à être dans de tels lieux pour tous les individus. A ce stade, on a pour les personnes se trouvant chez elles une propension et une localisation (donnée par le RP et que l'on utilisera directement), et pour le reste des personnes, un nombre total de personnes au travail et à l'extérieur sur l'ensemble de la zone d'étude. Il nous reste donc à répartir ces effectifs au travail et à l'extérieur sur l'ensemble de la zone par des clés de répartition spatiale, pour estimer le nombre de personnes de cette catégorie dans chaque carreau.

2.3.4 - Une désagrégation finale de ces totaux, et une combinaison avec les résultats sur les personnes restées à leur domicile

On reconstitue grâce à des clés de répartitions spatiales plus ou moins schématiques quelles peuvent être leurs localisations à ces créneaux., d'où un positionnement global de la population présente au sein de la commune. Pour chaque carreau et pour chaque créneau et période, on aura une estimation du nombre de personnes par répartition socio-démographique (sexe, cs tranche d'âge) et par localisation (habitation, travail et extérieur). On somme alors simplement pour chaque carreau et pour chaque créneau et période le nombre de personne quelque soit la catégorie socio -démographique et la localisation.

⁹ En plus des lieux de résidence, de travail ou d'études, de consommation (achats) on a également les lieux du cercle des relations (familles, amis) et les lieux de loisirs. Ceux-ci sont bien décrits et distinguables dans l'enquête emploi du temps, mais en revanche on ne dispose d'aucune clé de répartition spatiale pour les répartir (à l'exception très minoritaire des cinémas).



Explication de la méthode

2.3.5 - Synthèse

Encadré sur la méthode¹⁰ retenue :

On repère dans un premier temps l'ensemble des individus enquêtés par le recensement en lien avec la commune (soit qu'ils y résident, soit qu'ils y travaillent ou y étudient).

Le but est de déterminer pour chacun une propension à fréquenter leur lieu de résidence, leur lieu de travail ou encore l'extérieur pendant un créneau horaire et une période (semaine ou week-end) donnée. Ces propensions, combinées au budget temps moyen (en part) consacré par l'individu à ces différentes localisations au sein du créneau puis calées de façon à sommer à 1, permettent alors de connaître à quelle hauteur l'individu participe aux populations présentes dans les différents types de lieux. Cet exercice est en fait pratiqué par profils socio-démographiques des individus selon des variables assez naturelles (fait d'avoir ou non des enfants, vie en couple, sexe, catégorie socio-professionnelle combinée avec l'âge¹¹), et ce afin de renforcer la synchronisation des individus au sein d'un profil pour la fréquentation des différents lieux. Cette segmentation va dès lors conduire à des contributions différentes des individus aux populations des différents lieux en fonction de leurs profils.

¹⁰ Une alternative à cette méthode pourrait consister à effectuer non plus une modélisation au niveau individuel mais une modélisation fondée sur les courbes de fréquentation horaires des bâtiments tirées de l'enquête emploi du temps. Toutefois, cette méthode serait en réalité plus fruste que la première car elle ne tiendrait pas compte de la structure de la population à placer. Cette méthode n'est donc pas envisagée.

¹¹ Les variables âge, sexe et catégorie socio-professionnelle seront examinées en particulier dans ce cadre, car ce sont des variables dont on peut raisonnablement penser qu'elles influent sur la localisation de l'individu. Pour l'enquête emploi du temps, ces variables signalétiques sont bien présentes car elles font partie du tronc commun de l'enquête. Les covariables à utiliser dans le modèle doivent absolument être des variables communes à la source enquête emploi du temps et au recensement de la population.

Techniquement, on va donc procéder en deux étapes. La phase de modélisation fera appel à la technique de régression logistique pour une variable dépendante non ordonnée (modèle à choix discret). La modélisation économétrique¹² (avec les covariables indiquées supra) va s'écrire :

$$\ln\left(\frac{p_fréquenter_lieu_i}{1-p_fréquenter_lieu_i}\right) = \alpha * sexe + \beta * couple + \gamma * enfant + \delta * categorie$$

où catégorie est une variable captant simultanément l'effet âge et l'effet catégorie socio-professionnelle (cs) :

catégorie distingue ainsi les actifs ouvriers, employés, de professions intermédiaires ou cadres, les élèves (de moins de 18 ans), les étudiants (d'au moins 18 ans), les chômeurs de moins de 45 ans, les chômeurs de plus de 45 ans, les retraités de moins de 75 ans et les retraités de plus de 75 ans.

Cette façon d'opérer s'est avérée plus robuste car les effets du sexe, du fait d'avoir un enfant ou de vivre en couple sont estimés à travers tout l'échantillon et non simplement pour une catégorie, ce qui permet de limiter l'impact de l'erreur d'échantillonnage¹³.

Puis le calage opéré correspondra à :

$P_calée_fréquenter_lieu_i = (p_fréquenter_lieu_i * budg \text{ tps moyen passé au lieu } i) / (\text{somme sur } j \text{ } p_fréquenter_lieu_j * budg \text{ tps moyen passé au lieu } j)$.

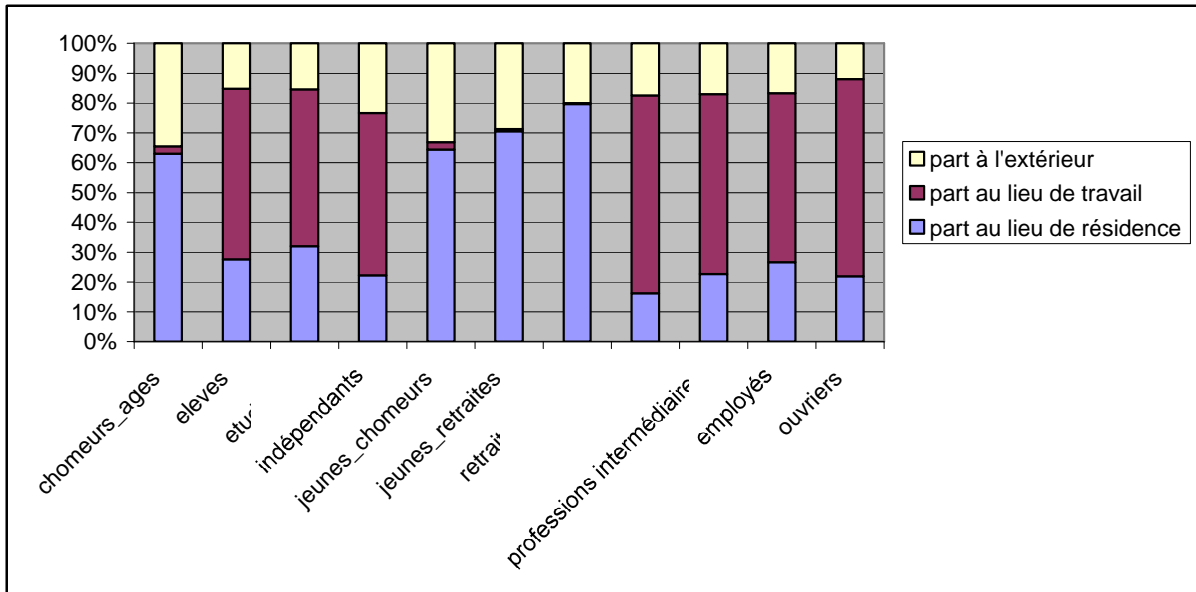
De la sorte on évite bien les doubles comptes.

Ensuite, des totaux sont formés par type de lieu, le total des personnes au travail étant calculé sur l'ensemble des personnes travaillant au sein de la commune considérée, et le total des personnes à l'extérieur étant effectué de sorte qu'une personne soit comptabilisée via la part d'elle-même qui fréquente l'extérieur pendant le créneau si elle réside et travaille ou étudie dans la commune, et comptabilisée via la moitié de cette part sinon. Cela correspond à l'hypothèse selon laquelle une personne qui ne travaille pas dans la commune où elle réside va fréquenter les commerces, restaurants, ... en moyenne une fois sur deux sur sa commune de résidence et une fois sur deux sur sa commune de travail.

Une sortie directe de la modélisation réalisée est par exemple la subdivision en parts des hommes célibataires et sans enfant, respectivement pour le créneau 8H-12H de la semaine ...

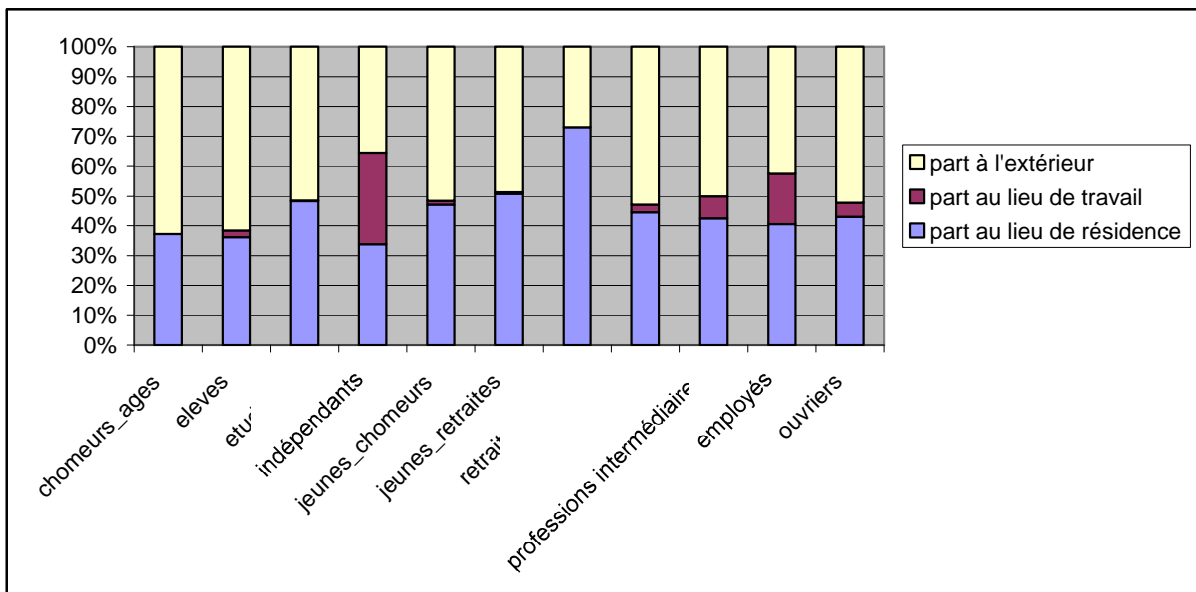
¹² Le modèle logit va correspondre au choix binaire (0/1) de fréquentation ou d'absence de fréquentation du type de lieu considéré.

¹³ La qualité de modélisation a été vérifiée sur la base du critère ROC, supérieur à 0,7 pour la grande majorité des modèles. Les résultats obtenus en terme d'indicateur ROC sont détaillés en annexe.



Pour lequel on constate que les cadres et les ouvriers sont très largement présents sur leur lieu de travail

et le créneau 14H-17H du week-end :



... pour lequel on constate que les indépendants consacrent une part non négligeable de ce créneau à leur travail.

2.4 - Analyse de la qualité de la modélisation

On dispose d'une variable dichotomique valant 1 si l'individu possède la caractéristique A et 0 sinon.

Si on se dote d'une règle d'affectation de chaque observation à une classe, par exemple en affectant à la classe + toutes les observations dont la probabilité d'appartenir à cette classe est supérieure à la probabilité de ne pas y appartenir (soit à 0.5), cette règle fournit une matrice de confusion et deux indicateurs : le taux de vrais positifs et le taux de faux positifs.

La matrice de confusion comptabilise les effectifs pour chaque croisement de modalité (constaté/prévu). Le taux de vrais positifs (encore appelé sensibilité) est le ratio des vrais positifs sur l'ensemble des positifs. On définit par ailleurs le taux de faux positifs comme le nombre de faux positifs rapporté au nombre total de positifs. La spécificité est définie comme le complément à 1 de ce taux.

Cette matrice de confusion et les deux indicateurs associés seraient différents pour un autre choix de la règle d'affectation.

Le diagnostic ROC (« Receiving Operating Characteristics ») associé aux régressions logistiques est obtenu en faisant varier un « seuil » ou « score » pour la probabilité d'appartenir à la classe + de 1 à 0.

La courbe ROC est alors la courbe paramétrique obtenue en mettant en relation le TFP en abscisse et le TVP en ordonnée. Un indicateur synthétique associé (noté couramment AUC) est l'aire comprise entre cette courbe et l'axe des abscisses. Dans le meilleur des cas elle vaut 1 et à l'opposé lorsque le modèle ne présente aucun pouvoir prédictif elle vaut 0.5. (ce qui équivaut à un simple tirage à pile ou face).

La comparaison de deux modèles au moyen des courbes ROC s'effectue très simplement : un modèle est meilleur qu'un autre (quel que soit la matrice de coût de mauvaise affectation) si et seulement si sa courbe ROC se situe au-dessus de celle de l'autre modèle. Il s'ensuit que les meilleurs modèles sont ceux qui participent à l'enveloppe convexe de l'ensemble des courbes ROC des différents modèles possibles. Les autres sont systématiquement dominés par un autre modèle et à ce titre peu intéressants.

Si on note nc le pourcentage de paires concordantes, nd le pourcentage de paires discordantes et t le nombre total de paires parmi les observations, on montre que $AUC = (nc + 0.5 * (t - nc - nd)) / t$.

Cet indicateur présente un avantage sur d'autres possibles notamment parce qu'il reste opérationnel même dans le cas de distributions très déséquilibrées.

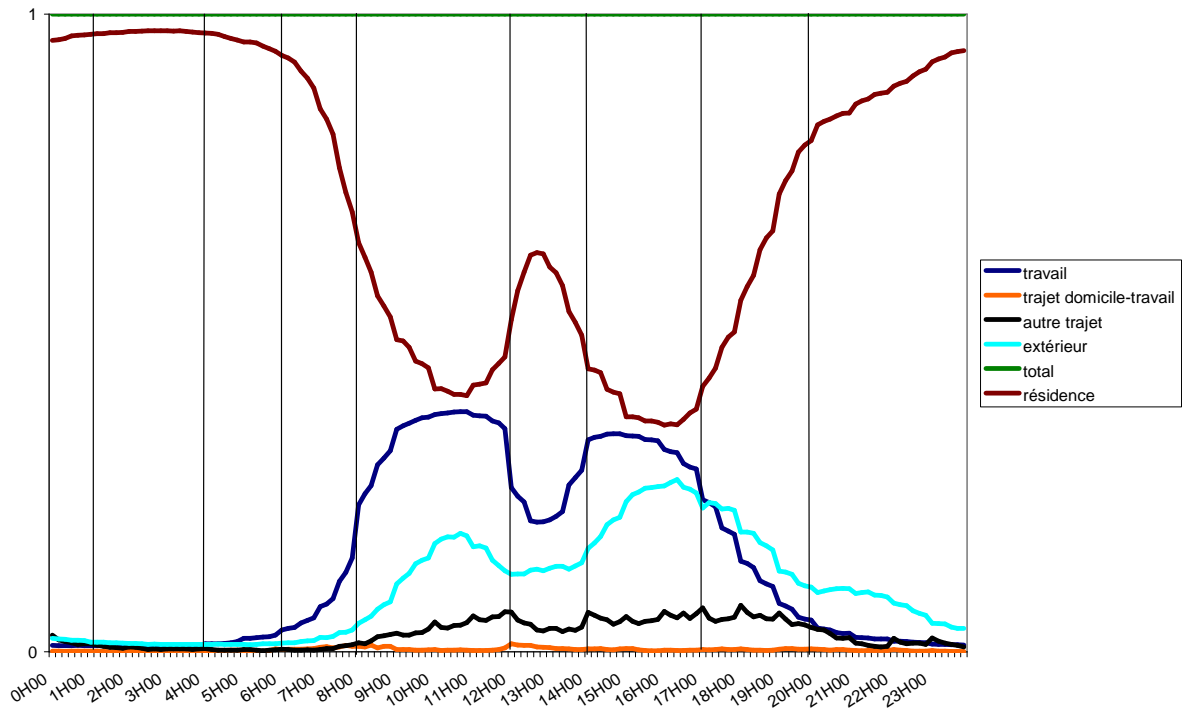
On reporte le lecteur à l'annexe pour l'examen des indicateurs ROC obtenus pour les 48 modèles.

3 - Les sources

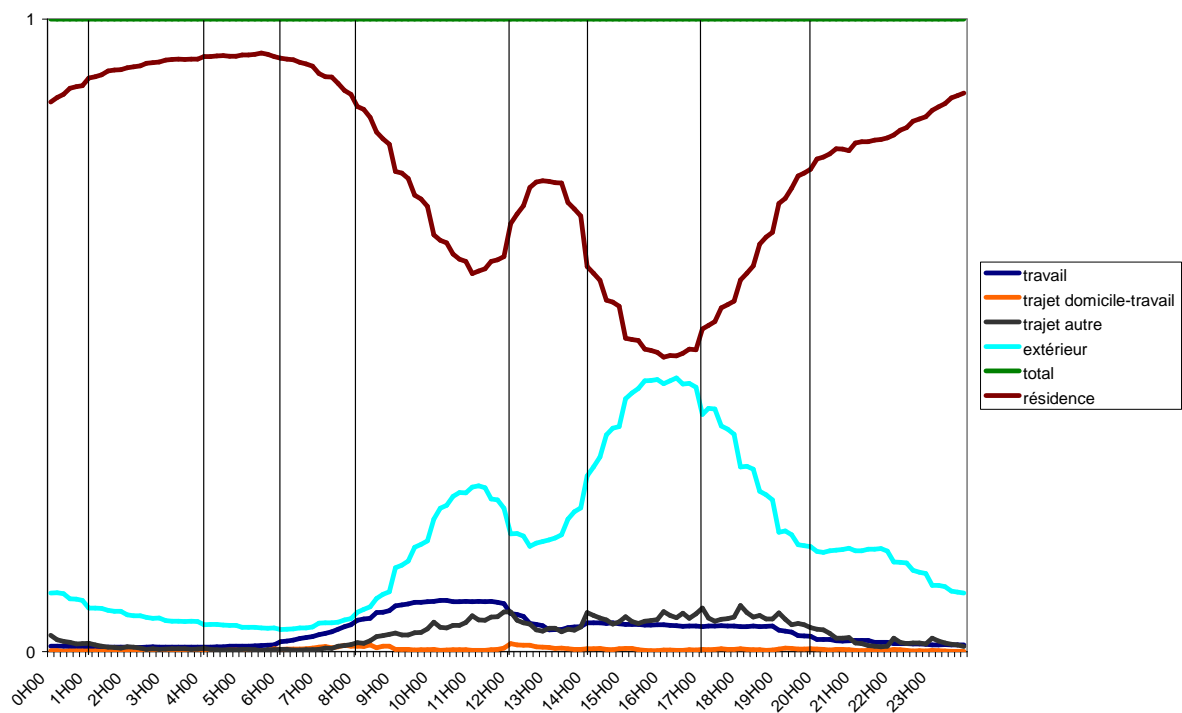
3.1 - Mesurer les localisations au cours de la journée : l'enquête emploi du temps

L'enquête emploi du temps de l'Insee est échantillonnée auprès de 12 000 résidences principales. Un carnet d'activité journalière est distribué à tous les individus âgés de 15 ans et plus, dans lesquels ils doivent décrire pour une journée qui leur est fixée toutes les activités et les lieux (résidence, étude ou travail, extérieur) où elles se déroulaient qui les ont occupés, et ce par tranches de 10 minutes.

On a représenté ci-dessous les courbes de fréquentation des différents lieux pour la semaine observées à partir de l'EDT 1999 toutes catégories et sexe de personnes confondus :



puis pour le week-end :



Comme attendu, on repère les comportements différenciés selon les tranches horaires, avec en particulier la pause méridienne qui se détache nettement.

Le découpage horaire adopté semble raisonnable au vu des courbes de fréquentation horaire, dans la mesure où il correspond à des périodes de stabilité relative des volumes de personnes dans les différents lieux.

Une question légitime qui se pose est la distinction éventuelle dans les modèles selon la taille de l'agglomération. Afin de juger de la pertinence de cette nouvelle variable dans la segmentation (variable comprise ici à trois modalités à savoir : petite agglomération de moins de 100 000 habitants, grande agglomération régionale de plus de 100 000 habitants, et agglomération parisienne), on procède en trois étapes.

- Pour pouvoir comparer les performances des modèles de fréquentation des différents lieux selon les variables de segmentation qui y sont incorporées, on définit l'indicateur du taux de bien classés selon les modèles. Plus précisément, on estime les modèles sur l'ensemble de l'échantillon des carnets de l'enquête emploi du temps et on regarde a posteriori quel pourcentage de personnes se voient prévoir par les modèles leur comportement réel (fréquentation ou non-fréquentation d'un lieu à un créneau donné).

La fréquentation ou non du lieu de résidence au créneau 5 est prévue à hauteur de 72 % si on introduit la variable de segmentation liée à l'agglomération de résidence contre 70 % si on ne l'introduit pas.

En revanche pour les autres créneaux étudiés (créneaux 4 et 7 en semaine¹⁴) les performances des modèles au regard du taux de bien classés sont identiques (respectivement de 78 % et 76 %).

Par conséquent les modèles de fréquentation sont estimés de manière aussi satisfaisantes dans les deux cas.

- On évalue l'impact sur les budgets temps moyens :

En particulier les budgets temps moyens pour le créneau 4 (8H-12H) en semaine au lieu de travail sont les suivants :

Avec prise en compte de l'agglomération :

¹⁴ On étudie trois créneaux pour lesquels l'impact a priori peut être palpable, à savoir le créneau 4 (8H-12H), le créneau 5 (12H-14H) et le créneau 7 (17H-20H), tous trois considérés en semaine

Catégorie	Sexe (1=homme, 2=femme)	Strate (de modalité croissante par taille de l'agglomération)	Budget temps pour le créneau 8H-12H alloué au travail en semaine
Salarié cadre	1	1	0,86
Salarié cadre	1	2	0,82
Salarié cadre	1	3	0,8
Salarié cadre	2	1	0,85
Salarié cadre	2	2	0,79
Salarié cadre	2	3	0,72
Salarié de profession intermédiaire	1	1	0,87
Salarié de profession intermédiaire	1	2	0,86
Salarié de profession intermédiaire	1	3	0,81
Salarié de profession intermédiaire	2	1	0,82
Salarié de profession intermédiaire	2	2	0,85
Salarié de profession intermédiaire	2	3	0,77
Salarié employé	1	1	0,91
Salarié employé	1	2	0,81
Salarié employé	1	3	0,8
Salarié employé	2	1	0,81
Salarié employé	2	2	0,83
Salarié employé	2	3	0,81
Salarié ouvrier	1	1	0,94
Salarié ouvrier	1	2	0,92
Salarié ouvrier	1	3	0,95
Salarié ouvrier	2	1	0,94
Salarié ouvrier	2	2	0,86
Salarié ouvrier	2	3	0,86

Sans prise en compte de l'agglomération :

Catégorie	sexe	Budget temps pour le créneau 8H-12H alloué au travail en semaine
Salarié cadre	1	0,83
Salarié cadre	2	0,78
Salarié de profession intermédiaire	1	0,85
Salarié de profession intermédiaire	2	0,81
Salarié employé	1	0,85
Salarié employé	2	0,82
Salarié ouvrier	1	0,94
Salarié ouvrier	2	0,91

La variable d'appartenance à un certain type d'agglomération semble avoir un impact sur les budgets temps. Ainsi les femmes cadres des petites agglomérations passent 85 % de la période 8H-12H en semaine sur leur lieu de travail contre 72 % pour celles de l'agglomération parisienne. Ces écarts de budgets temps semblent accréditer l'hypothèse que la variable de strate est une variable explicative du comportement de localisation des individus. Reste toutefois à mesurer globalement l'impact de cette variable.

- Calcul d'un indicateur d'écart entre les distributions par carreaux de population présente en semaine aux créneaux 4 (8H-12H) 5 (12H-14H) et 7 (17H-20H) :

Afin de juger de l'impact global de l'introduction de la variable de strate dans la modélisation globale, on a calculé successivement le nombre de personnes présentes aux créneaux 4 5 et 7 de la semaine par carreaux pour les deux modèles (avec et sans cette variable) puis pour un créneau donné considéré un indicateur synthétique défini comme le ratio de :

- au numérateur, la somme des différences absolues des effectifs par carreaux
- au dénominateur, la moyenne des populations présentes estimées avec et sans prise en compte de la variable agglomération.

Pour le créneau 4 en semaine, cet indicateur se monte à 0.92 %.

Pour le créneau 5 en semaine, il atteint 3.71 %.

Enfin pour le créneau 7 en semaine, il est de 1.96 %.

Conclusion : l'impact de la prise en compte de la variable agglomération apparaît donc mesuré.

Le gain lié à l'incorporation de la variable tranche d'unité urbaine dans l'estimation des modèles est faible et la petite taille de l'échantillon EDT conduit à limiter le nombre de variables explicatives afin de conserver un effectif suffisant dans chaque croisement de variables explicatives.

Cette variable n'a finalement pas été introduite dans l'analyse suite à l'arbitrage réalisé.

3.2 - Mesurer les univers socio-démographiques : le recensement¹⁵

Cette source très riche permet de connaître les caractéristiques détaillées des personnes qui sont introduits dans l'étude.

o Distinction entre les différentes « catégories » de population du RP

L'enquête emploi du temps ne couvre que les individus âgés d'au moins quinze ans de la population des ménages. D'où l'idée de distinguer au sein des individus enquêtés par le RP en fonction de la catégorie de population à laquelle ils appartiennent (population des ménages, mais également personnes effectuant un service de moyen ou long séjour, communautés religieuses, casernes, foyers étudiants, détenus, personnes en établissement social de court séjour, autres catégorie de population, habitations mobiles, sans abri, marinières) et de leur âge afin de n'appliquer une modélisation de type économétrique que sur le champ de l'enquête emploi du temps. Seront dès lors étudiés dans ce cadre les retraités, la population active occupée salariée, la population active occupée non salariés (agriculteurs non compris), la population active au chômage, et les élèves ou étudiants âgés d'au moins quinze ans.

En revanche les non actifs qui n'étudient pas et ne sont pas retraités, les enfants et adolescents âgés de moins de quinze ans, les indépendants salariés, les personnes qui ne vivent pas au sein d'un ménage et les agriculteurs ne feront pas l'objet d'une modélisation. On réservera à ces catégories de la population des comportements très simplifiés : à l'exception des foyers étudiants et des enfants et adolescents âgés de moins de quinze ans on considérera que toutes ces catégories demeurent à leur emplacement de nuit. En outre on fera des hypothèses simples sur la localisation des enfants et adolescents âgés de moins de quinze ans (placés à l'école sur les créneaux 4, 5 et 6 en semaine et à leur lieu de résidence le reste du temps).

3.3 - Quelles clés de répartitions spatiales ?

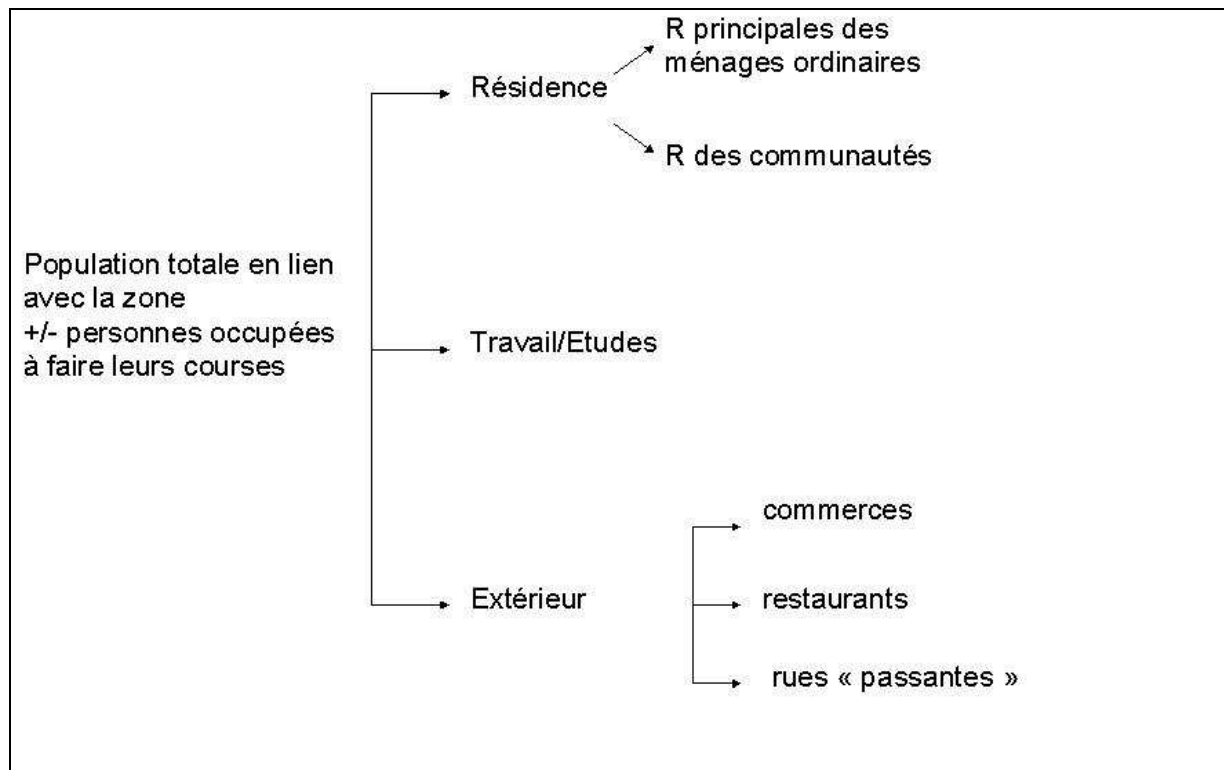
Concernant les lieux de résidence on ne dispose ici que des lieux de résidence enquêtés au cours du cycle quinquennal du recensement rénové de la population. On aurait pu envisager d'essayer de ventiler les effectifs recensés dûment pondérés au sein des adresses de la BSA (Base de sondage des adresses du recensement) mais au vu des objectifs du kit cela a semblé apporter une complication inutile dans les traitements et a donc été écarté.

Pour ce qui est des lieux de travail, on a mobilisé la source la plus riche au niveau du positionnement à un niveau fin des différents postes en fonction des caractéristiques de la personne pourvoyant l'emploi, à savoir le fichier « postes » de clap¹⁶ géolocalisé à l'aide de siren.

¹⁵ Les données fiscales présentent le grand avantage d'être exhaustives. Mais nous avons préféré utiliser le RP qui, bien que par sondage dans les communes de plus de 10 000 habitants, nous permet d'avoir la CS des personnes.

¹⁶ Connaissance locale de l'appareil productif

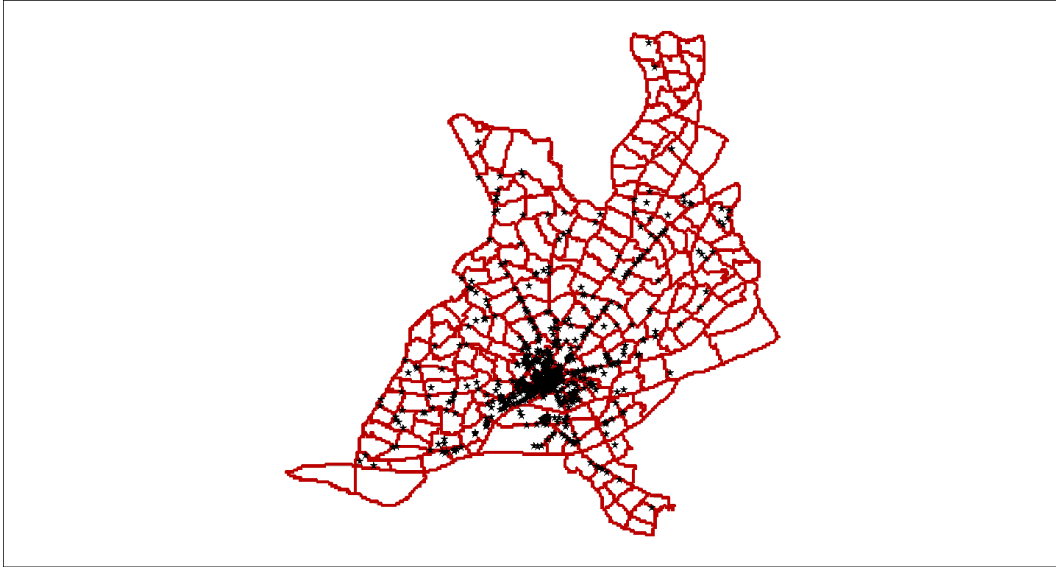
Son détail du nombre de postes au sein d'un établissement attribué à des personnes répondant toutes à un même profil sexe*catégorie professionnelle a donc été exploité ici.



Les autres lieux de travail (correspondant aux lieux de travail des indépendants non salariés) ont été assimilés à la localisation des services de la BPE (Base permanente des équipements de l'Insee) hormis : la police, la poste, la trésorerie, l'ANPE, la gendarmerie, mais également à la localisation des commerces et à celle des professions libérales travaillant dans le domaine de la santé.

Les lieux de scolarisation ont été tirés de la BPE. Dans un premier temps faute d'avoir la variable « capacité » de la BPE renseignée chaque établissement scolaire a été doté du même poids (égal à l'inverse du nombre total d'établissements scolaires sur le territoire étudié).

Les commerces et restaurants ont été tirés de la BPE. La population se trouvant ni à son lieu de résidence ni à son lieu de travail a été répartie entre commerces, restaurants et rues passantes (carreaux des sections cadastrales figurant les artères principales et pondérés en fonction du poids des commerces présents dans un petit disque de rayon 800 mètres).



Exemple : localisation des restaurants (en noir) et des sections cadastrales (en rouge) dans la commune de Nantes.

4 - Les résultats

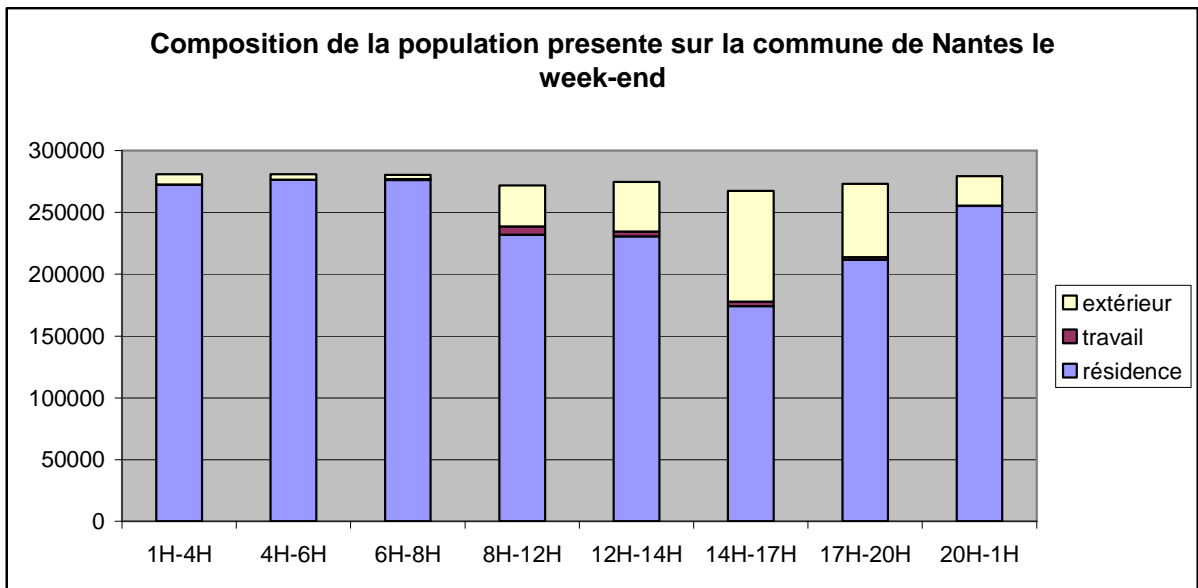
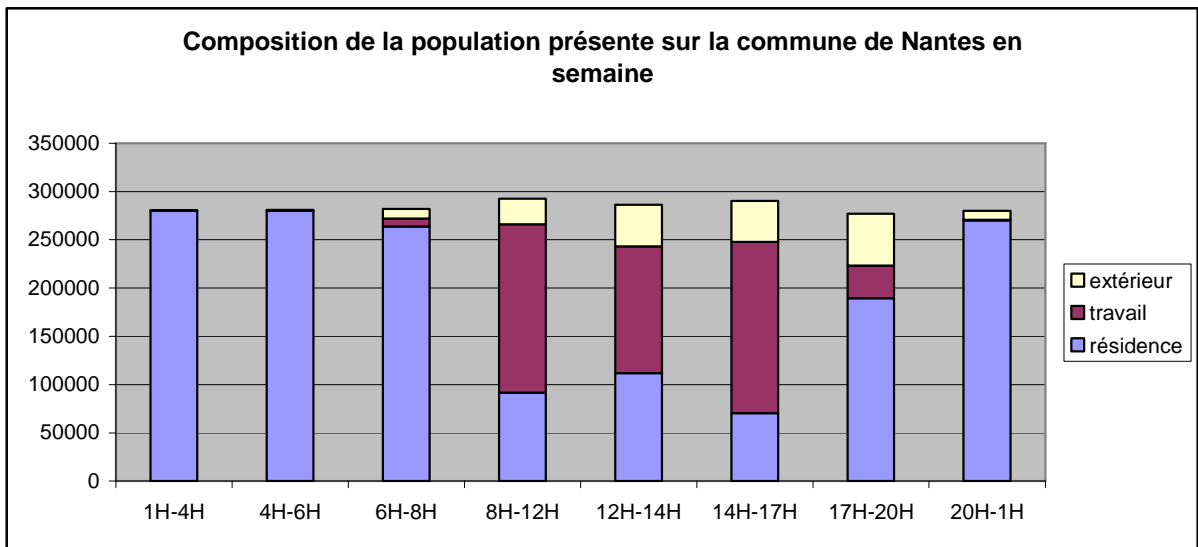
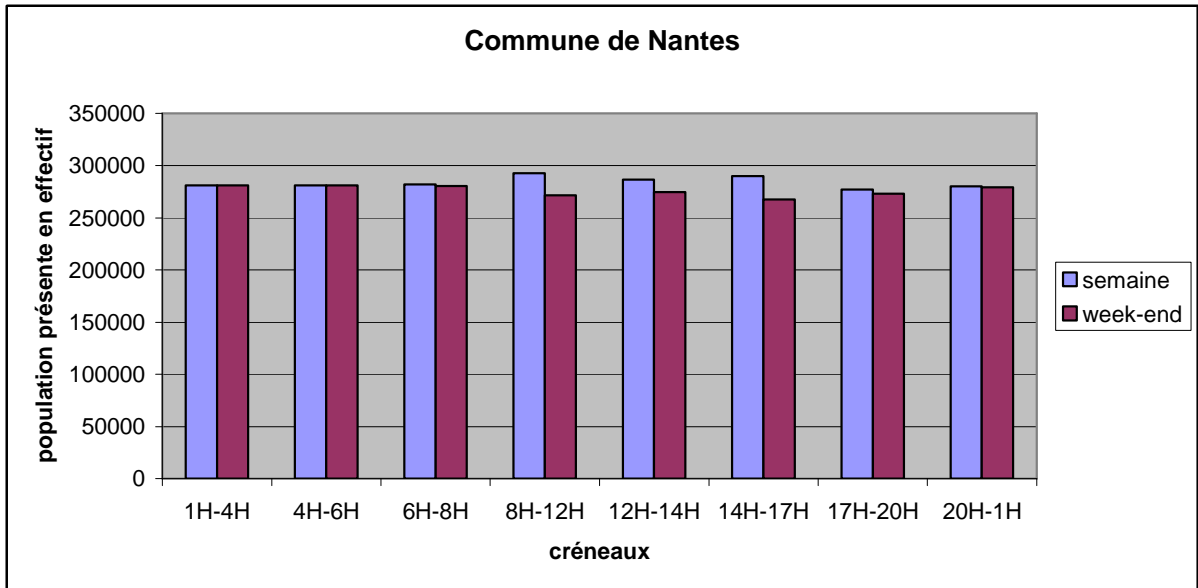
4.1 - Le choix de la maille de résolution

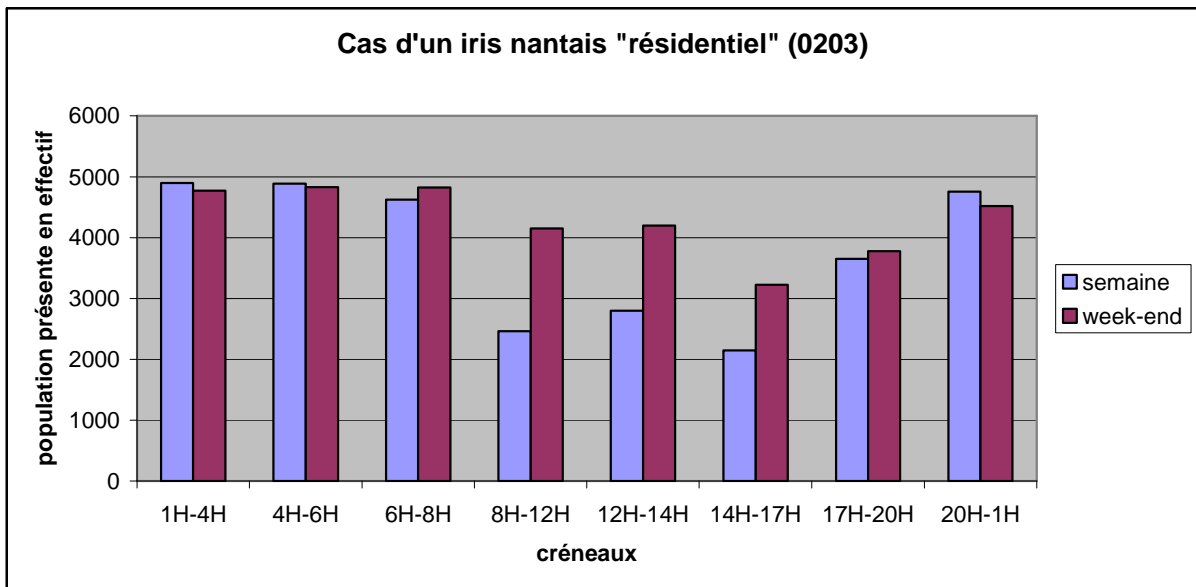
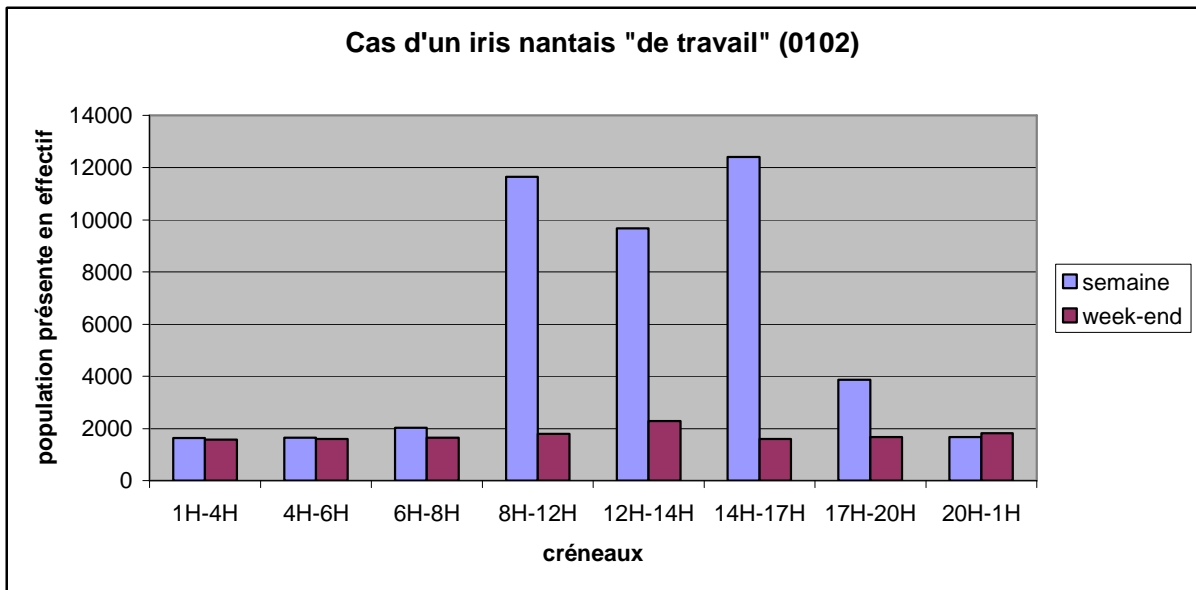
Le choix de la maille de résolution a résulté d'un arbitrage entre variabilité des cartes d'une part et protection de la confidentialité et précision des résultats d'autre part

Une maille de 100 mètres a finalement été retenue. En périphérie des villes il faudrait sans doute une maille plus grande (500 mètres par exemple) mais l'implémentation d'une maille variable selon que l'on est dans l'urbain, le périurbain ou le rural a été jugé techniquement trop complexe à mettre en oeuvre.

4.2 - Des chiffrages par agglomération/communes/iris ou zone à façon

Afin d'avoir un aperçu de l'évolution quotidienne de la population présente au sein de la commune ou encore au sein d'un iris précis, on présente ci-dessous quelques chiffrages obtenus par le kit :

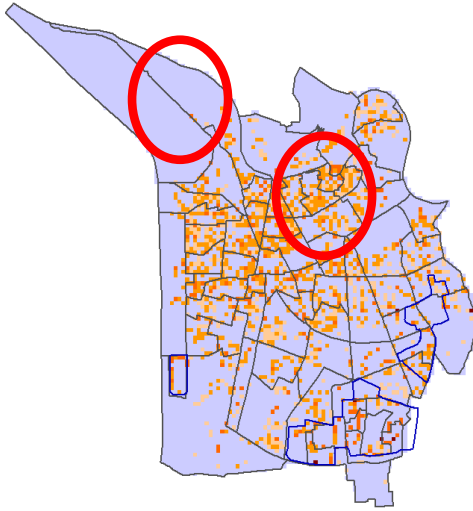




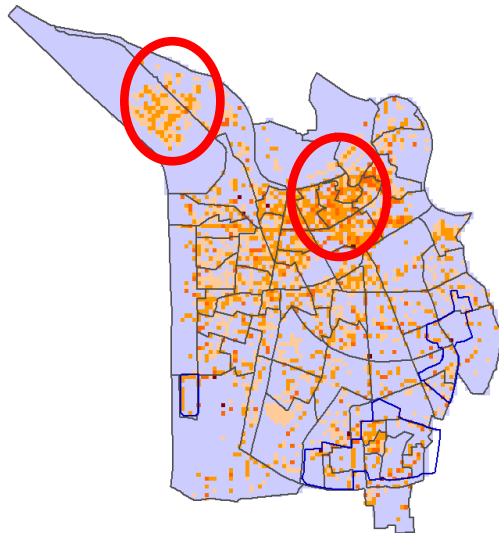
4.3 - cartes carroyées en niveau

Ci-dessous figurent trois cartes carroyées en niveau obtenues sur la commune de Grenoble la nuit, au créneau de travail du matin ou encore à la pause déjeuner.

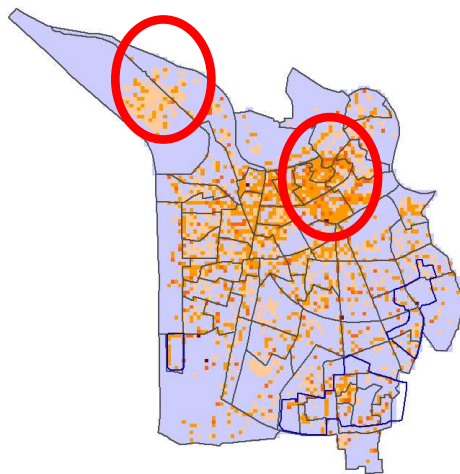
1H-4H
en semaine



8H-12H
en semaine



12H-14H
en semaine

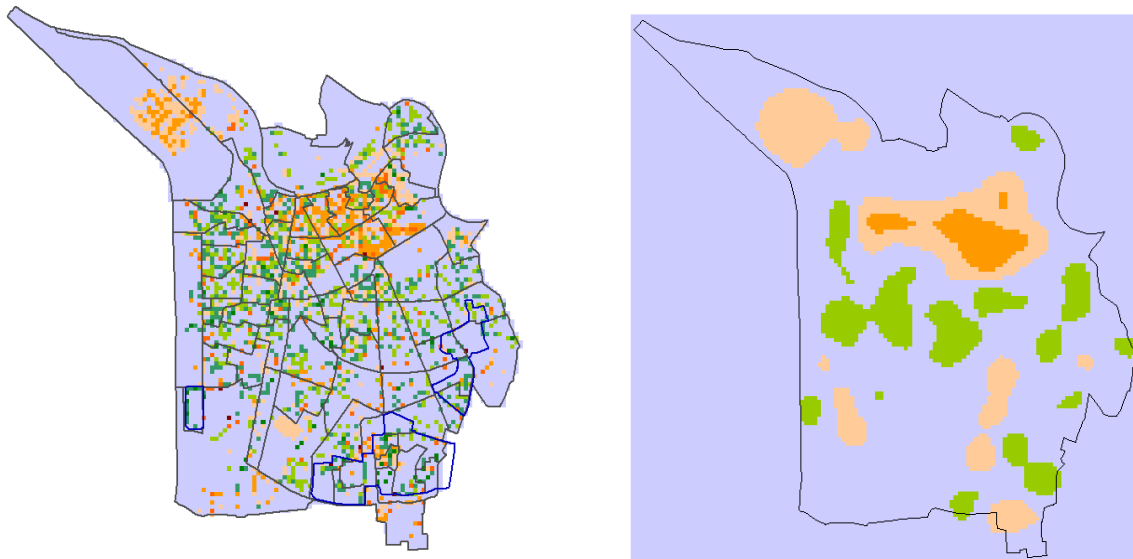


4.4 - Des cartes en différence carroyées

On peut prolonger l'utilisation de ces cartes en réalisant des cartes complémentaires dites "en différence" qui mesurent carreau par carreau les différences de population présente pour un même créneau et à deux périodes distinctes ou pour deux créneaux à une même période.

On opposera en particulier ainsi population de jour et de nuit, mais la richesse des cartes pouvant être produites est bien plus importante du fait des nombreuses combinaisons période/créneaux qui peuvent être sélectionnées.

On a représenté ci-dessous la différence S4-S1 entre population de jour (présente au créneau 4 soit 8H-12H) de la journée et population de nuit (présente au créneau 1 soit 1H-4H). On propose deux cartes, l'une simple différence des cartes carroyées, la seconde dans une version lissée car l'information significative réside plus dans l'analyse de l'ensemble d'une zone (reçoit-elle de la population, se dépeuple-t-elle et à quels créneaux ?) que carreau par carreau.



On constate la convergence de la population vers deux points focaux principaux : le centre historique (en orange foncé) ainsi qu'un pôle secondaire dans la pointe Nord-Ouest de Grenoble. Certaines « poches » en revanche sont structurellement déficitaires pendant les horaires de travail et apparaissent en vert.

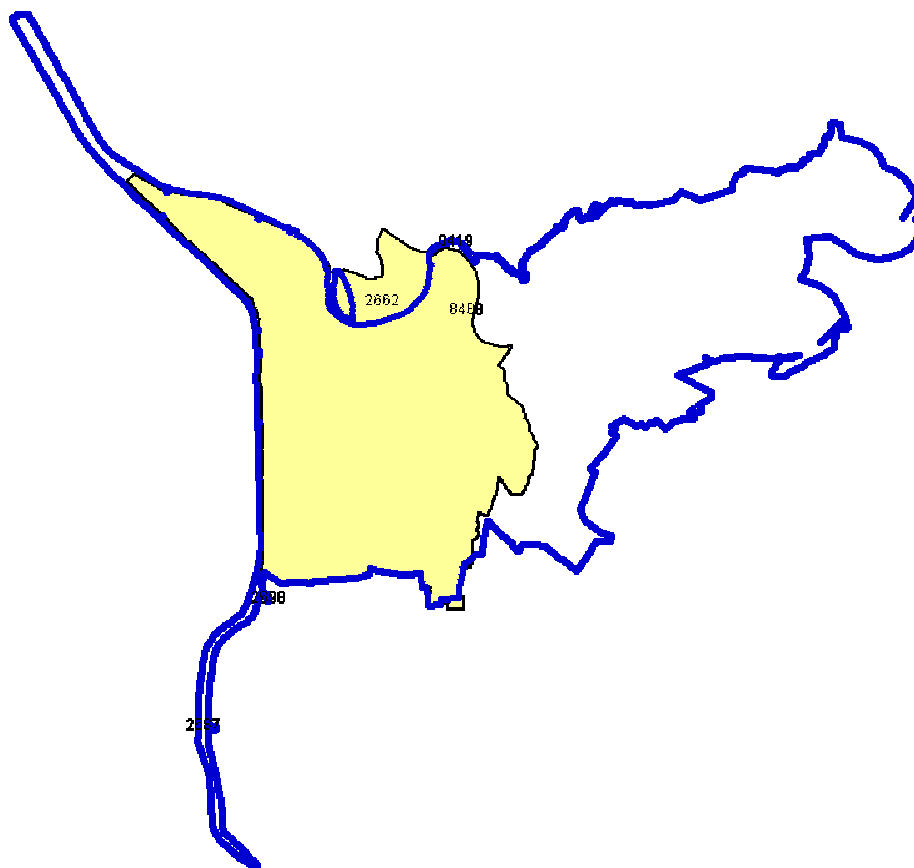
4.5. Précision des résultats

La précision théorique des résultats fournis par le kit se décompose en trois « composantes » à savoir : la précision liée au recensement, la précision liée à la modélisation économétrique et aux hypothèses très fortes formulées pour la population non couverte par cette modélisation, et enfin aux erreurs qui entachent les différentes clés de répartitions spatiales (mauvaise prise en compte des intérimaires et problème de dégroupement dans clap notamment, mais également clé assez sommaire pour le positionnement à l'extérieur).

Dans ce cadre seule la précision liée au recensement peut raisonnablement être étudiée, et ce grâce à la macro élaborée par G. Chauvet permettant d'estimer la précision sur une zone à façon. Elle permettra de donner une largeur « minimale » aux intervalles de confiance et donc de poser des seuils d'alertes sur certains effectifs insuffisants de population présente pour lesquelles tout commentaire sera à proscrire.

On l'applique aux variables de modélisation des comportements individuels au niveau adresse (précisément les parts calées des individus restées au lieu de résidence).

- o Application sur les zones inondables de la ville de Grenoble :



Les zones inondables sur Grenoble comprennent une petite zone au nord de l'Isère (notée 2662 dans ce qui suit), une grande au sud (notée 8488).

	zaf	ESTIM	CV
p_semaine_res_1_cal	2662	310	3,7
p_semaine_res_2_cal	2662	309	3,72
p_semaine_res_3_cal	2662	284	3,86
p_semaine_res_4_cal	2662	98	3,38
p_semaine_res_5_cal	2662	125	2,25
p_semaine_res_6_cal	2662	71	3,26
p_semaine_res_7_cal	2662	175	2,97
p_semaine_res_8_cal	2662	292	3,44
p_semaine_res_1_cal	8488	131657	0,27
p_semaine_res_2_cal	8488	131407	0,27
p_semaine_res_3_cal	8488	122492	0,27
p_semaine_res_4_cal	8488	50042	0,35
p_semaine_res_5_cal	8488	60698	0,34
p_semaine_res_6_cal	8488	37746	0,4
p_semaine_res_7_cal	8488	79485	0,29
p_semaine_res_8_cal	8488	125358	0,27

On constate que la précision liée à l'échantillonnage du recensement au niveau de la petite boucle 2662 est déjà trop faible pour que les résultats produits sur cette zone à façon puissent être considérés comme robustes.

En revanche les résultats sont satisfaisants pour l'autre zone.

Rappelons toutefois que seul le volet « précision de la population au lieu de résidence » est ici estimé.

5 - Les principales limites de l'exercice

Tout d'abord l'une des principales limites inhérente à la méthode est l'absence de prise en compte systématique de la population touristique, ce qui dans certaines zones peut fortement sous-estimer la population présente totale. Cette prise en compte est en effet malheureusement conditionnelle à la disponibilité locale de données tant en volume de la population touristique que de clés de répartitions ad hoc.

Par ailleurs, les personnes réalisant des navettes domicile-travail sur longues distances et qui de ce fait ne peuvent revenir chez elles que le week-end y sont placés de façon factice y compris sur cette période.

Les clés de répartition sont nécessairement entachées d'erreurs (problème de dégroupement de clap notamment, c'est-à-dire que certains salariés sont arbitrairement rattachés par exemple au centre de gestion administratif de l'entreprise et non à leur lieu de travail réel), ce qui accorde plus de poids en terme d'emplois à une localisation qu'à une autre de façon fallacieuse.

L'absence de prise en compte dans la segmentation réalisée d'une variable d'activité empêche de produire des chiffrages de population travaillant de nuit. Toutefois il aurait fallu pour cela travailler sur un échantillon beaucoup plus important.

La question du volume de population dans les transports n'a pas pu être abordée spécifiquement. La localisation alternative aux résidences et aux lieux de travail ou d'étude a volontairement été les sections cadastrales, présumées épouser relativement bien les grands axes y compris souterrains.

6 - Perspectives

L'une des limites tenait à l'ancienneté de l'enquête emploi du temps mobilisé : la mise à jour des modèles sera réalisée prochainement grâce à la nouvelle enquête de 2010.

Bibliographie

[1] Mille M., « Des densités habitantes aux densités mouvantes : l'exemple de la métropole lilloise », *Cybergeo*, 2000 (<http://cybergeo.revues.org/1866?lang=en>)

[2] Rakotomalala R., « Courbe Roc, Receiving Operating Characteristics, un autre moyen d'évaluer un modèle de prédiction », *Laboratoire ERIC*

[3] Klein O., « Modélisation et représentations spatio-temporelles des déplacements quotidiens urbains, application à l'aire urbaine Belfort-Montbéliard », 11 juin 2007

[4] Laurent Jardinier, Cete de Lyon, Régis de Solere, Certu, « Les pulsations urbaines, une analyse originale des enquêtes ménages déplacement », 3 février 2011

Webographie

[5] http://geodepot.statcan.ca/Diss/Maps/ThematicMaps/placeofwork_f.cfm

[6] http://www.espacedestemps.grandlyon.com/Diagnostiquer/diagnostiquer_chrono_gerland.htm

Annexe : valeurs obtenues pour l'indicateur ROC par les 48 modèles

Tableau récapitulatif sur les critères de qualité des 48 modèles :

Indicateur Roc pour les 48 modèles

Modélisation de la probabilité de fréquenter ...

	le lieu de résidence en semaine	le lieu de résidence le week-end	le lieu de travail en semaine	le lieu de travail le week-end	l'extérieur en semaine	l'extérieur le week-end
1H-4H	0,688	0,781	0,802	0,876	0,717	0,77
4H-6H	0,679	0,747	0,86	0,861	0,79	0,765
6H-8H	0,778	0,72	0,807	0,796	0,738	0,637
8H-12H	0,8	0,717	0,837	0,79	0,622	0,615
12H-14H	0,768	0,642	0,824	0,811	0,657	0,623
14H-17H	0,764	0,619	0,83	0,829	0,613	0,574
17H-20H	0,721	0,651	0,815	0,835	0,675	0,598
20H-1H	0,719	0,78	0,764	0,824	0,699	0,698

Note de lecture : les deux-tiers des modèles présentent un indicateur Roc qui peut être considéré comme très largement satisfaisant (supérieur à 0.7). Seulement deux modèles ont obtenus un indicateur ROC inférieur à 0,6, ce qui n'a pas semblé invalidant pour avoir une perception globale des fluctuations journalières de population assises sur ces modèles.