

Génération des émissions et attractions à l'échelle des communes

Nous avons étudié jusqu'ici les différentes possibilités en génération de trafic. Il s'agit maintenant de faire le choix de méthodologies adaptées à notre territoire d'étude, aux objectifs du modèle de génération, aux objectifs plus généraux du modèle à quatre étapes et aux données disponibles.

La présentation de la construction des données, de l'accès que nous en avons eu ainsi que de l'extrapolation de certaines informations est nécessaire pour comprendre les choix méthodologiques que nous avons réalisés. Cette présentation permettra également d'explicitier clairement les positions que nous avons choisies quant à la modélisation effectuée par la suite. Enfin, nous présenterons les résultats en terme d'émissions et attractions des déplacements à l'échelle communale du modèle ainsi mis en place.

1. Les données disponibles

Différentes sources de données sont accessibles concernant la mobilité en France. Elles sont principalement contenues dans la base de données de l'Insee *Communes mobilité*. Ces données ne sont toutefois pas exhaustives et ne présentent qu'une partie restreinte de l'information.

Afin d'approcher au plus près la mobilité et sa quantification à un instant de référence, différentes techniques ont été employées : Une enquête axée sur les transports en commun a tout d'abord été réalisée ; à laquelle sont venues s'ajouter les données de l'enquête ménages – déplacement de la Communauté d'agglomération du Grand Besançon (CAGB) qui nous ont permis d'extrapoler les données inhérentes au déplacement en voiture personnelle sur l'ensemble de la région de Franche-Comté. Nous allons présenter ici les différents protocoles, d'enquête et d'extrapolation, qui nous ont fourni les données nécessaires à la modélisation, après avoir expliqué les objectifs du modèle qui ont conditionnés nos choix.

1.1. Objectifs du modèle

L'objectif du modèle de génération de trafic est, en premier lieu, de pouvoir prévoir, de la manière la plus fine possible, l'évolution de la mobilité à court et moyen terme. Prévoir à long terme de ce genre de phénomène s'avère être un exercice périlleux, de par la difficulté d'estimation des différents paramètres liés à la mobilité quotidienne.

Remis dans le cadre plus complet du modèle à quatre étapes, le modèle de génération de trafic a également pour but de pouvoir estimer l'impact d'une modification de l'offre de transport sur les comportements de mobilité des individus. Cette problématique n'est au demeurant qu'assez peu liée à l'étape spécifique de la génération de trafic. Nous considérons en effet ici que l'individu fait le choix d'effectuer ou non un déplacement sans prendre en compte les potentialités liées à l'offre de transport. Il va de soi que la réalité du processus décisionnel est plus complexe et que les informations liées aux possibilités modales influencent la décision de mobilité. Le modèle de génération de trafic étant à terme inclus dans un modèle à quatre étapes, la séparation des étapes reste une contrainte forte, comme nous l'avons vu dans le premier chapitre. Les postulats de départ du modèle à quatre étapes nous obligent donc à ignorer les phénomènes d'induction qui sont responsables de l'augmentation ou la diminution de la mobilité à la suite de mise en service de nouvelles prestations ou de nouvelles infrastructures. Les inactifs motorisés

connaissent par exemple une forte évolution de leurs déplacements, spécifiquement pour des motifs non contraints, en fonction de l'évolution de l'offre de transport en commun [Bonnell, 1995a].

La modélisation développée ici est donc une modélisation à but prévisionnel, chargée de quantifier, à un instant $t+1$, les émissions et attractions de zones géographiques préétablies. Ces émissions et attractions pourront par la suite, au cours des autres étapes du modèle, permettre de tester les incidences de politiques spécifiques de transport.

1.2. Les données disponibles

1.2.1. Les données « Communes mobilité »

Les données *communes mobilité* ne couvrent qu'une partie des motifs de déplacement. Ces données sont réparties selon les motifs travail et études et fournissent diverses informations attachées à cette mobilité (annexe 1). On connaît notamment la répartition modale de cette mobilité. Même si la question du mode n'est pas abordée dans notre modélisation, elle reste importante dans la quantification des déplacements. En effet, si on connaît l'ensemble des déplacements concernant les deux motifs détaillés ici, nous allons voir par la suite que les diverses enquêtes et données dont nous disposons ne nous permettent pas d'accéder aux données concernant la mobilité en voiture personnelle attachée à certains motifs.

1.2.2. L'enquête TER Franche-Comté

La construction de l'enquête

Dans le cadre de la modélisation de la mobilité régionale, une enquête a été réalisée durant l'automne 2004. Cette enquête a été effectuée sur l'ensemble des services de transport en commun (cars et trains) à la charge de la Région de Franche-Comté. Plusieurs réflexions ont été menées à cette occasion sur la qualité de l'information à collecter et donc sur le fond et la forme de cette enquête.

Le premier point à soulever, conditionnant l'ensemble des autres interrogations concernant l'enquête, porte sur le choix entre un questionnaire auto administré

ou avec l'intervention d'un enquêteur. L'avantage certain de l'intervention d'un enquêteur réside dans la compréhension de l'enquête par l'individu enquêté. Comme nous l'avons déjà vu, l'interprétation de l'enquête peut fournir un biais important sur les résultats. Mais ces considérations ne constituent pas les seuls critères de choix relatifs à la voie d'administration du questionnaire. Le coût tout d'abord, plus important dans le cas d'un entretien dirigé, pousse à préférer, sur une enquête de cette ampleur, la solution d'un questionnaire que les usagers des transports en commun peuvent remplir seuls. Le choix ici d'une enquête exhaustive amène donc, dans tous les cas, à préférer un questionnaire auto administré. De plus, vouloir cerner l'ensemble des usagers des transports en commun amène à considérer également la longueur du questionnaire. Beaucoup de trajets étant réalisés sur de petites distances, et donc des laps de temps courts, il est nécessaire que le questionnaire soit conçu avec un nombre de questions limitées et que son remplissage se fasse avec une certaine rapidité (annexe 2).

Il est possible, au demeurant, que bien que conçu pour l'ensemble des usagers, tous les individus concernés par l'enquête ne fassent pas le choix de remplir le questionnaire distribué. Un comptage aux arrêts des différentes montées et descentes est donc nécessaire pour d'éventuels redressements statistiques.

Concernant le comptage des déplacements, la réalisation du questionnaire nous amène à nous interroger sur la définition retenue pour ces déplacements. Nous reviendrons par la suite sur la période et la durée de l'enquête, mais quelque soit la durée choisie, il est possible d'enquêter plusieurs fois le même individu, se situant dans une chaîne de déplacements ou une sortie. Il est alors important de pouvoir situer l'enquête dans cette chaîne afin de ne pas comptabiliser plusieurs fois les caractéristiques liées à un aller-retour. De même, dans le cas d'enquête réalisée sur plusieurs jours, il est nécessaire de savoir si l'enquêté a déjà ou non été pris en compte lors des précédentes journées. Le choix a été fait ici de quantifier le nombre de réponses déjà données par voie de questionnaire. Ainsi, par croisement avec les profils des individus et les motifs de déplacement, les allers-retours pourront être pris en compte, ainsi que les trajets répétés d'une journée sur l'autre.

Vient ensuite la question liée à l'unité de mesure des migrants. Même si le profil de l'individu enquêté peut être remis dans le cadre de son ménage assez aisément par voie de questionnaire, qualifier la mobilité totale de ce ménage paraît être difficile. L'enquête réalisée ici est donc individu centré et ne considère que les caractéristiques liées à l'individu remplissant le

questionnaire. De plus, le but principal du modèle de génération de trafic étant la prévision, travailler à l'échelle de l'individu paraît être la solution la plus judicieuse au vu des avantages et inconvénients expliqués au début de ce chapitre.

Concernant une enquête dont le but est de quantifier des déplacements, le choix de la période et de la durée d'enquête s'avère être de première importance. Toute période de vacances scolaires est à exclure dès le départ : la mobilité quotidienne se trouve radicalement modifiée à ces périodes, autant pour les populations scolaires que pour les actifs. On pourrait se poser la question de la saison de l'enquête. La mobilité paraît naturellement être différente selon la météo et la température extérieure. Toutefois, les moyens financiers et temporels ne nous permettant pas de réaliser une enquête parfaite dans la théorie, nous ferons le postulat d'une relative constance de la mobilité au cours de l'année, hors périodes de vacances. Le mois d'octobre a finalement été retenu.

A propos de la durée, toute échelle temporelle d'enquête inférieure à une journée a été écartée. La mobilité, qu'il s'agisse de mobilité contrainte ou non, s'inscrit traditionnellement dans une période journalière. Pour autant, tous les jours de la semaine ne se ressemblent pas en terme de quantité et de nature de déplacements. Hormis les jours de week-end qui sont radicalement différents du reste de la semaine, les débuts et fins de semaine « active » (à savoir les lundi et vendredi) proposent également des différences notables avec les jours de milieu de semaine. Il est donc important de pouvoir avoir un aperçu de l'ensemble de ces plages temporelles. La Région de Franche-Comté, financeur de cette enquête, a ici fait le choix d'enquêter le lundi et le jeudi de plusieurs semaines consécutives. On retrouve ici certains réflexes hérités des objectifs passés de la modélisation de trafic qui vise à quantifier les déplacements aux instants où le trafic est le plus dense afin de dimensionner au mieux l'offre de transport et les infrastructures. Adopter de tels choix d'enquête nous amène à renoncer à comprendre la mobilité marginale du week-end au profit d'un gain relatif de précision sur les jours présentant les plus gros contingents d'usagers. La question peut également se poser d'enquêter sur la totalité d'une semaine plutôt que sur deux jours choisis comme référents sur plusieurs semaines consécutives. Etaler l'enquête sur plusieurs semaines offre l'opportunité de lisser l'impact d'évènements ponctuels comme certaines manifestations commerciales ou culturelles, certains retards sur divers services ou encore une météo spécifique. Toutefois, faire le choix de jours référents, -choix qui se

justifie pleinement dans le cas où l'enquête est limitée à deux jours maximum pour des raisons financières ou techniques-, amène une approximation certaine pour les autres jours de la semaine. Le mercredi par exemple, spécifique par les horaires des scolaires auxquels s'ajoutent souvent des emplois du temps aménagés de certains actifs ayant des enfants à charge, n'est que mal pris en compte. De même, le vendredi, avec les modifications évoquées en première partie en lien avec les transformations des emplois du temps, n'est plus tout à fait comparable au lundi.

Enfin, la question de la perception de l'individu est tout à fait intéressante à traiter dans le cadre d'une enquête de cette ampleur. Nous parlons ici de la perception intrinsèque de la qualité de certains aspects du service comme de leur importance.

Au-delà de la modélisation de la génération de trafic, que ce soit pour l'étape de répartition modale ou celle de l'affectation de trafic, la perception par chaque individu du coût, de la rapidité, du confort et de l'écologie relatifs à son déplacement influence les choix qu'il réalise en amont ou pendant le déplacement. De plus, concernant ce type de données, il est difficile de se référer à des données existantes. Les mentalités liées à ces questions évoluent rapidement dans le temps et ne sont que rarement comparables d'un territoire à un autre. On peut donc regretter que pour raccourcir au maximum le questionnaire la Région de Franche-Comté ait écarté toute question se référant à l'individu sous l'angle de la perception de son déplacement.

Toutefois, ces données existent sous une autre forme, à travers des enquêtes « satisfaction » réalisées chaque année par la SNCF sur des échantillons de population. Réalisées par voie téléphonique, ces enquêtes détaillent la perception qu'a l'individu des critères de rapidité, de confort et de coût sur son trajet. Nous n'avons hélas pas pu avoir accès à ces données.

Premiers traitements concernant l'enquête : nettoyage de la base de données

Les premiers traitements concernant l'enquête nous ont révélé certaines invraisemblances et nous avons rapidement constaté qu'une certaine proportion des vingt mille enquêtes réalisées contenait des réponses erronées.

La première étape a consisté à évaluer le degré de cohérence de chaque questionnaire pour extraire les réponses risquant de fausser l'analyse. Trois critères ont été retenus pour faire émerger divers degrés d'incohérence :

- Le nombre de réponses nous a fourni un premier élément d'analyse. Nous avons considéré que sur huit jours d'enquête, le nombre de fois qu'un individu peut répondre à l'enquête est limité. Nous avons donc appliqué une note dégressive lorsque ce nombre augmentait avec un seuil de départ fixé à seize, soit deux réponses par jours. Il est toujours possible, en théorie, qu'un individu ait été amené à répondre un plus grand nombre de fois au questionnaire. Il nous paraît cependant probable que toute réponse supérieure à ce chiffre soit le reflet de réponses erronées.
- L'âge ensuite, nous a permis d'écarter certains questionnaires. Tout âge inférieur à sept ans a été évalué comme présentant un degré d'incohérence.
- Enfin, le croisement des informations du profil de l'individu avec le nombre de fois où le questionnaire a déjà été rempli nous a permis d'affiner encore notre analyse : A profil constant, le nombre de réponses déjà fourni doit présenter une continuité. Un individu ne peut pas avoir rempli 9 fois le questionnaire sans l'avoir rempli une septième et une huitième fois. Nous avons toutefois considéré qu'une erreur de comptage de la part de l'enquêté était toujours possible. Nous avons donc noté de manière dégressive les réponses suivant la taille de l'erreur constatée.

Nous avons par la suite agrégé ces différents indices d'incohérence pour obtenir une évaluation de chaque questionnaire d'enquête. Plus de 1500 enquêtes sur les 19928 de départ ont révélé au moins deux facteurs d'incohérence simultanés. Cette proportion assez forte (plus de 7,5 % de réponses erronées) nous a amené à pousser plus loin cette analyse et à tenter de comprendre les causes de ces invraisemblances. Il n'est pas possible d'écarter l'ensemble de ces questionnaires nous paraissant présenter des réponses fausses sans en comprendre la raison.

Notre raisonnement intuitif nous a tout d'abord amené à analyser la répartition de ces réponses dans le temps en fonction de la représentation des catégories socioprofessionnelles aux différents moments de la journée. La figure 22 présente la répartition de ces groupes en fonction du temps avec le degré moyen de cohérence des enquêtes par heure.

La répartition des scolaires paraît expliquer, en premier lieu, au moins en partie la proportion de réponses erronées : plus la proportion des scolaires augmente, plus le nombre de réponses recelant des degrés d'incohérence semble croître.

Cependant, cette proportion de scolaires est elle-même fortement corrélée avec d'autres facteurs.

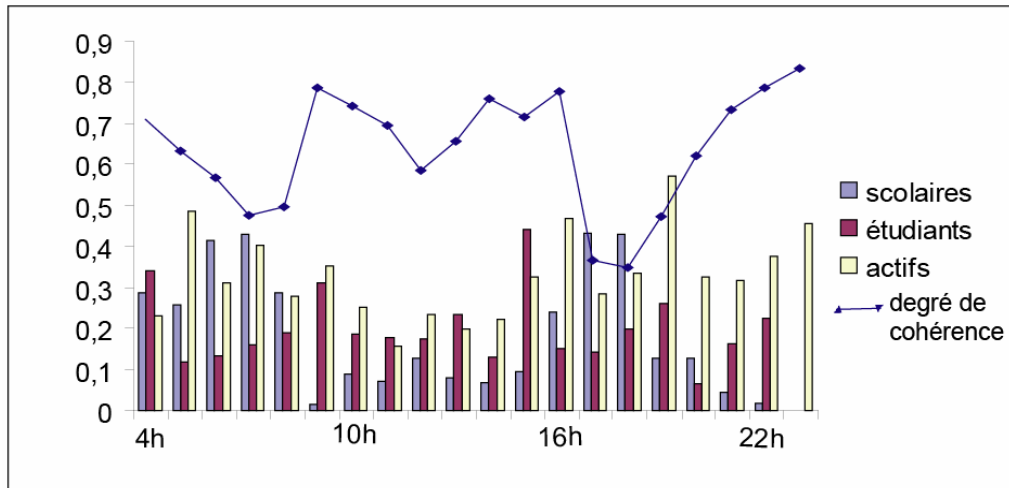


Figure 22 : Les principales catégories sociales et les réponses erronées aux enquêtes TER

Source : enquête TER Franche-Comté

La figure 23 montre par exemple la fréquentation des services TER aux différentes heures de la journée.

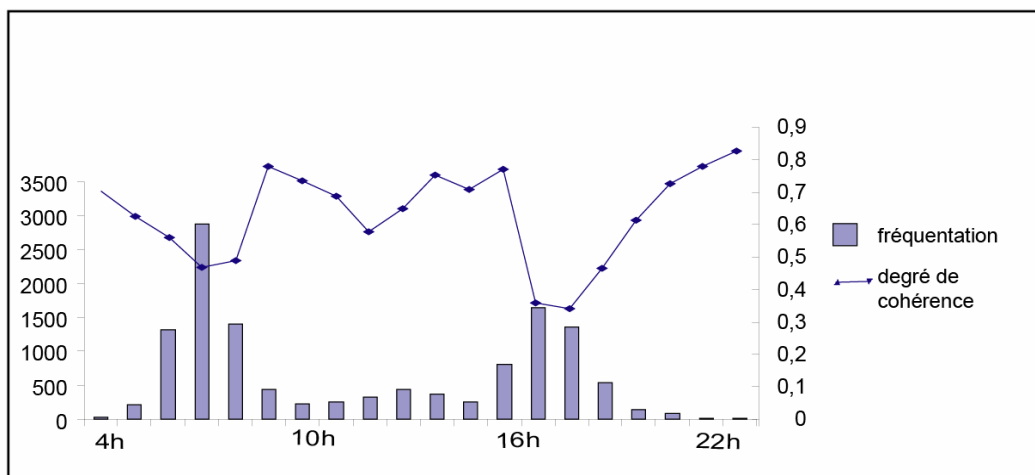


Figure 23 : Cohérences de réponses et fréquentation du service enquêté

Source : enquête TER Franche-Comté

On constate également ici une relation forte entre les pics de fréquentation et le pourcentage de réponses erronées. On remarque tout de même une proportion plus forte de réponses présentant des incohérences pour le pic de fin de journée,

pourtant moins important en terme de fréquentation que le début de la journée. Il est donc possible que la proportion de scolaires ne soit que peu liée au phénomène de réponses erronées et que la fréquentation explique cette manifestation d'erreurs.

L'analyse segmentée des lundi et jeudi tend effectivement à confirmer cette théorie. La figure 24 développe cette idée en mettant en évidence les degrés de cohérence de chacun des jours avec la fréquentation qui lui est affiliée.

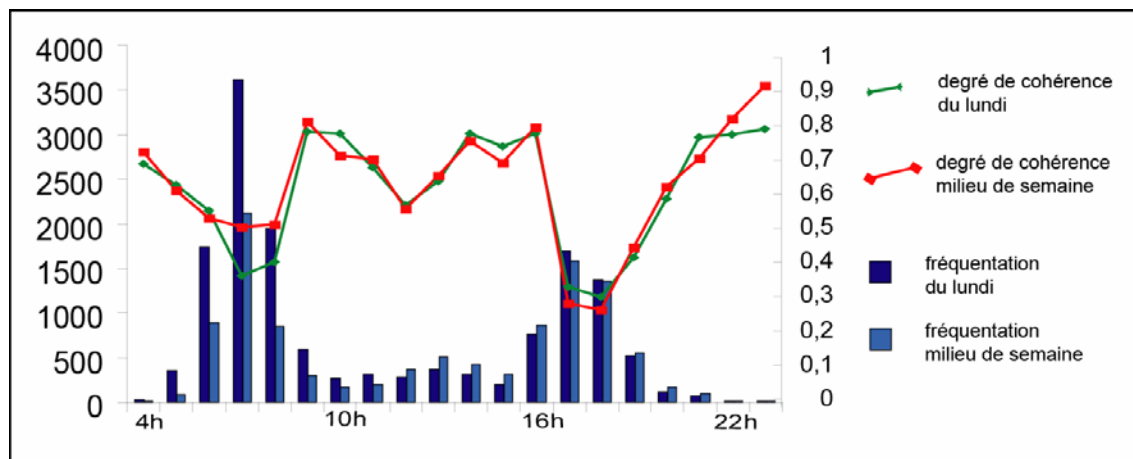


Figure 24 : analyse comparative des fréquentation et cohérence des réponses de début et milieu de semaine

Source : enquête TER de Franche-Comté

On voit effectivement le lundi, pour lequel la fréquentation est plus forte en début de journée, présenter un plus fort taux de questionnaires incohérents à cette même heure. Il est néanmoins difficile de déterminer avec précision si ce phénomène est dû ou non à la fréquentation ou à la proportion de scolaires, puisque cette dernière est également plus importante le lundi.

On peut raisonnablement penser que les populations scolaires ont une plus forte tendance, de par leur jeune âge, à fournir des réponses erronées ; mais qu'une forte fréquentation favorise ce phénomène, en couvrant plus ou moins ce genre de comportement dans la foule. On imagine en effet plus volontiers remplir le questionnaire de manière apocryphe lorsqu'on le rend en même temps que trente autres individus que lorsqu'on le rend seul à l'enquêteur.

Un autre phénomène vient également s'ajouter à ceux-ci. L'avancée du temps au cours de la journée semble correspondre à une croissance progressive du nombre de réponses incohérentes. La figure 25 est construite en pondérant le nombre de réponse par la fréquentation.

On voit nettement la qualité des réponses se dégrader avec le temps et le pic de fin de journée reparaître. La fatigue peut certainement expliquer en grande partie ce phénomène.

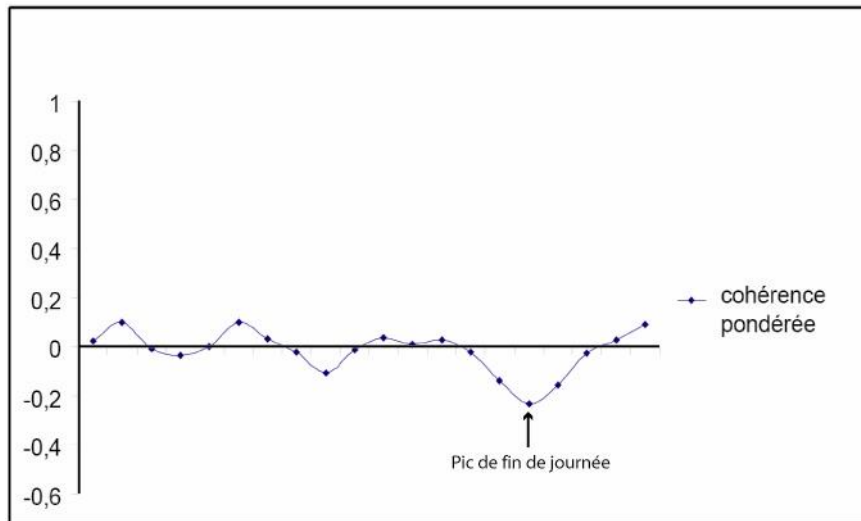


Figure 25 : Evolution du degré de cohérence moyen pondéré par la fréquentation au cours de la journée

Source : Enquête TER Franche-Comté

Nous avons donc établi une base de données fiable sur les transports en commun sur l'ensemble de la région, couplée à la base *communes mobilité* de l'Insee. Manque alors à ce recueil l'ensemble des informations de déplacements liés aux motifs différents du travail ou des études non contenus dans la base de l'Insee et réalisés par un autre mode que le transport en commun, ces derniers étant présents dans la base issue de l'enquête TER de la Région. Ces données n'existant pas en l'état, nous avons dû les construire par conjecture à partir d'un territoire test.

2.1.3. La construction des données manquantes

Nous avons vu, au cours de ce chapitre, le fonctionnement des modèles normatifs. Ces modèles consistent à appliquer à un territoire d'étude, des normes obtenues à partir de calculs sur un territoire référent.

Ne disposant pas d'une norme récente sur un territoire comparable au nôtre, nous avons pris le parti de concevoir une norme des attractions et émissions par commune sur le territoire référent de la Communauté d'Agglomération du Grand Besançon (CAGB). L'enquête ménages déplacement réalisée au cours de l'année 2005 sur la CAGB nous fournit en effet la totalité des informations

nécessaires à la construction d'une telle norme de calcul. Elle pose toutefois le problème de ne pas avoir été conçue à la même date que nos autres sources de données. Nous considérerons pourtant que ce décalage temporel de recueil de données n'influence que peu les résultats obtenus sur les déplacements des individus.

Bien que nommée communauté d'agglomération, la CAGB présente un panel très large de communes que l'on peut, en terme de mobilité, classer en plusieurs catégories. Il va de soi que les communes, suivant leur profil, ne sont pas soumises aux mêmes comportements de mobilité. Il convient donc d'établir une typologie afin de pouvoir construire une norme différente pour chaque profil ainsi dégagé. Nous pourrions par la suite appliquer à chaque commune, suivant le type auquel elle appartient, le comportement référent.

La mise en place d'une typologie de communes

La typologie de commune qu'il convient d'établir ici doit se baser sur la mobilité quotidienne et plus précisément sur les déplacements liés aux motifs autres que travail ou études.

Pour cela, nous avons dans un premier temps dégagé l'ensemble des variables explicatives de la mobilité quotidienne liée aux motifs nous concernant ici en nous basant sur la littérature (tableau 5)¹⁵. Puis, à l'aide d'une analyse factorielle (tableau 10), nous avons dégagé deux axes factoriels nous permettant de situer les communes de notre territoire dans un espace à deux dimensions.

Composante	Valeurs propres initiales			Sommes des carrés chargées		
	Total	% de la variance ==	% cumulés	Total	% de la variance ==	% cumulés
1	5,954	76,931	76,931	6,954	86,931	86,931
2	1,976	22,204	99,135			

Tableau 10 : Analyse factorielle à partir des variables explicatives de la mobilité quotidienne

Puis nous avons appliqué aux communes une méthode de groupement à partir de cet espace bidimensionnel en faisant appel à la méthode de classification autour des centres mobiles appelée également méthode des nuées dynamiques (annexe 3).

¹⁵ Rapport PREDIT, 2000, *Evaluation des modèles de prévisions de trafic*, GESMAD

Nous avons donc obtenu une classification en quatre classes. Il est difficile d'attribuer une signification à chacune de ces classes. Elles regroupent chacune des communes présentant des profils de mobilité semblables mais ne correspondent pas forcément à un profil préétabli de commune (rural, périurbain...). Chacune de ces classes étant représentée au sein de la CAGB (figure 26), il est possible de fixer une norme à partir de ces groupes.

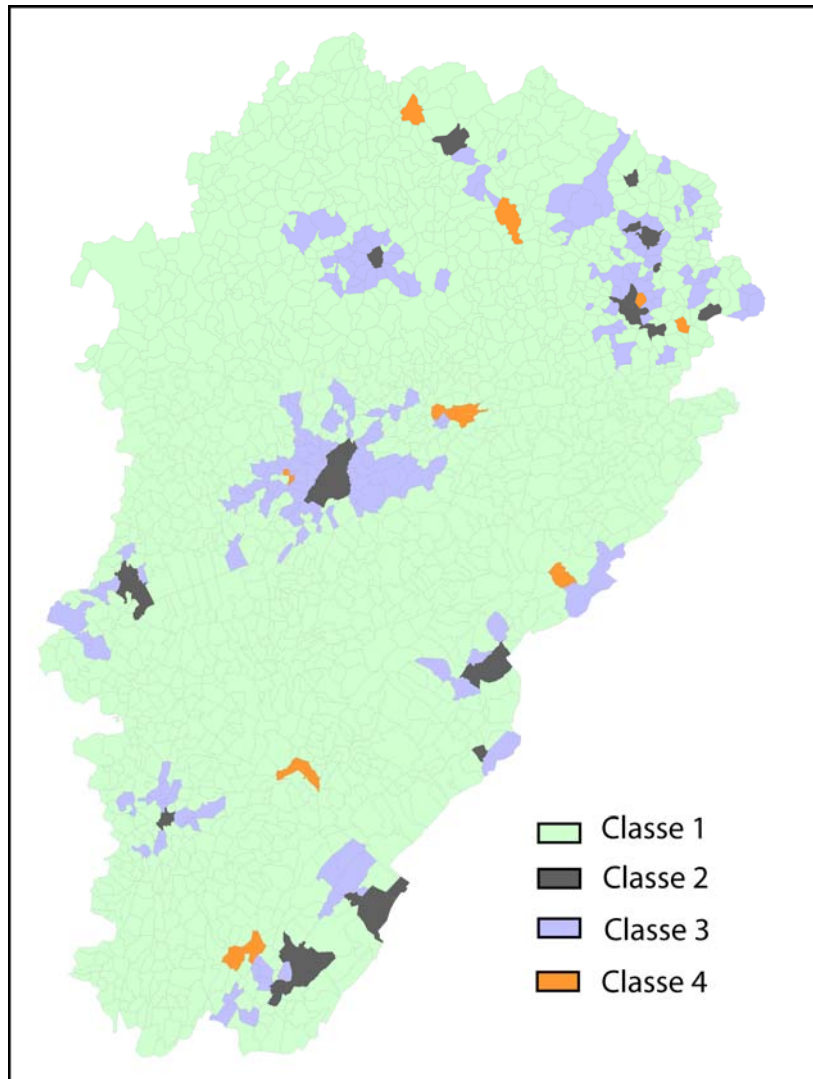


Figure 26 : Typologie des communes franc-comtoises en fonction de leur comportement de mobilité

Source : enquête TER Franche-Comté, INSEE

L'extrapolation des données

Nous avons, par la suite, calculé des profils types concernant les émissions et attractions de chaque groupe en calculant la moyenne arithmétique des émissions et attractions connues sur la CAGB.

Enfin, nous avons appliqué ces profils à chaque commune de Franche-Comté afin de conjecturer des émissions et attractions sur le territoire d'étude concernant les motifs non traités par les données de l'Insee.

On peut regretter de n'avoir pas appliqué ici une norme plus détaillée afin de prévoir au mieux les déplacements inhérents aux motifs concernés, à l'image des travaux déjà cités de Horowitz [Horowitz, 1993]. Toutefois, tenter d'approcher à un tel niveau de détail des informations pour lesquelles nous n'avons aucun moyen de vérification nous paraît être superflu. Car même si la CAGB peut fournir, faute de mieux, un territoire de référence, on ne peut certifier que les comportements de mobilité d'une Région sont strictement identiques à celui d'une portion de ce même territoire régional.

Il est pourtant nécessaire de disposer de données concernant cette part modale de la mobilité attachée aux motifs précédemment cités. Il n'est pas possible, en fait, de réaliser deux modèles parallèles, l'un à niveau de détail élevé et ne prenant en compte que les données sur les transports en commun, et l'autre plus agrégé fondé sur les données manquantes. Faire un tel choix équivaut à poser le postulat de départ que le mode est la première variable intervenant dans le processus de choix lié à la génération de trafic. On peut alors, avec cette hypothèse, accepter le fait que deux groupes de population distincts aux comportements de mobilité différents sont identifiables *a priori* en fonction de leur mode de transport. La réalité est pourtant tout autre et les expériences passées, comme nous l'avons déjà vu au cours de ce chapitre (paragraphe 1.1), nous poussent à considérer le motif du déplacement comme première variable explicative. Nous avons donc fait le choix de garder le niveau d'analyse le plus détaillé possible en conjecturant une partie de la donnée générale et en gardant à l'esprit le caractère approximatif de cette information.

2. Le cadre méthodologique

Nous disposons maintenant d'une base de données complète sur l'ensemble des émissions et attractions pour l'ensemble des modes et l'ensemble des motifs de déplacement.

Il s'agit maintenant de mettre en place une méthodologie adaptée à notre problématique afin de comprendre les émissions et attractions par motif pour pouvoir les prévoir sur chaque commune de notre aire d'étude.

Ne disposant pas de données suffisamment fines sur l'ensemble de la Franche-Comté concernant les déplacements multi motifs, nous avons pris le parti de traiter le déplacement en tant que mouvement effectué par un individu sur une voie publique, d'une commune à une autre, pour un motif principal (et potentiellement un ou plusieurs motifs secondaires), à l'aide d'un ou plusieurs modes de transport, qu'il soit basé ou non au domicile.

Par le fait, les émissions sont comptabilisées comme le mouvement de l'origine à la destination, liée au motif principal du déplacement en ne considérant que la zone émettrice. Les attractions sont comptabilisées quant à elles comme le mouvement de l'origine à la destination, liée au motif principal du déplacement en ne considérant que la zone réceptrice.

Le détail des données dont nous disposons nous autorise toutefois à mettre en place un modèle de régression afin d'approcher au plus près la réalité concernant la mobilité individuelle sur le territoire franc-comtois. Comme nous l'avons vu, le choix de cette méthodologie pose certains problèmes qu'il convient de résoudre pour la mise en place de notre protocole.

2.1. L'indépendance des individus

La première hypothèse à respecter au cours d'un tel traitement statistique est l'indépendance des individus. Comme nous l'avons déjà évoqué, les individus de notre série statistique représentés par les communes franc-comtoises ne sont pas indépendants de par leur contiguïté et leurs échanges permanents. On peut en conséquence tenter de diminuer ce biais en échantillonnant nos individus de façon à réduire la continuité du territoire. Ainsi, on amenuise les chances de contiguïté entre les communes et l'intensité de la relation entre les individus : les individus présentent de la sorte une relative indépendance.

Reste alors la question du protocole d'échantillonnage. On peut tout simplement choisir aléatoirement les communes en émettant l'hypothèse que le

hasard fournit une chance acceptable de choisir des individus indépendants les uns des autres. Une autre solution consiste à contraindre l'échantillon à des communes ne présentant pas de contiguïté et ainsi tenter de réduire au maximum les relations entre les individus. Une issue différente à ce problème est d'obliger l'échantillonnage à sélectionner le même pourcentage de communes dans chaque département ou canton afin de garantir une bonne répartition spatiale des individus. Enfin, on peut sélectionner de manière dirigée des communes dans chaque groupe de communes présentant un comportement de mobilité différents. Pour cela, il convient de réaliser, comme pour l'extrapolation de données, une typologie des communes franc-comtoises basée sur l'ensemble des variables explicatives de la mobilité quotidienne.

Cette dernière méthode permet d'éviter une sur représentation des communes appartenant au type le plus présent sur le territoire. En effet, une fois la typologie établie, on se rend compte qu'une seule classe concerne plus de 80 % des communes franc-comtoises. Cette classe concerne principalement des communes rurales caractérisées par une population faible. Cette méthode permet donc de rééquilibrer l'échantillonnage en fonction de la population représentée par chaque commune. Ainsi, on ne donne pas une importance trop grande à des communes ne représentant qu'une part faible de la population totale. Les communes représentant une part importante de la population sont évidemment beaucoup moins nombreuses que les communes à faible poids démographique.

Comme nous allons le voir maintenant, le choix de l'échantillonnage dépend également des choix effectués concernant la linéarité du modèle.

2.2. La linéarité du modèle

La seconde hypothèse est la linéarité du modèle. Cette hypothèse, ici non respectée, peut se solutionner de différentes façon comme nous l'avons déjà vu. Dans le cas de la génération de trafic, le nombre de variables explicatives pouvant être important, il est préférable d'écarter tout de suite la possibilité de réaliser des régressions non linéaires. Le modèle mathématique s'en trouverait compliqué et beaucoup plus difficile à évaluer.

Reste alors l'option du modèle catégoriel ou celle de la linéarisation par transformation des variables. Découper la population en classes pose plusieurs problèmes déjà évoqués précédemment. Seule l'analyse nous permettra donc de définir la solution optimale à la résolution de ce problème.

L'application du modèle catégoriel trouve en effet son application principale lorsque la linéarisation et la régression qui l'accompagne ne fournissent pas de solution satisfaisante. Il est alors utile de découper la population en classes afin d'affiner au mieux les régressions linéaires sur chaque classe. Il ne faut toutefois pas perdre de vue que l'objectif principal de notre modèle est la prévision, domaine pour lequel le modèle catégoriel pose des problèmes d'application.

Une simple constatation théorique nous permet de voir à quelle transformation nous devons faire appel pour la transformation des variables non linéaires. Nous sommes ici dans la configuration d'une équation de type :

$$y = \alpha x^\beta$$

Il est donc nécessaire d'utiliser le logarithme de cette fonction pour obtenir ainsi la fonction :

$$y' = \log \alpha + \beta x'$$

Avec : $y' = \log y$ et $x' = \log x$

Le logarithme d'une variable présente également l'avantage de normaliser sa distribution [Besse, 2003]. On répond mieux, de ce fait, aux hypothèses d'un modèle de régression linéaire, à savoir la normalité des variables explicatives. Ainsi, parmi les différentes possibilités d'échantillonnage étudiées auparavant, le choix aléatoire paraît justifié. Il n'existe en effet plus dans le logarithme de la variable, de valeur qui diffère fortement du reste des données. Il ne paraît donc pas nécessaire de sélectionner les individus en fonction de leur profil afin de représenter l'ensemble des spécificités présentes au sein de la population statistique. Une autre solution, concernant ces individus, appelés « points leviers »¹⁶, qui influence considérablement les estimateurs par leur excentricité, consiste à les écarter de la série statistique. Nous avons préféré ici conserver ces valeurs et réduire leur influence par la normalisation de la distribution. Nous avons toutefois fait le choix de réaliser deux échantillons parallèles, l'un contraint par les classes de population, l'autre aléatoire, afin de pouvoir comparer les résultats fournis par ces deux méthodes.

¹⁶ Nous nous référons ici à P. Besse, 2003, Modélisation statistique et apprentissage, publications du laboratoire de statistique et probabilités, Université Paul Sabatier.

2.3. L'identification des variables explicatives et la multi colinéarité

La multi colinéarité et l'identification des variables, expliquant le processus de choix de l'individu sur sa volonté de déplacement, sont deux problèmes profondément liés. Il s'agit en réalité de comprendre le lien entre l'ensemble des variables ; ce lien peut se décomposer en deux catégories, le lien entre toutes les variables (qu'elles soient explicatives ou non) et la variable à expliquer, et les liens qu'entretiennent les variables explicatives entre elles.

Les variables explicatives étant supérieures à deux quel que soit le motif traité, il est nécessaire de raisonner en terme de régression multiple. Plusieurs possibilités sont envisageables en terme de traitement statistique pour identifier les différentes variables et leur implication dans le modèle final de génération.

Nous avons opté pour une solution déjà éprouvée et offrant des aménagements possibles en fonction du but de la modélisation. De plus, la méthode employée ici nous permet de mettre en place des indicateurs de fiabilité et de significativité sans faire appel à des calculs très complexes et souvent difficilement reproductibles dans le cas de prévisions.

Nous avons donc fait appel à la régression multiple pas-à-pas. Cette démarche consiste « à recueillir un assez grand nombre de variables jugées explicatives et à choisir parmi elles l'ensemble le plus explicatif. Le but est d'obtenir une régression optimale : le plus fort R^2 pour un minimum de variables explicatives. » [Groupe Chadule, 2001]. Il existe plusieurs méthodes pour éliminer les variables les moins intéressantes pour le modèle.

- La méthode backward ou méthode descendante.

Cette méthode fonctionne sur les bases d'un algorithme démarrant du modèle complet et en supprimant progressivement les variables les moins significatives pour le modèle. On calcule donc dans un premier temps la régression pour toutes les variables jugées comme explicatives. A chaque étape, la variable associée à la plus grande p-value du test de student (probabilité critique du test) est éliminée. La procédure s'arrête lorsque l'ensemble des variables restant dans le modèle ont des p-values plus petites que le seuil fixé au départ.

Cette méthode est couramment utilisée lorsque le raisonnement inductif seul ne permet pas de dégager une série de variables explicatives satisfaisante. Elle présente l'inconvénient majeur d'écartier définitivement une variable lorsqu'elle a été supprimée du modèle. Une variable non significative à une étape du modèle peut le redevenir

lorsque d'autres variables avec lesquelles elle était fortement corrélée ont été à leur tour écartées.

- La méthode forward ou méthode ascendante

Cette méthode procède dans le sens inverse de la précédente. On examine le modèle avec une seule variable explicative, puis on introduit une à une d'autres variables explicatives. On effectue donc toutes les régressions possibles avec une seule variable jugée comme explicative et on retient la variable pour laquelle le test de Student est le plus significatif.

On effectue par la suite toutes les régressions possibles avec deux variables explicatives parmi lesquelles se trouve la variable explicative sélectionnée auparavant et rechercher la nouvelle variable la plus significative. On réitère le processus jusqu'à ce qu'aucune d'entre elles ne soit significative relativement au seuil fixé au début par la p-value.

L'inconvénient de cette méthode est du même ordre que pour la méthode backward : une fois introduite dans le modèle, une variable ne peut plus en être éliminée. Ainsi, le modèle final peut contenir des variables qui ne sont plus significatives une fois toutes les variables introduites. On rencontre ce genre de problème principalement concernant les variables soumises à une très forte multi colinéarité.

Ce problème trouve sa solution dans le Stepwise

- Le stepwise ou méthode mixte

Cette méthode fonctionne sur le même principe d'ajout de variables que la méthode forward. Cette procédure diffère toutefois de la précédente puisqu'elle fonctionne sur le réexamen à chaque étape de toutes les variables introduites précédemment. La méthode Stepwise propose donc, après l'introduction d'une nouvelle variable, de réexaminer les tests de Student pour chaque variable explicative anciennement admise, et de retirer du modèle une ou plusieurs variables si elles s'avèrent n'être plus significatives dans le modèle. Le processus s'arrête lorsque aucune variable ne peut être introduite ou retirée du modèle.

En raison des très fortes corrélations entre les variables explicatives, nous avons pris le parti d'utiliser la méthode des Stepwise. Nous verrons par la suite,

lors de leur application, les différents indices et tests qui vont nous permettre d'évaluer le modèle calculé.

3. Emissions et attractions à l'échelle communale

Nous avons tenté, pour chaque motif de déplacement, de conserver un maximum de variables jugées explicatives afin de permettre un choix très large de sélection par le Stepwise.

Les résultats des régressions multiples sont présentés par motif, en détaillant les émissions des attractions.

3.1. Emissions pour le motif travail

Les déplacements liés au travail, fortement contraints, s'expliquent certainement plus facilement que les autres. C'est le seul motif pour lequel le Stepwise n'a retenu qu'une variable explicative (tableau 11).

En effet, le nombre d'actifs de la zone conditionne beaucoup le nombre d'individus se déplaçant pour un motif professionnel. On obtient donc très logiquement une régression de très haute qualité et une part d'explication du modèle qui s'élève à 99 %.

Modèle de régression							
Variable à expliquer : E_{Trav} = Emissions pour le motif travail							
Variable explicative : $RATA$ = Nombre d'actifs dans la zone							
$\log(E_{Trav}) = 0,993 \log(RATA) - 0,0212$							
Variance				Indices et test d'évaluation			
	Somme des carrés	dl	Carré moyen		R ²	R ² ajusté	R ² predict
Modèle	134,121	1	134,121	NCo	0,999	0,999	0,999
Résidus	0,197	523	0,000376	Co	0,999	0,999	0,999
Total	134,318	524			F=	356663 (NCo)	738340 (Co)
Elasticité							
Variables	$RATA$						
Echantillon non contraint	0,993 (0,002)						
Echantillon contraint	0,988 ((0,001)						

Tableau 11 : Modèle de régression pour les émissions liées au motif travail

On obtient ici un modèle très satisfaisant, que l'échantillon soit contraint ou non. Nous ne ferons ici qu'un bref rappel concernant les différents indices et

tests d'évaluation du modèle. Nous nous référons à des ouvrages traitant cette question de manière exhaustive.¹⁷

Le R^2 est le coefficient de détermination, il permet de traduire la variance expliquée par le modèle. Il exprime l'intensité de la liaison entre la variable à expliquer et l'ensemble des régresseurs. Ce coefficient présente l'inconvénient majeur d'augmenter à mesure que l'on ajoute des variables dans le modèle. En clair, un modèle avec un nombre élevé de variables va tendre vers un R^2 qui peut parfois être surestimé. On fait appel dans ce cas au R^2 ajusté qui est corrigé par les degrés de liberté.

Le $R^2_{predict}$ est un coefficient de prédiction défini par Turchin et Ellner [Turchin et Ellner, 2000] qui équivaut à la proportion de la variance des densités expliquées par le modèle. Il est calculé sur le contre échantillon de données et est défini par :

$$R^2_{predict} = 1 - \frac{\text{Moyenne des erreurs au carré}}{\text{Variance des données prédites}}$$

Si $R^2_{predict} \approx R^2$, alors les données prévues possèdent une certaine validité. Lorsque $R^2_{predict} < R^2$, alors il est probable que le modèle n'ait pas saisi toute la prévisibilité des données, suggérant que le modèle est basé sur des principes ou des paramètres du modèle incorrects. Enfin, si $R^2_{predict} > R^2$, on peut considérer que ce coefficient traduit une approbation forte du modèle. Il suggère que les données prédites, fondées sur le modèle, fournissent une exactitude encore meilleure que les données elles-mêmes. Ce résultat s'explique lorsque l'on dispose de beaucoup d'individus et que le modèle permet d'isoler certains individus dont le comportement échappe à toute logique.

Enfin, si ces différents coefficients de détermination mettent en évidence l'intensité de la liaison entre la variable à expliquer et les variables explicatives, ils ne permettent pas de savoir si le modèle est statistiquement pertinent, pour un seuil d'erreur α donné. Ici, nous avons fixé la valeur de α à 0,5, soit 5 % de risque d'erreur acceptable. Nous avons donc calculé un test de significativité F de la régression multiple afin de vérifier si la liaison mise en évidence avec la régression n'est pas un simple artefact.

F est défini par :

¹⁷ Y. Dodge, V. Rousson, *Analyse de régression appliquée*, Dunod, 2004 ; TOMASSONE RICHARD, *La régression : nouveaux regards sur une ancienne méthode statistique*, 1983

$$F = \frac{R^2/p}{(1-R^2)/(n-p-1)}$$

Avec p : nombre de variables explicatives insérées dans le modèle

n : nombre de valeurs de chaque variable

Si F calculé est supérieur à F théorique défini en annexe 5, le coefficient de détermination multiple est significatif pour le degré d'erreur fixé.

Dans le cas des émissions pour le motif travail, le F calculé est très largement supérieur au F théorique. La bonne qualité de la régression vue auparavant est donc considérée comme significative.

Afin de comparer au mieux les résultats obtenus selon les deux méthodes choisies d'échantillonnage, nous avons également calculé la variance entre les données prédites sur les contre échantillons issues du modèle contraint et celles du modèle non contraint :

$$E(\text{var})_{\text{travail}} = 1,5065E - 05$$

Nous avons donc une variance très faible au regard de l'ordre de grandeur des valeurs. Le modèle est donc robuste, quel que soit la technique d'échantillonnage choisie, en ce qui concerne les émissions liées au motif travail. Nous verrons à la fin de ce chapitre, une fois les variances liées à chaque motif calculées, si cette constatation peut-être faite sur l'ensemble de l'analyse.

La qualité du modèle peut également être éprouvée graphiquement, en croisant les données de mobilité prédites aux données de mobilité réelles (figure 27).

La mobilité prédite par le modèle est représentée sur l'axe des ordonnées, la mobilité réelle est portée sur l'axe des abscisses. On retrouve ici les caractéristiques déjà évoquées par les différents indices et tests de la régression qui nous amène à une excellente adéquation entre ces deux jeux de données. Nous avons pris le parti ici de représenter les valeurs prédites et réelles encore logarithmées afin de mieux se rendre compte graphiquement du résultat. En effet, le retour aux valeurs réelles fait disparaître la répartition normale des individus, amène des valeurs très fortes et a tendance à regrouper graphiquement le reste des observations ce qui amenuise la qualité de l'observation (figure 28)

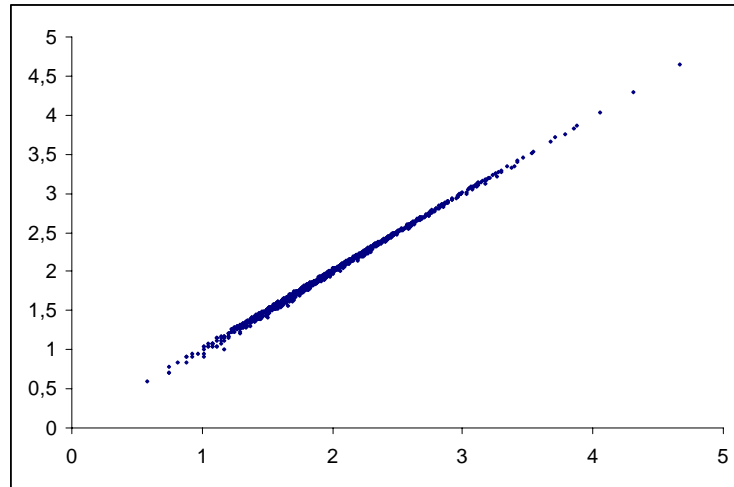


Figure 27 : Validation du modèle E_{Trav} (logarithme des valeurs) par le croisement des flux prédits avec les flux observés

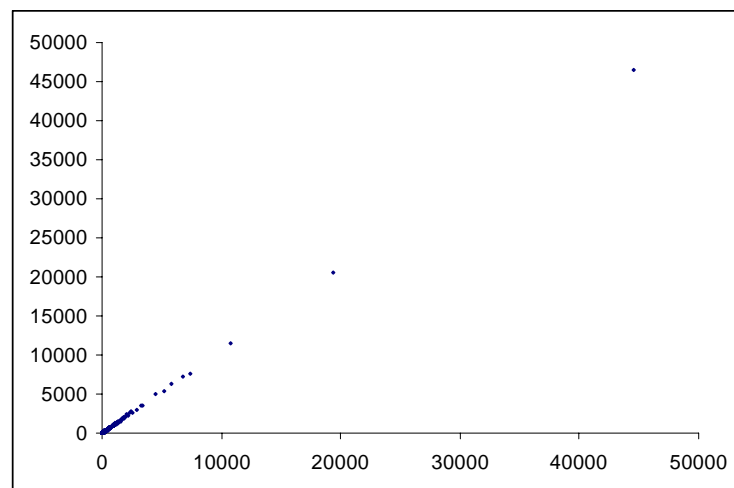


Figure 28 : Validation du modèle E_{Trav} (valeurs réelles) par le croisement des flux prédits avec les flux observés

3.2. Attractions pour le motif travail

Comme nous l'avons déjà dit, le motif travail trouve son explication à travers des variables déjà connues et facilement identifiables. Comme pour les émissions, le Stepwise n'a trouvé concernant les attractions liées au travail, qu'une seule variable explicative. Le nombre d'emplois de la zone explique en effet très logiquement l'attraction de cette zone concernant le motif professionnel (tableau 12).

Modèle de régression							
Variable à expliquer : A_{Trav} = Attractions pour le motif travail							
Variable explicative : EET = Nombre d'emplois dans la zone							
$\log(A_{Trav}) = 0,999 \log(EET) - 0,01403$							
Variance				Indices et test d'évaluation			
	Somme des carrés	dl	Carré moyen		R ²	R ² ajusté	R ² predict
Modèle	203,058	1	203,056	NCo	0,996	0,996	0,996
Résidus	0,908	523	0,001736	Co	0,998	0,998	0,996
Total	203,967	524			F=	116935 (NCo)	211384 (Co)
Elasticité							
Variables	EET						
Echantillon non contraint	0,999 (0,005)						
Echantillon contraint	1,001 ((0,002))						

Tableau 12 : Modèle de régression pour les attractions liées au motif travail

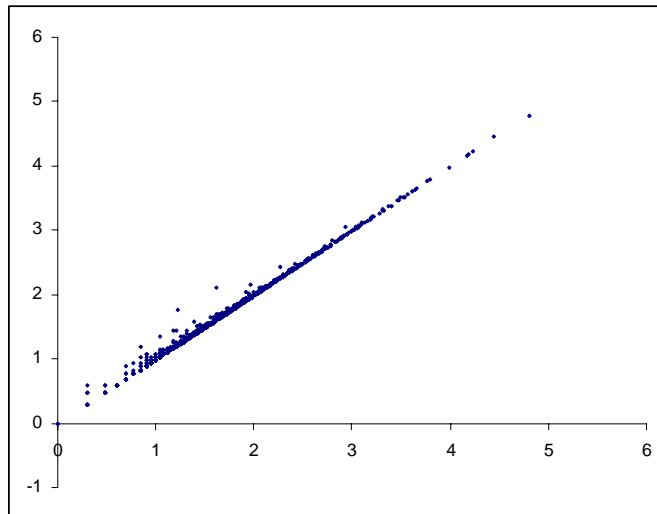


Figure 29 : Validation du modèle A_{Trav} par le croisement des flux prédits avec les flux observés

L'ensemble des indices et tests de cette régression nous permettent de conclure à la validité du modèle. La figure 29 montre une fois encore la très bonne adéquation entre les données prédites et les valeurs réelles.

3.3. Emissions pour le motif études

La mobilité quotidienne liée au motif études reste une mobilité déjà souvent appréhendée et bien connue des modélisateurs. Le modèle défini par la démarche pas à pas trouve sa pertinence maximale avec trois variables explicatives (tableau 13). Il s'agit très logiquement de la part de la population fortement concernée par un tel motif.

L'ensemble des indices ainsi que la figure 30 nous permettent d'approuver ce modèle et de juger sa validité comme satisfaisante. Aucun écart significatif, que ce soit dans la validité ou dans le paramétrage du modèle, n'apparaît quant aux deux méthodes d'échantillonnage.

Modèle de régression							
Variable à expliquer : E_{Etu} = Emissions pour le motif études							
Variables explicatives : $RAF7$ = Nombre d'enfants scolarisés de 7 à 15 ans							
$RDPO15$ = Population totale de 15 à 19 ans							
$RAF3$ = Nombre d'enfants scolarisés de 3 à 6 ans							
$\log(E_{Etu}) = 0,532 \log(RAF7) + 0,310 \log(RDPO15) + 0,193(RAF3) + 0,392$							
Variance				Indices et test d'évaluation			
	Somme des carrés	dl	Carré moyen		R ²	R ² ajusté	R ² predict
Modèle	142,322	3	47,441	NCo	0,991	0,991	0,990
Résidus	1,305	521	0,002504	Co	0,996	0,996	0,994
Total	143,626	524			F=	18946 (NCo)	38853 (Co)
Elasticité							
Variables	$RAF7$	$RDPO15$	$RAF3$				
Echantillon non contraint	0,532 (0,015)	0,310 (0,013)	0,193 (0,010)				
Echantillon contraint	0,554 (0,015)	0,301 (0,013)	0,177 (0,011)				

Tableau 13 : Modèle de régression pour les émissions liées au motif études

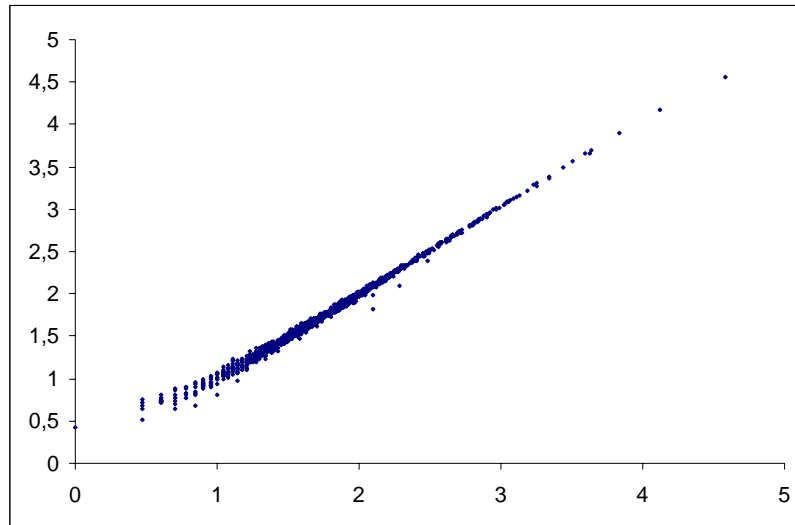


Figure 30 : Validation du modèle E_{Etu} par le croisement des flux prédits avec les flux observés

3.4. Attractions pour le motif études

Modèle de régression							
Variable à expliquer : A_{Etu} = Attractions pour le motif études							
Variables explicatives : $EFFP$ = Effectifs en établissement d'enseignement primaire							
$EFFS$ = Effectifs en établissement d'enseignement secondaire et universitaire							
$\log(A_{Etu}) = 0,875 \log(EFFP) + 0,180 \log(EFFS) + 0,215$							
Variance				Indices et test d'évaluation			
	Somme des carrés	dl	Carré moyen		R ²	R ² ajusté	R ² predict
Modèle	395,296	2	197,648	NCo	0,955	0,955	0,948
Résidus	18,553	522	0,03554	Co	0,966	0,966	0,959
Total	413,849	524			F=	5561 (NCo)	12650 (Co)
Elasticité							
Variables	$EFFP$	$EFFS$					
Echantillon non contraint	0,875 (0,010)	0,180 (0,015)					
Echantillon contraint	0,895 (0,007)	0,150 (0,008)					

Tableau 14 : Modèle de régression pour les attractions liées au motif études

L'attraction liée aux études est naturellement liée à la capacité d'accueil des populations scolarisées de chaque zone. La régression présentée ici est satisfaisante par l'ensemble des tests présentés ici (tableau 14) et paraît concluante au regard de la figure 31.

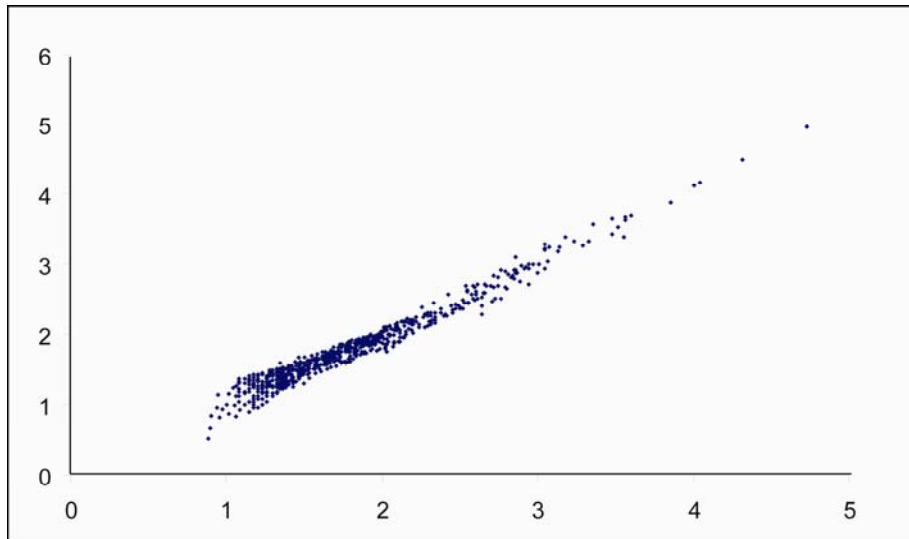


Figure 31 : Validation du modèle A_{Etu} par le croisement des flux prédits avec les flux observés

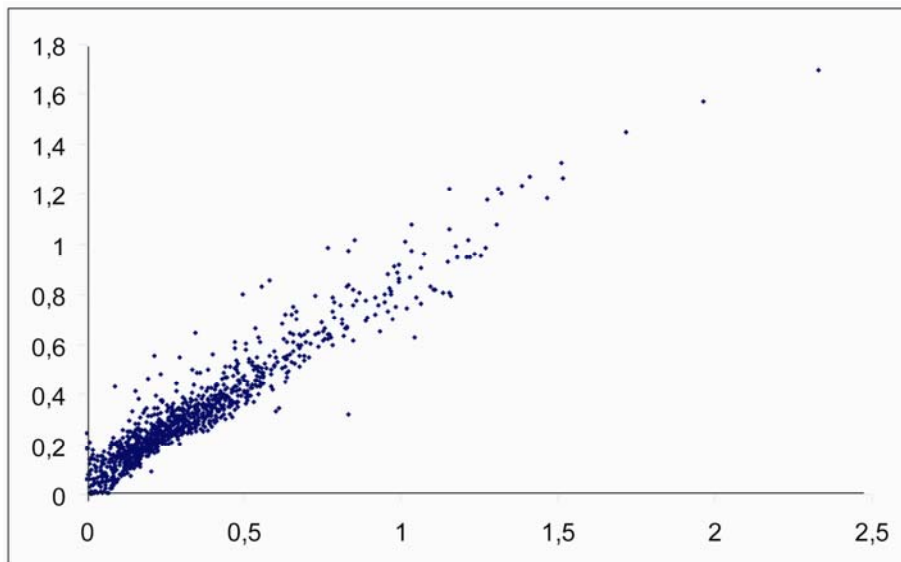
Les résultats présentés ci-après résument les différents modèles mis en place en fonction des motifs de déplacement. Ils présentent un ensemble concluant que nous ne commenterons pas en détail comme nous l'avons fait jusqu'ici pour réserver nos commentaires aux modèles moins performants.

3.5. Les déplacements liés au motif achats

Il est à noter ici que le nombre de commerces de détail défavorise les émissions pour le motif achats. On le voit par l'élasticité négative constatée par rapport à cette variable. On peut s'interroger sur l'élasticité positive liée à la présence de grandes surfaces commerciales. Cette tendance peut certainement s'expliquer par la localisation de ces espaces commerciaux, souvent situés sur des communes périphériques aux grandes villes n'accueillant souvent que très peu d'autres équipements commerciaux.

Modèle de régression							
Variable à expliquer : E_{Achat} = Emissions pour le motif achats							
Variables explicatives : $RATA$ = Nombre d'actifs dans la zone							
$GSPC$ = Grandes surfaces à prédominance commerciale							
$ACDMON$ = Autres commerces de détail en magasin ou non							
$\log(E_{Achat}) = 0,352 \log(RATA) + 0,338 \log(GSPC) - 0,105(ACDMON) - 0,464$							
Variance				Indices et test d'évaluation			
	Somme des carrés	dl	Carré moyen		R ²	R ² ajusté	R ² predict
Modèle	33,259	4	8,315	NCo	0,912	0,911	0,902
Résidus	3,227	520	0,006206	Co	0,918	0,918	0,910
Total	36,486	524			F=	1339 (NCo)	2234 (Co)
Elasticité							
Variabes	$RATA$	$GSPC$	$ACDMON$				
Echantillon non contraint	0,352 (0,011)	0,338 (0,065)	-0,105 (0,021)				
Echantillon contraint	0,383 (0,012)	0,204 (0,050)	-0,100 (0,023)				

Tableau 15 : Modèle de régression pour les émissions liées au motif achats

Figure 32 : Validation du modèle E_{Achat} par le croisement des flux prédits avec les flux observés

Modèle de régression							
Variable à expliquer : A_{Achat} = Attractions pour le motif achats							
Variables explicatives : $PCRA$ = Petits commerces et réparation automobile							
$ACDMON$ = Autres commerces de détail en magasin ou non							
EEC = Total des emplois commerciaux							
$MASON$ = Magasins d'alimentation, spécialisés ou non							
$\log(A_{Achat}) = 0,326 \log(PCRA) + 0,224 \log(ACDMON) + 0,124 \log(EEC) + 0,336 \log(MASON) - 0,446$							
Variance				Indices et test d'évaluation			
	Somme des carrés	dl	Carré moyen		R ²	R ² ajusté	R ² predict
Modèle	78,566	4	15,713	NCo	0,764	0,761	0,748
Résidus	24,316	520	0,04685	Co	0,854	0,852	0,837
Total	102,882	524			F=	335 (NCo)	605 (Co)
Elasticité							
Variables	$PCRA$	$ACDMON$	EEC	$MASON$			
Echantillon non contraint	0,326 (0,067)	0,224 (0,055)	0,124 (0,023)	0,336 (0,076)			
Echantillon contraint	0,269 (0,059)	0,142 (0,056)	0,130 (0,025)	0,312 (0,066)			

Tableau 16 : Modèle de régression pour les attractions liées au motif achats

Certains éléments peuvent ici nous surprendre. L'absence des grandes surfaces pour expliquer l'attraction des communes en terme commercial paraît en effet surprenante. Cette absence s'explique certainement par deux aspects : n'est prise en compte ici que la mobilité intercommunale, et on peut supposer qu'un grand nombre de déplacements dans ces espaces commerciaux sont intra communaux d'une part ; d'autre part, ne sont prises en compte que les données d'enquête déclarées par les individus enquêtés. On peut alors supposer que l'individu se rendant dans une grande surface commerciale va associer sa destination à l'agglomération à laquelle se rattache la grande surface. Hors, comme nous l'avons vu auparavant, ces grandes surfaces ne sont pas obligatoirement situées dans ces agglomérations mais dans les communes qui leur sont périphériques. Il se peut alors que la perception de l'enquêté vis-à-vis de sa destination influence le résultat.

La régression, bien que la part d'explication qu'elle fournit soit moins importante que pour les régressions précédentes (85% pour le modèle issu de l'échantillon contraint), reste satisfaisante. On peut toutefois noter l'influence moins importante des commerces de faible taille pour le modèle correspondant à l'échantillon contraint.

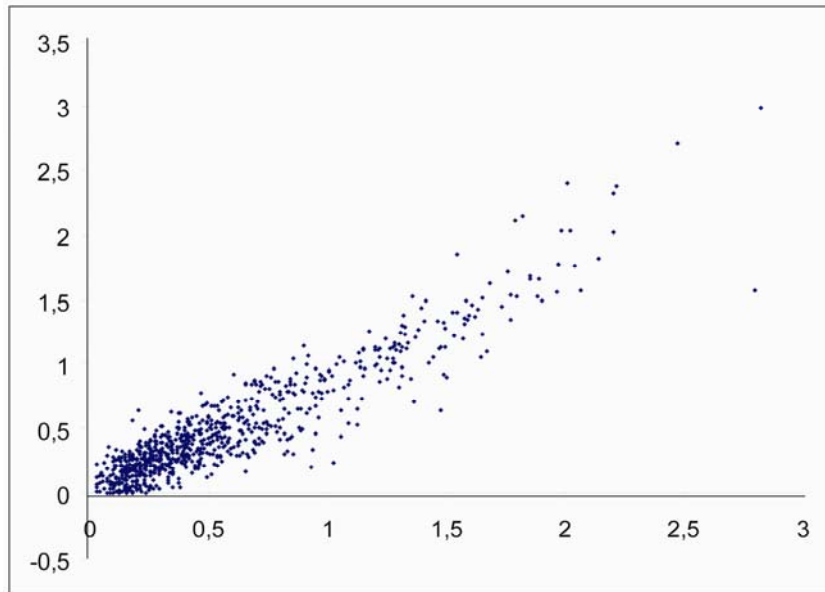


Figure 33 : Validation du modèle A_{Achat} par le croisement des flux prédits avec les flux observés

3.6. Les déplacements liés au motif loisirs

Les résultats liés à la mobilité inhérente aux loisirs nécessitent une plus grande attention. Ils diffèrent selon que l'on travaille sur l'échantillon contraint ou non. En effet, le Stepwise ne retient que trois variables explicatives à partir de l'échantillon contraint contre quatre dans l'autre cas. On retrouve toutefois cette variable si on augmente la tolérance d'introduction d'une nouvelle variable (diminution de la p-value). Il est difficile d'émettre une hypothèse pour expliquer cet écart d'un modèle à l'autre.

Modèle de régression							
Variable à expliquer : E_{Lois} = Emissions pour le motif loisirs							
Variables explicatives : $RDPOP$ = Population totale de la zone							
$ELAV$ = Equipement de loisir ou agence de voyage							
$RD20$ = Part de la population âgée de 20 à 40 ans							
$RD40$ = Part de la population âgée de 40 à 60 ans							
$\log(E_{Lois}) = -0,173 \log(RDPOP) - 0,839 \log(ELAV) + 0,209 \log(RD20) + 0,191 \log(RD40) - 0,108$							
Variance				Indices et test d'évaluation			
	Somme des carrés	dl	Carré moyen		R ²	R ² ajusté	R ² predict
Modèle	21,002	9	2,334	NCo	0,870	0,868	0,857
Résidus	3,127	515	0,006072	Co	0,840	0,838	0,825
Total	24,129	524			F=	384 (NCo)	680 (Co)
Elasticité							
Variables	$RDPOP$	$ELAV$	$RD20$	$RD40$			
Echantillon non contraint	-0,173 (0,074)	-0,839 (0,106)	0,209 (0,037)	0,191 (0,040)			
Echantillon contraint	-0,092 (0,049)	-1,025 (0,094)	0,187 (0,027)				

Tableau 17 : Modèle de régression pour les émissions liées au motif loisirs

Un second aspect de ce modèle nécessite une explication : il s'agit de l'élasticité négative liée à la population totale de la zone. Cette élasticité, prise indépendamment, signifie que plus la population augmente moins on a d'émissions liées aux loisirs. En réalité, il est nécessaire de cerner le modèle dans son ensemble pour en comprendre le sens. Si on remonte aux itérations précédentes du Stepwise, on se rend compte que l'élasticité de cette variable change de signe avec l'introduction des autres variables explicatives liées à la population. En fait, cette élasticité négative renforce le sens de l'élasticité positive des autres variables liées aux caractéristiques de la population.

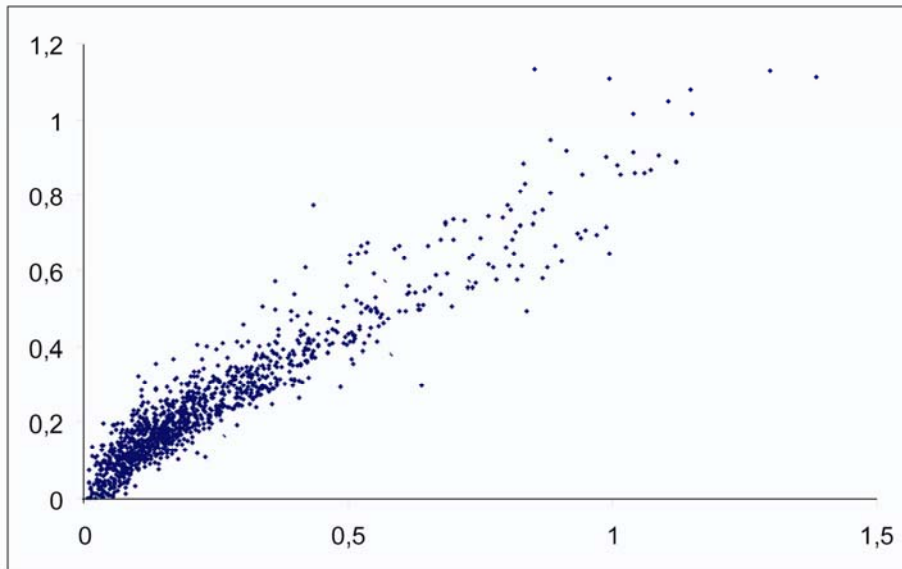


Figure 34 : Validation du modèle E_{Lois} par le croisement des flux prédits avec les flux observés

Modèle de régression							
Variable à expliquer : A_{Lois} = Attractions pour le motif loisirs							
Variables explicatives : $EETEV$ = Total des emplois tertiaires							
HER = Hôtels et restaurants							
$ACDMON$ = Autres commerces de détail en magasin ou non							
$PCRA$ = Petits commerces et réparation automobile							
$ELAV$ = Equipement de loisir ou agence de voyage							
$\log(A_{Lois}) = 0,108 \log(EETEV) + 0,334 \log(HER) + 0,210 \log(ACDMON) + 0,271 \log(PCRA) + 0,124 \log(ELAV) - 0,270$							
Variance				Indices et test d'évaluation			
	Somme des carrés	dl	Carré moyen		R ²	R ² ajusté	R ² predict
Modèle	73,956	5	14,791	NCo	0,775	0,773	0,768
Résidus	21,457	519	0,04134	Co	0,845	0,843	0,825
Total	95,413	524			F=	357 (NCo)	564 (Co)
Elasticité							
Variables	$EETEV$	HER	$ACDMON$	$PCRA$	$ELAV$		
Echantillon non contraint	0,108 (0,022)	0,334 (0,053)	0,210 (0,051)	0,271 (0,063)	0,124 (0,045)		
Echantillon contraint	0,113 (0,025)	0,221 (0,048)	0,178 (0,053)	0,211 (0,062)	0,118 (0,038)		

Tableau 18 : Modèle de régression pour les attractions liées au motif loisirs

Il peut paraître surprenant, concernant les attractions liées au motif loisirs, de ne trouver la variable *ELAV* qu'en dernière variable introduite par le Stepwise. En réalité, c'est la très forte corrélation avec la première variable introduite (*EETEV*) qui explique une si faible importance pour cette variable.

On peut noter ici une relative amélioration du modèle avec le recours à l'échantillon contraint. Cette amélioration trouve certainement son explication par la meilleure représentation des données dans ce type d'échantillon, la plupart des équipements sportifs étant localisé dans les communes pourvues d'une population importante.

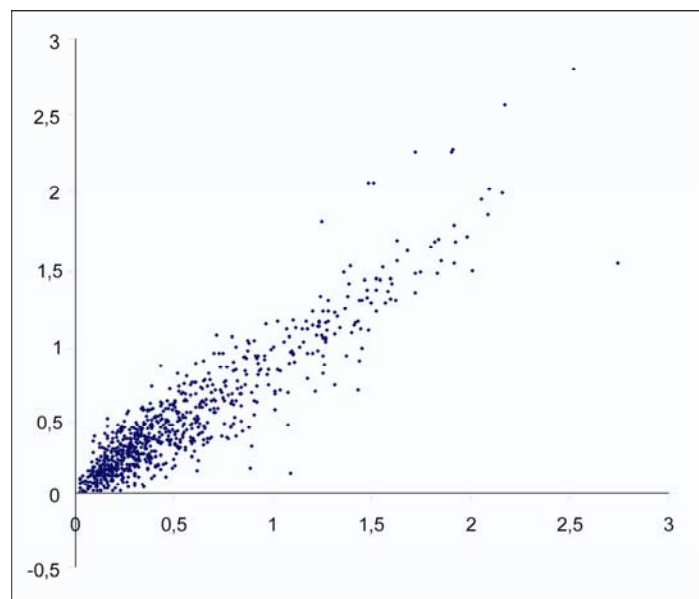


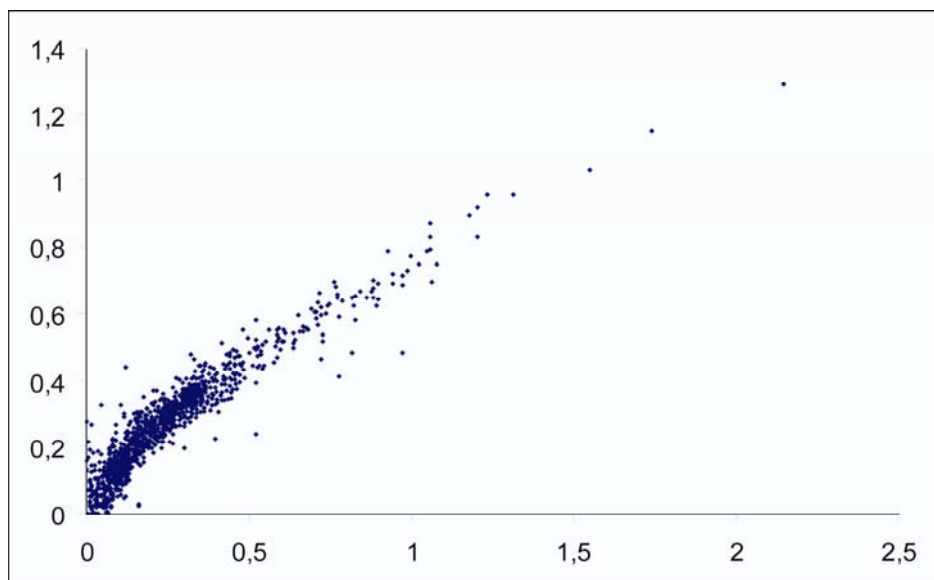
Figure 35 : Validation du modèle A_{Lois} par le croisement des flux prédits avec les flux observés

3.7. Les déplacements liés au motif visites

L'élasticité négative constatée ici pour la population totale de la zone est à rapprocher du phénomène vu précédemment pour les émissions liées aux loisirs. On constate donc ici que ce sont les populations jeunes ou âgées qui expliquent le mieux la mobilité liée aux visites à des parents ou amis.

Modèle de régression							
Variable à expliquer : E_{Visi} = Emissions pour le motif visites							
Variables explicatives : $RDPOP$ = Population totale de la zone							
$RD75$ = Part de la population âgée de 75 ans et plus							
$RD20$ = Part de la population âgée de 20 à 40 ans							
$RLIV$ = Part de la population à posséder un véhicule par foyer							
$\log(E_{Visi}) = -0,265 \log(RDPOP) + 0,06028 \log(RD75) + 0,248 \log(RD20)$							
$+ 0,187 \log(RLIV) - 0,161$							
Variance				Indices et test d'évaluation			
	Somme des carrés	dl	Carré moyen		R ²	R ² ajusté	R ² predict
Modèle	18,398	5	3,680	NCo	0,848	0,846	0,840
Résidus	3,306	519	0,006371	Co	0,905	0,905	0,897
Total	21,705	524			F=	577 (NCo)	824 (Co)
Elasticité							
Variables	$RDPOP$	$RD75$	$RD20$	$RLIV$			
Echantillon non contraint	-0,265 (0,072)	0,06028 (0,018)	0,248 (0,038)	0,187 (0,036)			
Echantillon contraint	-0,236 (0,067)	0,110 (0,021)	0,132 (0,032)	0,241 (0,038)			

Tableau 19 : Modèle de régression pour les émissions liées au motif visite

Figure 36 : Validation du modèle E_{Visi} par le croisement des flux prédits avec les flux observés

Modèle de régression							
Variable à expliquer : A_{visi} = Attractions pour le motif visites							
Variables explicatives : $PCRA$ = Petits commerces et réparation automobile							
$RD75$ = Part de la population âgée de 75 ans et plus							
$RD20$ = Part de la population âgée de 20 à 40 ans							
TRV = Services de transport routier de voyageurs							
$\log(A_{visi}) = 0,443 \log(PCRA) + 0,117 \log(RD75) + 0,353 \log(RD20) + 0,220 \log(TRV) - 0,379$							
Variance				Indices et test d'évaluation			
	Somme des carrés	dl	Carré moyen		R ²	R ² ajusté	R ² predict
Modèle	62,899	5	12,580	NCo	0,753	0,751	0,741
Résidus	20,603	519	0,03970	Co	0,831	0,830	0,824
Total	83,502	524			F=	316 (NCo)	519 (Co)
Elasticité							
Variables	$PCRA$	$RD75$	$RD20$	TRV			
Echantillon non contraint	0,443 (0,058)	0,117 (0,036)	0,353 (0,033)	0,220 (0,095)			
Echantillon contraint	0,367 (0,055)	0,193 (0,040)	0,346 (0,037)	0,105 (0,082)			

Tableau 20 : Modèle de régression pour les attractions liées au motif visite

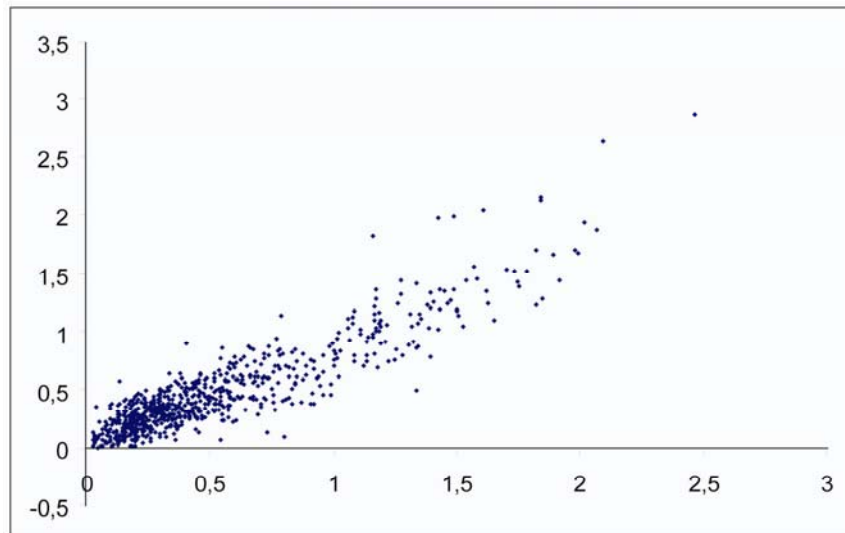


Figure 37 : Validation du modèle A_{visi} par le croisement des flux prédits avec les flux observés

On retrouve très logiquement les mêmes groupes de population que ceux concernés par les émissions liées aux visites. La part de petits commerces semble également être une variable favorisant l'attraction d'une zone en terme

de visites. Le regroupement d'activités au sein d'une même sortie peut peut-être expliquer en partie ce résultat, bien qu'il paraisse surprenant qu'un individu puisse choisir sa destination de visite en fonction de la proximité d'un espace commercial dans la zone d'arrivée de son déplacement.

La présence de service de transport de voyageurs semble également avoir une incidence positive sur l'attraction des zones pour le motif nous intéressant ici. La population âgée, concernée ici par ce phénomène, représente en effet une population captive des transports en commun et peut donc expliquer ce phénomène. Nous avons donc ici une mobilité qui peut être influencée par un choix modal contraint.

Le modèle paraît obtenir de relatifs meilleurs résultats par l'échantillon contraint dans ce cas.

3.8. Les déplacements liés au motif santé

Modèle de régression							
Variable à expliquer : $E_{Santé}$ = Emissions pour le motif santé							
Variables explicatives : ARS = Activité relative à la santé							
$ARAS$ = Activité relative à l'action sociale							
$RD20$ = Part de la population âgée de 20 à 40 ans							
$RD40$ = Part de la population âgée de 40 à 60 ans							
$RLIV$ = Part de la population à posséder un véhicule par foyer							
$\log(E_{Santé}) = -0,132 \log(ARS) - 0,400 \log(ARAS) + 0,0570 \log(RD20) + 0,0341 \log(RD40) + 0,0500(RLIV) - 0,147$							
Variance				Indices et test d'évaluation			
	Somme des carrés	dl	Carré moyen		R ²	R ² ajusté	R ² predict
Modèle	4,714	5	0,943	NCo	0,768	0,765	0,745
Résidus	1,428	519	0,002751	Co	0,734	0,732	0,715
Total	6,142	524			F=	342 (NCo)	328 (Co)
Elasticité							
Variables	ARS	$ARAS$	$RD20$	$RD40$	$RLIV$		
Echantillon non contraint	-0,132 (0,013)	-0,400 (0,063)	0,0570 (0,018)	0,0341 (0,016)	0,0500 (0,019)		
Echantillon contraint	-0,158 (0,020)	-0,386 (0,066)	0,0478 (0,017)	0,0594 (0,025)	0,0813 (0,027)		

Tableau 21 : Modèle de régression pour les émissions liées au motif santé

Les émissions liées à la santé sont fortement corrélées à la présence ou non d'un service de santé dans la zone concernée –l'élasticité négative nous montre que la présence d'un service de santé défavorise les émissions de déplacement pour un tel motif– et s'explique également par la population faisant plus volontiers appel à ces services. L'absence de population âgée peut nous surprendre. Cette déficience peut s'expliquer par une mobilité réduite de ce groupe d'individus et la possibilité de déplacer le service plutôt que la personne. On peut également émettre l'hypothèse que les personnes les plus soumises à la fréquentation régulière des services de santé, comme les personnes âgées par exemple, modifient leur lieu d'habitat en fonction de ces services et n'ont, de fait, plus de déplacements intercommunaux à réaliser concernant ce motif.

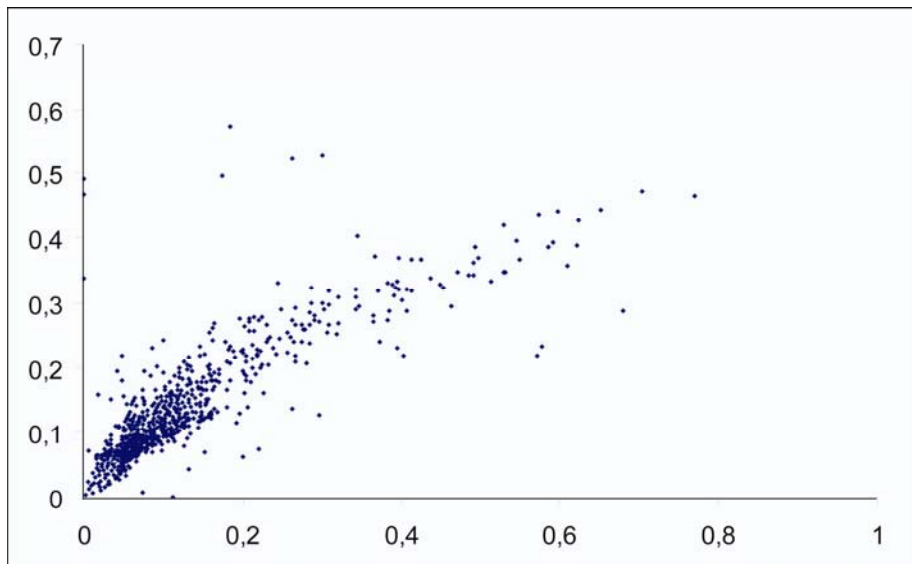


Figure 38 : Validation du modèle $E_{santé}$ par le croisement des flux prédits avec les flux observés

Concernant les attractions de déplacements liées au motif de santé, la présence ou non de services liés à ce motif suffit à expliquer ce phénomène. Nous avons tenté, devant les résultats légèrement moins probants de ces deux derniers modèles, d'employer un modèle catégoriel pour améliorer la qualité de la régression. Cependant, le détail des relations de chaque variable explicative avec la variable à expliquer ne nous a fourni aucun résultat nous permettant de catégoriser la population statistique de cette série.

Modèle de régression							
Variable à expliquer : $A_{Santé}$ = Attraction pour le motif santé							
Variables explicatives : ARS = Activité relative à la santé							
$ARAS$ = Activité relative à l'action sociale							
$\log(A_{Santé}) = 0,695 \log(ARS) + 0,448 \log(ARAS) + 0,01499$							
Variance				Indices et test d'évaluation			
	Somme des carrés	dl	Carré moyen		R ²	R ² ajusté	R ² predict
Modèle	8,839	2	4,419	NCo	0,659	0,658	0,652
Résidus	4,574	522	0,008763	Co	0,761	0,760	0,752
Total	13,413	524			F=	504 (NCo)	832 (Co)
Elasticité							
Variables	ARS	$ARAS$					
Echantillon non contraint	0,448 (0,016)	0,695 (0,112)					
Echantillon contraint	0,592 (0,017)	0,612 (0,105)					

Tableau 22 : Modèle de régression pour les attractions liées au motif santé

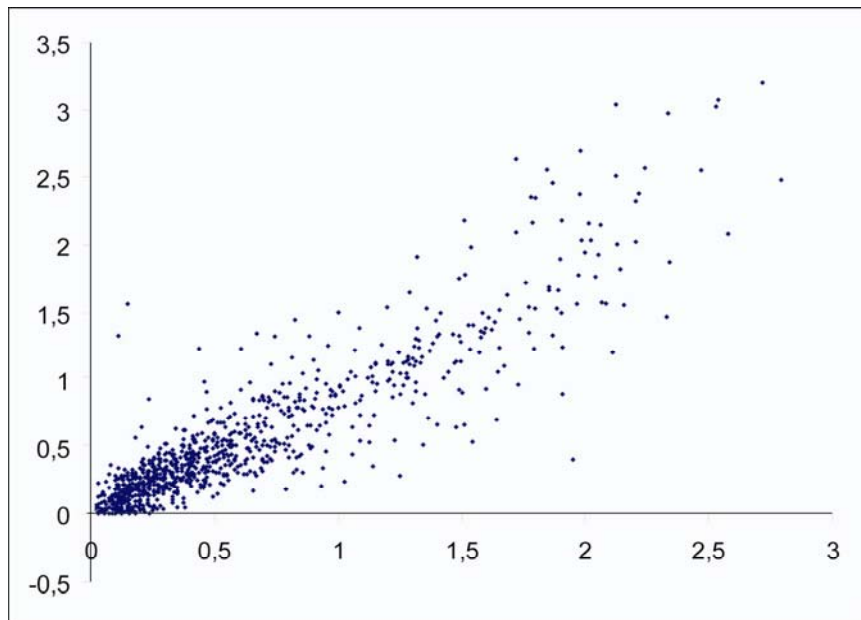


Figure 39 : Validation du modèle $A_{Santé}$ par le croisement des flux prédits avec les flux observés

L'influence de l'échantillonnage, bien que visible parfois à travers les modèles que nous venons de présenter, ne nous apparaît que peu importante. On semble toutefois gagner globalement en qualité de régression par le biais de l'échantillon contraint par un tirage réparti dans les groupes de communes présentant des comportements de mobilité différents.

CONCLUSION

Nous disposons maintenant de l'évaluation des émissions et attractions par motif à l'échelle communale. Nous avons fait ici le choix de méthodologie nous permettant d'identifier finement les variables explicatives liées à l'origine des déplacements, sans faire appel à des formalisations trop lourdes. Nous avons donc pris le parti de conserver un pouvoir d'évaluation de la méthode.

Cette évaluation, construite à partir de données en grande partie issues d'enquêtes, paraît fournir des résultats satisfaisants et présentant une certaine robustesse statistique.

Cependant, la détermination des émissions et attractions à l'échelle communale n'est pas une fin en soi, même si elle peut trouver sa place dans certaines simulations de la mobilité régionale, transcendant le simple service TER.