



**HAL**  
open science

# Quelle prise en compte des dynamiques urbaines dans la prévision de la demande de transport ?

Jorge Cabrera Delgado

## ► To cite this version:

Jorge Cabrera Delgado. Quelle prise en compte des dynamiques urbaines dans la prévision de la demande de transport ?. Economies et finances. Université Lumière - Lyon II, 2013. Français. NNT : . tel-00877044

**HAL Id: tel-00877044**

**<https://theses.hal.science/tel-00877044>**

Submitted on 25 Oct 2013

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

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

Université Lumière Lyon 2  
Faculté de Sciences Économiques et de Gestion

# Quelle prise en compte des dynamiques urbaines dans la prévision de la demande de transport ?

Thèse

pour l'obtention du

Doctorat de Sciences Économiques mention Économie des Transports

Présentée et soutenue publiquement par

Jorge E. CABRERA DELGADO

le 1<sup>er</sup> juillet 2013

Membres du Jury :

---

M. Alain AYONG LE KAMA, Professeur de Sciences Économiques à l'Université  
Paris Ouest – Nanterre La Défense

---

M. Patrick BONNEL, Directeur de Recherches au Laboratoire d'Économie des  
Transports, ENTPE – Université Lumière Lyon 2, Directeur de Thèse

---

M. Eric CORNELIS, Professeur à l'Université de Namur, Rapporteur

---

M. Yves CROZET, Professeur de Sciences Économiques à l'Institut d'Études  
Politiques de Lyon

---

M. Michel DIMOU, Professeur de Sciences Économiques à l'Université du Sud  
Toulon-Var, Rapporteur



---

# Remerciements

Je dois commencer par exprimer toute ma reconnaissance à Patrick BONNEL, qui m'a confié ce travail de thèse et l'a encadré. Ses relectures rigoureuses, ses conseils, sa grande disponibilité et, sur la fin, ses encouragements, ont été d'une aide précieuse. Je lui exprime ma très profonde gratitude.

Je remercie Eric CORNELIS et Michel DIMOU, qui ont accepté d'être rapporteurs de cette thèse, ainsi que Alain AYONG LE KAMA et Yves CROZET de l'honneur qu'ils me font en participant au jury.

Merci à Jean-Pierre NICOLAS et aux autres personnes avec qui j'ai eu le plaisir de travailler au sein du projet SIMBAD. J'ai une pensée particulière pour Cécile GODINOT, Thippaphone KHENMINGMONGKHON et Mériem NOUREDDINE sans l'aide de qui la codification des réseaux de transport aurait été un obstacle infranchissable.

Merci aussi à tous les membres du Laboratoire d'Économie des Transports pour leur accueil, leur disponibilité et les excellentes conditions de travail dont ils m'ont fait bénéficier. Je remercie tout spécialement Florent, mon collègue du bureau T125, et Sonia qui ont partagé avec moi les plaisirs et les peines du parcours du doctorant. Je remercie également Louafi, Pascal et Caroline pour leurs conseils avisés et leur gentillesse, Martine SEFSAF pour son aide dans mes recherches documentaires, ainsi que Didier, Florence, Loulou, Marie et Olivier pour leur sympathie et leur bonne humeur.

Je suis redevable à la Région Rhône-Alpes et à l'ADEME d'avoir soutenu financièrement cette recherche, respectivement, par une allocation doctorale dans le cadre des Clusters de Recherche, et par un contrat de recherche (N° 08 66 C0084) dans le cadre du PREDIT, ainsi qu'à la Faculté de Sciences Économiques et de Gestion de l'Université Lumière Lyon 2 de m'avoir permis d'exercer la fonction d'ATER.

Finalement, je dois exprimer ma gratitude envers mes proches, en France et en Bolivie. Amis, vous avez toujours eu le bon mot ou le bon geste pour m'encourager. Famille, votre soutien sans faille, même à distance, a sans doute été l'élément le plus important pour que cette thèse puisse voir le jour.



---

# Sommaire

<b>Introduction</b>	<b>7</b>
<b>1 Objectifs et principes de la modélisation à quatre étapes</b>	<b>13</b>
1.1 Prévoir la demande de transport : l'approche des modèles économiques	13
1.2 Familles de modèles pour la spécification des quatre étapes . . . . .	25
1.3 Synthèse et remarques conclusives . . . . .	33
<b>2 Transférabilité temporelle des modèles de demande de transport : état des savoirs</b>	<b>35</b>
2.1 Validation des modèles prospectifs ou étude de leur transférabilité temporelle . . . . .	35
2.2 Évidence sur la transférabilité temporelle des modèles de prévision de la demande de transport . . . . .	38
2.3 Conclusion . . . . .	48
<b>3 Terrain d'étude et données pour la modélisation</b>	<b>49</b>
3.1 L'agglomération lyonnaise : reflet des mutations dans les politiques d'aménagement des transports en France . . . . .	50
3.2 Les données relatives à la demande de transport . . . . .	52
3.3 Données relatives à l'offre de transport . . . . .	62
3.4 Synthèse et remarques conclusives . . . . .	90
<b>4 Génération</b>	<b>93</b>
4.1 Une méthodologie pour étudier les évolutions temporelles . . . . .	93
4.2 Étude de la mobilité moyenne par segment de population : est-ce que les émissions moyennes obtenues dans les modèles de classification croisée sont transférables dans le temps ? . . . . .	107
4.3 Le modèle de génération par régression zonale . . . . .	130
4.4 Le modèle de génération par régression à l'échelle des ménages . . . . .	135
4.5 Classification automatique : est-ce optimal de conserver les mêmes segments de population en prévision ? . . . . .	143
4.6 Remarques conclusives . . . . .	144

<b>5</b>	<b>Distribution</b>	<b>147</b>
5.1	Méthodologie . . . . .	147
5.2	Résultats . . . . .	154
5.3	Conclusion . . . . .	160
<b>6</b>	<b>Choix modal</b>	<b>163</b>
6.1	Méthodologie pour comprendre l'évolution du choix du mode . . . . .	164
6.2	L'hypothèse de stabilité des paramètres est-elle acceptable dans le cas des modèles de choix modal? . . . . .	176
6.3	Conclusion . . . . .	185
	<b>Conclusion</b>	<b>187</b>
	<b>Bibliographie</b>	<b>191</b>
 <b>Annexes</b>		
<b>A</b>	<b>Compléments méthodologiques</b>	<b>209</b>
A.1	Dérivation du modèle gravitaire selon l'approche classique dite de maximisation de l'entropie . . . . .	209
A.2	Zonage agrégé commun pour la génération et le choix modal . . . . .	213
<b>B</b>	<b>Définition de la variable « cycle de vie »</b>	<b>214</b>
<b>C</b>	<b>Réseau TC modélisé</b>	<b>215</b>
<b>D</b>	<b>Compléments sur les estimations des modèles de génération</b>	<b>218</b>
D.1	Modèle de génération du projet SIMBAD . . . . .	218
D.2	Représentation graphique des données pour la régression zonale . . . . .	219
D.3	Résultats de la régression au niveau des ménages sans la prise en compte du plan de sondage . . . . .	221
D.4	Arbres de classification . . . . .	223
	<b>Liste des figures</b>	<b>247</b>
	<b>Liste des tables</b>	<b>251</b>
	<b>Table des matières</b>	<b>253</b>

---

# Introduction

*But they will teach us that eternity is the standing still of the present time, a nunc-stans, as the Schools call it; which neither they nor any else understand, no more than they would a hic-stans for an infinite greatness of place.*

— Thomas Hobbes, *Leviathan*, IV, 46

La quantité considérable de ressources publiques mobilisées pour faciliter les déplacements des personnes suscite naturellement le besoin de disposer d'un mécanisme permettant une gestion avisée, si ce n'est optimisée, de celles-ci. Le calcul économique se présente comme un outil d'aide à la décision puissant en ce sens. En effet, reprenant la formule de Marcel Boiteux (2001), « l'intérêt de faire du bilan socioéconomique, non le critère, mais le moyen d'estimation de la valeur d'un projet, c'est de permettre une analyse des raisons pour lesquelles on est conduit à s'écarter de la solution à laquelle ce seul bilan aurait conduit, et de pouvoir ainsi chiffrer le surcoût de la décision ». Flyvbjerg (2005) remarque que, au cœur des estimations de la viabilité financière des projets d'infrastructures de transport, de l'évaluation de leurs impacts socioéconomiques et environnementaux ainsi que de celle de politiques utilisant d'autres leviers que la capacité, on trouve des prévisions de la demande de transport<sup>1</sup>. Ce travail s'intéresse à la capacité des modèles de prévision de la demande de transport à produire des estimations fiables.

## Enjeux de la justesse des prévisions pour la planification des transports en urbain

Le rôle central des estimations de demande de transport ne rend pas les erreurs de prévision moins nombreuses ni moins importantes. Flyvbjerg *et al.* (2005) mettent en évidence des sous-estimations de demande pour les projets routiers de 9,5 % en moyenne et des surestimations de 51,4 % en moyenne pour les projets ferroviaires. Cette étude suggère que les causes de la mauvaise estimation de la demande dans les projets routiers sont plutôt d'ordre technique (différents aspects de la modélisation), alors que pour les projets ferroviaires des enjeux exogènes semblent l'emporter, sans que les aspects techniques soient mis hors de cause. Le *Skytrain* de Bangkok est un exemple type, donné par Flyvbjerg (2005), des conséquences que peuvent avoir les

---

1. Aussi appelées prévisions de trafic.

erreurs de prévision. Avec un coût de plus de deux milliards de dollars, l'infrastructure était largement surdimensionnée (prévisions 2,5 fois supérieures à la fréquentation observée à l'ouverture), de telle sorte que (page 523) :

« *As a result, station platforms are too long for the shortened trains that now operate the system, a large number of trains and cars are idly parked in the train garage because there is no need for them, terminals are too large, etc. The project company has ended up in financial trouble and even though urban rail is probably a good idea for a congested and air-polluted city like Bangkok, overinvesting in idle capacity is hardly the best way to use resources, and especially not in a developing nation where capital for investment is scarce.* »

La planification des transports de personnes fait face, en urbain, à ce que les observateurs appellent souvent une « double crise de congestion et de financement » (Masson, 2000 ; Bonnafous, 2004). Le volet « congestion » de cette double crise est lié au fort développement de la mobilité en voiture particulière, qui a conduit à une dégradation des conditions de circulation dans les quartiers centraux et dans les axes radiaux des grandes agglomérations<sup>2</sup>. Le volet « financement » de la double crise affecte particulièrement les transports en commun. Il est lié à une rareté accrue des fonds publics pour les investissements en infrastructures et à des perspectives de recettes faibles pour les opérateurs des transports en commun. En réalité, les investissements augmentant la capacité des réseaux routiers urbains ont été décidés principalement dans une période de relative prospérité économique (années 1960). Le ralentissement de la croissance dans les pays développés, que l'on connaît depuis les années 1970, s'est naturellement traduit par une raréfaction des fonds publics disponibles pour les investissements. Celle-ci a, par ailleurs, coïncidé avec le constat que l'augmentation de capacité était insuffisante pour absorber la croissance du trafic<sup>3</sup>. Les autorités publiques ont donc été amenées à tenter de mieux orienter la demande de déplacements sur le plan modal. Bonnafous remarque que l'on a ajouté une « *mission de fluidité*<sup>4</sup> » à la traditionnelle « *mission de solidarité*<sup>5</sup> » des transports en commun. En France, cette nouvelle mission de fluidité a donné lieu au développement des réseaux de RER à Paris et de métro en province (Lyon, Lille, Marseille...). Cependant, même si ces investissements se sont traduits par des améliorations de part de marché des transports en commun, celle-ci est restée relativement faible, ce qui a conduit à freiner leur développement.

Aux enjeux classiques de congestion et de financement des infrastructures, s'ajoutent de nouvelles considérations relatives à l'efficacité globale du système, à la sécurité, à la pollution, etc. Ainsi, par exemple, comme le soulignent Nicolas *et al.* (2009) (page 3),

---

2. La congestion du réseau de voirie n'a pas un impact uniquement sur la voiture particulière. Elle dérègle également la circulation des transports en commun partageant cette voirie.

3. Dans un article récent, Duranton et Turner (2011) montrent que, aux États-Unis, les investissements en voies rapides ont eu une forte demande induite et qu'ils n'ont pas conduit à une fluidification des conditions de trafic, sur la période 1983-2003.

4. En italique dans le texte d'origine.

5. En italique dans le texte d'origine. La mission de solidarité des transports en commun fait référence à la nécessité d'assurer une mobilité minimale à une population captive. Notons que, avec la forte augmentation de la motorisation, celle-ci a une portée plus réduite.

« la loi sur l'air en 1996, la loi d'Orientation sur l'Aménagement et le Développement Durable de 1999 ainsi que la loi Solidarité et Renouvellement Urbains de 2000 insistent toutes largement sur la notion de développement durable, avec l'exigence d'une meilleure prise en compte des problèmes environnementaux qui ne néglige pas pour autant les aspects économiques et sociaux des questions traitées. »

La gestion efficace des déplacements urbains se présente comme un défi qui, s'il n'est pas relevé, risque de conduire à une « asphyxie pour la ville » (Masson, 2000). La prise de conscience du fait que « l'évolution du système de déplacements est fortement liée aux changements des modes de vie et aux transformations des formes urbaines », en particulier de l'étalement urbain, et que, « de manière réciproque, le mouvement d'étalement urbain a été en partie entretenu par l'accroissement des facilités de transport à destination des zones périphériques (diffusion de l'automobile, développement des projets d'investissements routiers, etc.) » a motivé l'intérêt croissant porté sur des outils de modélisation aptes à intégrer dans l'analyse ces deux types d'interactions. C'est dans ce sens que l'on a assisté en France, pendant les années 2000, au développement des modèles d'interaction transport-urbanisme (ou LUTI<sup>6</sup> de par les sigles de leur dénomination en anglais). Cependant, il convient de noter ici que les modèles LUTI ont une composante consacrée à la modélisation des déplacements et leur émergence ne fait qu'augmenter la nécessité d'étudier la capacité prédictive de cette composante, en raison des horizons d'étude plus éloignés.

## Modélisation statique et hypothèse de stabilité paramétrique en question

En milieu urbain, l'outil quasi-incontournable pour la prévision de la demande de transport de personnes, à des fins de planification de moyen-long terme, est la structure de modélisation séquentielle à quatre étapes<sup>7</sup> — génération, distribution, répartition modale et affectation — (CERTU, 2002 ; TRB, 2007). Une caractéristique importante du modèle à quatre étapes, dans un contexte de prévision, est qu'il s'agit d'une approche statique. En effet, la séquence usuelle de modélisation des déplacements urbains repose sur trois hypothèses fondamentales dans son utilisation en prévision (Smith et Cleveland, 1976) :

- (i) les variables exogènes peuvent être prévues de manière précise ;

---

6. *Land Use and Transport Interaction models*. Les principaux modèles LUTI français sont présentés dans l'ouvrage collectif édité par Antoni (2011).

7. Le modèle à quatre étapes est présenté en détail dans le chapitre suivant. Il est utilisé en France pour prévoir le trafic de nouvelles infrastructures (dimensionnement ou évaluation) ou pour simuler l'évolution des déplacements sur un territoire dans le cadre des études de planification (Plan de Déplacements Urbains, Dossier Voirie de l'Agglomération, Schéma de Cohérence Territoriale...). Les nouvelles contraintes institutionnelles et les nouveaux besoins en matière de gestion d'infrastructures ont conduit à des raffinements importants des modèles, sans pour autant remettre fondamentalement en cause « l'esprit » (Gaudry, 2007) de la séquence à quatre étapes. En effet, parmi ces raffinements on peut citer notamment le développement des modèles désagrégés — choix discrets — (Domencich et McFadden, 1975 ; Ben-Akiva et Lerman, 1985) ou des modèles de schémas d'activités (Ettema et Timmermans, 1997).

- (ii) les modèles fournissent une bonne simulation, juste du point de vue des comportements, des déplacements observés pendant la période de référence ;
- (iii) la structure des modèles en matière de variables explicatives ainsi que les paramètres sont stables au cours du temps.

Si les deux premières hypothèses semblent assez classiques dès lors que l'on a recours à une modélisation, la troisième est problématique en raison du caractère statique du modèle à quatre étapes et des évolutions observées au niveau des mobilités quotidiennes sur des périodes analogues à celles des horizons de prévision<sup>8</sup>.

D'un point de vue conceptuel, une modélisation dynamique, prenant appui sur des données en séries temporelles, serait la manière la plus cohérente, en termes d'objectifs, de prendre en compte les sensibilités de la mobilité urbaine par rapport à une modification du système de transport (Kitamura, 1990). Ortúzar et Willumsen (2001) remarquent que l'approche statique suppose qu'une mesure de la réponse à un changement dans le temps peut être trouvée en calculant des dérivées de la fonction de demande en fonction des variables concernées par une certaine politique. Dit autrement, on fait implicitement l'hypothèse qu'une relation stimulus-réponse peut être dérivée à partir de modèles estimés à un point dans le temps. Ceci serait raisonnable si on avait toujours un nombre conséquent de personnes modifiant leurs choix, de mode ou destination par exemple, dans les deux sens et sans effet d'habitude ou autres liens temporels. On peut argumenter que la modélisation à quatre étapes, statique, est en partie restée incontournable car elle est compatible avec les données disponibles. Dans le transport de longue distance, l'accès à certaines séries chronologiques a permis de mettre en place des modèles dynamiques, permettant de prendre en compte des évolutions temporelles (Cabanne, 2005). La contrainte matérielle joue, en urbain, un rôle déterminant :

- la principale source de données pour étudier les comportements de mobilité est constituée par des enquêtes déplacements (en France, on a, par exemple, les enquêtes ménages déplacements, CERTU, 2008). Le coût élevé de celles-ci fait qu'elles sont réalisées avec une périodicité d'une dizaine d'années, dans le meilleur des cas, ce qui représente déjà un effort considérable ;
- les données de niveau de service des transports sont souvent établies en faisant une codification des réseaux. Cette opération est extrêmement coûteuse et chronophage, ce qui fait que ces données sont produites au cas par cas, et sûrement pas de manière standardisée à des intervalles de temps réguliers. En théorie, il est possible de reconstituer et de coder les réseaux à plusieurs dates. L'opération demande toutefois un travail d'archive conséquent, car il n'existe pas de bases de données ou d'organisme collectant l'historique des réseaux<sup>9</sup>.

Les études de planification visent à prévoir ou à simuler l'évolution du système de transports à moyen-long terme, autrement dit à 10 ans ou plus. Sur ce type d'horizon temporel, l'analyse des données d'enquête permet de mettre en évidence certaines dynamiques qui interrogent la pertinence de l'hypothèse de stabilité des paramètres

---

8. L'hypothèse (iii) est d'autant plus problématique que le caractère séquentiel du modèle peut conduire à des amplifications des erreurs à chaque étape (Zhao et Kockelman, 2002).

9. Ainsi, par exemple, une recherche précédente conduite par la SEMALY et le LET avait dû se contenter d'une reconstitution sommaire reposant sur une représentation très agrégée et schématique des réseaux (Lichère et Raux, 1997a,b).

des modèles de prévision. Au niveau du modèle de génération, le poids des motifs travail et études régresse au profit des motifs considérés comme moins contraints, tel le loisir. Au niveau de la distribution, on constate un allongement régulier des distances sans pouvoir établir avec précision s'il résulte exclusivement de l'accès plus fréquent aux modes motorisés et de l'accroissement des vitesses moyennes de déplacements, ou s'il n'y a pas en plus un besoin (souhaité ou subi selon les interprétations) des individus d'élargir leur aire de pratique des activités quotidiennes. Au niveau de la répartition modale, jusqu'à la fin des années 1990, on a observé une progression de la part de la voiture, au détriment de la marche à pied et des transports en commun. Cette tendance s'est inversée à partir du début des années 2000 (CERTU, 2012).

## Déroulement du travail

Cette recherche aborde la question de la pertinence de l'hypothèse de stabilité des paramètres des modèles de prévision de trafic. Son objectif est d'apporter des éléments empiriques permettant de juger de la validité de cette hypothèse pour la pratique de la planification des transports. Cet objectif peut être décomposé en deux sous-objectifs :

- (i) Le premier sous-objectif concerne la nécessité de se doter de données temporelles permettant d'alimenter des modèles de prévision. Si des enquêtes ménages déplacements sont disponibles à plusieurs dates, ce n'est pas le cas des réseaux. L'agglomération lyonnaise est retenue comme terrain d'application, compte tenu de la connaissance que nous avons du terrain et des acteurs concernés, d'une part, et des outils de modélisation qui ont déjà pu être développés sur cette agglomération, d'autre part. Nous y effectuons la recherche de données sur l'évolution des réseaux de transport et réalisons leur codifications aux différentes dates des enquêtes ménages déplacements disponibles. La codification reprend les principes du projet SIMBAD (Nicolas *et al.*, 2009 ; Nicolas, 2011). Nous devons vérifier que les données disponibles sont comparables dans le temps et, pour les données à produire, établir une méthodologie permettant d'assurer cette même condition.
- (ii) Le deuxième sous-objectif concerne la modélisation, une fois que les données ont été rendues disponibles. Nous estimons des modèles de génération, de distribution et de répartition modale, noyaux de la modélisation séquentielle pour les trois dernières enquêtes ménages déplacements disponibles sur l'agglomération lyonnaise (1985, 1995 et 2006). Le calage de modèles permet d'étudier l'évolution des coefficients estimés et de tirer des enseignements en matière d'évolution des comportements. L'analyse doit tenir compte de l'évolution des périmètres d'enquête et de la précision des données des enquêtes.

La présentation est organisée de la manière suivante : un premier chapitre permet de faire une description des objectifs et des principes théoriques de la modélisation des déplacements de personnes, dans une approche économique. Il présente la démarche de modélisation séquentielle à quatre étapes et les familles de modèles retenues dans le cadre de ce travail. Un deuxième chapitre permet de faire un état des connaissances, en matière empirique, par rapport à l'hypothèse de stabilité des paramètres du modèle

à quatre étapes. Il présente la problématique de la validation temporelle des modèles de prévision et le concept de transférabilité temporelle et fait une revue des études empiriques, avec une perspective temporelle, portant sur des composantes des modèles de prévision. Un troisième chapitre permet de présenter les données utilisées pour cette étude. Il établit les conditions nécessaires pour pouvoir utiliser les données des enquêtes ménages dans une étude diachronique. Il permet également de décrire une méthode reproductible de codification de l'historique des réseaux assurant une consistance des résultats d'affectation. Un quatrième chapitre aborde la question de la stabilité des paramètres de plusieurs méthodes de la modélisation de la génération. Un cinquième chapitre aborde la capacité du modèle de distribution gravitaire à reproduire l'allongement des distances observé ces 20 dernières années. Un sixième chapitre étudie la stabilité paramétrique d'un modèle de choix discrets de type logit. Il évalue également la capacité de celui-ci à reproduire le regain des parts de marché des transports en commun observé à partir du début des années 2000.

---

# Objectifs et principes de la modélisation à quatre étapes

Le modèle à quatre étapes demeure l'outil incontournable de la modélisation de la demande de transports. En réalité, il serait plus juste de parler de séquence de modélisation que de modèle, puisque les manières de formaliser chacune des étapes sont très variées.

Ce chapitre présente les principes de théorie économique qui servent de fondement à l'approche de modélisation de la demande de transport. Il essaye de faire le lien entre les concepts fondamentaux de la théorie économique pertinents pour l'analyse des déplacements urbains et la structure typique retenue dans la modélisation.

Dans un premier temps, il présente les objectifs de la modélisation de la demande de déplacements et les principes économiques sous-jacents. Dans un deuxième temps, il présente les familles de modèles qui seront étudiées, par la suite, dans le présent travail.

## 1.1 Prévoir la demande de transport : l'approche des modèles économiques

Orfeuil (2002) remarque que « [l]'étude des pratiques de déplacement n'est pas une discipline académique établie, elle met en jeu des disciplines variées (anthropologie, sociologie, psychologie et psychosociologie, géographie, économie, démographie, socio-économie, ingénierie), avec des difficultés de dialogue entre les différents groupes ». Dans le cas particulier de la modélisation, celle-ci est encore loin d'être un sujet de « niche » de l'économie. Les publications sur le sujet se font principalement dans des revues de génie civil, de géographie (et, plus généralement, de ce que les anglo-saxons appellent *urban planning*) et d'économie urbaine ou des transports. Or, entre la fin des années 1950 et notre époque, des progrès considérables ont eu lieu dans l'approche économique des comportements de mobilité. Cette section présente les objectifs de la modélisation de la demande de transport de personnes et les concepts fondamentaux de la théorie économique pertinents pour l'analyse des déplacements.

### 1.1.1 Objectifs de la modélisation

Manheim (1980) remarque que le service transport est accompli par un processus extrêmement complexe d'interactions entre individus, systèmes techniques, institutions... Il s'en suit que, pour appréhender cette « réalité » complexe et pour agir sur elle, l'analyste doit être en mesure de prévoir les conséquences que les actions envisagées peuvent avoir sur le système qu'il étudie. Pour relever ce « défi du multidimensionnel » (Bonnafous, 1989), il est amené à se doter d'un outil qui, par abstraction, simplifie cette « réalité » complexe, en ne retenant que ses traits essentiels : le modèle<sup>1</sup>. Bonnafous distingue deux attributs principaux d'un modèle. Le premier est son objet, qui n'est autre chose que « la part de réalité qu'il s'efforce de représenter ». Le deuxième attribut d'un modèle est son objectif, « qui est de mieux comprendre cette part de réalité, d'anticiper son évolution ou encore d'agir sur elle ». Il est possible, à partir de ce deuxième attribut, de proposer un critère pour évaluer la performance de la procédure de simplification intrinsèque à la modélisation. En effet, puisqu'un modèle a un objectif, il peut être évalué selon sa capacité à atteindre cet objectif. Dans cette optique, Bonnafous introduit le concept d'« opérationnalité » du modèle. Ainsi, « [un] modèle apte à réaliser les performances que l'on attend de lui sera dit opérationnel. L'opérationnalité du modèle constitue donc sa qualité fondamentale ». Il y a trois conditions à l'opérationnalité d'un modèle :

Cohérence : interne et d'objectifs. L'aspect « interne » de la cohérence fait référence à un agencement des parties constituantes du modèle en accord rigoureux avec des principes logiques (ce qui est souvent facilité par le recours aux mathématiques). La cohérence d'objectifs veut simplement que l'agencement des parties constituantes du modèle soit orienté selon les objectifs de celui-ci. Bonnafous donne l'exemple d'un modèle dont l'objectif est la prévision de la demande pour lequel « il faudra bien que sa formalisation exprime la détermination de cette demande en fonction de facteurs qui sont, eux, plus aisément prévisibles ».

Pertinence : il s'agit d'une conformité du modèle à ce qu'on croît, à un moment donné, savoir de la réalité.

Mesurabilité : cette condition est nécessaire pour évaluer la pertinence du modèle. Elle doit être vérifiée sur trois aspects de celui-ci : les relations causales, les quantités mises en jeu (variables) et les paramètres.

Afin de caractériser l'objet et les objectifs — la portée — des modèles de prévision de la demande de transport de personnes, il est utile, à ce stade, de proposer une représentation schématique de la ville (inspirée de Bonnafous et Puel, 1983) comme un système composé de trois sous-systèmes. Un sous-système des transports ( $\mathcal{T}$ ) comprenant l'ensemble des infrastructures de transport avec leurs caractéristiques techniques : technologies, capacités, vitesses, types de liens, etc. Un sous-système de localisation des activités ( $\mathcal{L}$ ) comprenant l'inscription spatiale des équipements où les individus réalisent leurs activités. Un sous-système d'activités et relations sociales ( $\mathcal{A}$ ) qui comprend l'ensemble des comportements des individus (qui donnent, en partie, lieu à des déplacements). Bonnel (2004) note que, dans les modèles de prévision de la demande de transport, le sous-système de localisations est considéré

---

1. Guitton (1963), cité par Bonnafous (1989), propose une définition du modèle proche de celle suggérée par Manheim puisqu'il évoque une « construction simplifiée destinée à expliquer la réalité ou à agir sur elle ».

comme exogène. Ces modèles s'intéressent aux relations causales entre sous-système de transport et sous-système d'activités, pour un sous-système de localisations donné. Ces relations causales correspondent donc à leur objet. L'objectif des modèles peut être conçu comme l'évaluation des impacts d'une modification du sous-système des transports ( $\Delta\mathcal{T}$ ) sur le sous-système d'activités (Manheim, 1980). En général, on prend également en compte les modifications du sous-système de localisations ( $\Delta\mathcal{L}$ ), mais celles-ci sont prévues de manière indépendante (par d'autres procédures allant du simple prolongement de tendance à la modélisation fine des comportements de localisation).

### 1.1.2 Cadrage économique de l'analyse de la demande de transport de personnes

Suivant la présentation de Meyer et Straszheim (1971), on peut supposer que, dans une économie de marché, la demande de transport est le résultat des comportements rationnels des agents économiques ; plus précisément, des décisions prises par les ménages et les entreprises visant à maximiser leur utilité ou profit, respectivement. À un moment donné, les agents économiques et les différents biens de l'économie sont répartis dans l'espace (ils ne coexistent pas en un point). Il faut donc que les uns ou les autres se déplacent afin que les besoins des premiers puissent être satisfaits. L'étude de la demande de transport de personnes porte, bien entendu, sur les déplacements réalisés par les personnes.

Il est nécessaire de noter, avant de nous plonger dans la présentation de ces modèles, que l'analyse des transports est une composante relativement réduite de la science économique. En général, les analyses introduisant l'espace contiennent également une certaine forme de coûts de transport. Par exemple, les modèles d'économie géographique comportent des représentations explicites des transports même si celles-ci restent très simples (par exemple, les représentations de type « iceberg » à la Samuelson, voir Dimou, 2003)<sup>2</sup>. Dans le cas particulier des modèles de prévision de trafic, il en est autrement ; les besoins opérationnels font qu'une représentation plus détaillée des caractéristiques des transports est nécessaire (cohérence d'objectifs).

#### 1.1.2.1 Interdépendance entre offre et demande de transport de personnes : le transport comme service

En réalité, il est difficile d'appréhender la demande de transport sans comprendre avant certaines particularités de l'offre. Le transport est un service : il ne peut pas être stocké pour être utilisé, par exemple, pour palier à des périodes de demande plus forte. Il doit être consommé au lieu et au moment de sa production. Pour Ortúzar et Willumsen (2001), de cette propriété fondamentale découle la constitution de  $\mathcal{T}$ , où l'on trouve :

- une infrastructure (par exemple un réseau viaire) ;
- un système de gestion, composé d'un ensemble de règles (par exemple conduire à droite) et de stratégies de régulation (par exemple les feux de signalisation) ;

---

2. Les objectifs de ces modèles, à savoir l'étude des phénomènes auto-entretenus d'agglomération, ne rendent pas nécessaire une description plus fine qui ne changerait pas fondamentalement leurs propriétés principales.

– un ensemble de solutions techniques de transport et leurs opérateurs.

Le service transport possède un certain nombre de caractéristiques propres dont deux sont d'intérêt particulier pour l'analyse de la demande :

- (i) présence attributs non monétaires : dans les transports, des attributs autres que le prix ont un rôle important dans les choix des utilisateurs. Pour Manheim (1980), du point de vue de l'utilisateur (consommateur), l'élément d'intérêt principal d'une solution de transport est son niveau de service, contenant le temps de trajet, le coût, le temps d'attente, la fiabilité des temps, etc. ;
- (ii) existence d'interrelations et effets de rétroaction au niveau des attributs non monétaires : les valeurs de certains attributs non monétaires des transports ne sont pas seulement déterminées par les décisions de l'organisme en charge de l'offre (opérateur) mais elles dépendent des réponses des utilisateurs ou consommateurs (par exemple, l'utilisation massive d'une infrastructure peut conduire à sa congestion et à une augmentation des temps de parcours).

Prenant en compte ces particularités, Bonnel (2004), à la suite de Manheim (1979), propose une formalisation de l'équilibre de court terme dans le marché des transports, sur l'espace défini par le niveau de service et la quantité des déplacements (trafic). La fonction d'offre  $\Omega$  est donnée par :

$$s = \Omega(\mathcal{T}, q) \quad (1.1)$$

où  $s$  est le niveau de service de l'offre de transport,  $\mathcal{T}$  représente l'ensemble des caractéristiques du sous-système de transport et  $q$  est le volume de trafic observé sur  $\mathcal{T}$ . Lorsque  $\mathcal{T}$  est constant,  $s$  décroît à la suite d'une augmentation de  $q$ . La fonction de demande  $\Upsilon$  est donnée par :

$$q = \Upsilon((\mathcal{L}, \mathcal{A}), s) \quad (1.2)$$

où  $\mathcal{L}$  correspond au sous-système de localisation et  $\mathcal{A}$  correspond au sous-système d'activités et relations sociales. Lorsque les conditions de transport se dégradent ( $s$  diminue), les usagers réduisent leur utilisation du système de transport (toute chose égale par ailleurs). Sur l'espace  $(q, s)$ , on a donc un point où offre et demande coïncident : l'équilibre du marché des transports. La figure 1.1 donne une représentation de cet équilibre en 1.1(a). L'équilibre se situe au point de coordonnées  $(q^*, s^*)$ . On note qu'il s'agit d'un cas où les sous-systèmes  $\mathcal{L}$ ,  $\mathcal{A}$  et  $\mathcal{T}$  sont constants. La figure 1.1(b) représente le cas où l'on modifie le sous-système de transport d'un état en 0 vers un état en 1 ( $\Delta\mathcal{T}$ ).

Bonnel remarque que l'enjeu de la modélisation de la demande de transports consiste en la définition des relations de cet équilibre.

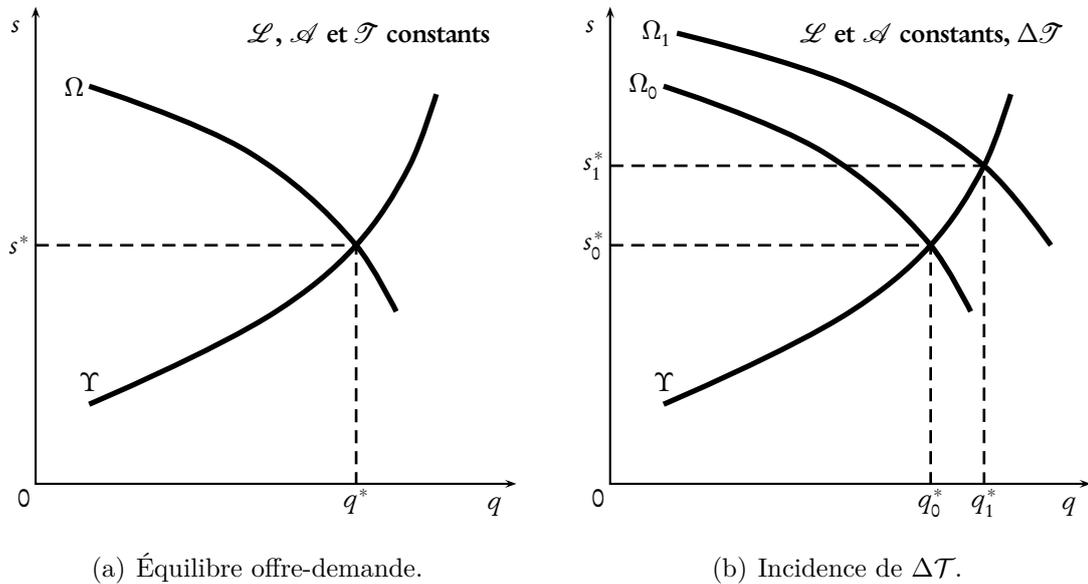
### 1.1.2.2 Espaces et Activités : la demande de transports comme une demande « dérivée »

La nécessité de prendre en compte les composantes non monétaires du transport — en particulier du temps de parcours — fait que, pour l'analyse de la demande de celui-ci, il est convenable de retenir dans un premier temps<sup>3</sup>, en guise de fond

---

3. L'approche plus contemporaine, celle des modèles d'utilité aléatoire et de choix discrets, sera introduite un peu plus loin dans l'exposé (section 1.1.4).

FIGURE 1.1 – Équilibre offre-demande de transport de Manheim.



Source : à partir de Manheim (1979) et Bonnel (2004).

théorique, la « nouvelle théorie du consommateur » (Lancaster, 1966). De manière succincte, Crozet (2005) note qu'au cœur du processus décisionnel des individus on trouve le temps : l'affectation de celui-ci dans les programmes d'activités des agents économiques « conduit à des niveaux d'utilité variables entre lesquels l'*homo œconomicus* est capable d'arbitrer ». L'analyse de l'allocation du temps des consommateurs, depuis la contribution séminale de Becker (1965), fait appel à la notion de commodité (de l'anglais *commodity*). Les commodités sont des combinaisons de biens (ou services) et de temps, de telle sorte que les fonctions d'utilité des consommateurs en dépendent directement. Reprenant la présentation faite par Crozet (pp. 32–33) :

« [...] l'utilité d'un individu ne provient pas seulement de la quantité des biens et services consommés, mais des commodités auxquelles ils correspondent (repas, garde des enfants, soins personnels, partie de golf, soirée au cinéma...). Le consommateur n'est donc pas un être passif, il est en réalité le producteur des commodités qu'il consomme. Or la production de ces commodités nécessite des biens et services, mais aussi du temps. Ces deux inputs sont à relier à deux dotations dont disposent les individus, une dotation en temps et une dotation en monnaie. »

À l'équilibre du consommateur, on obtient les quantités de biens consommées, mais également le temps nécessaire à la production des commodités. Au cœur de l'analyse se trouve l'arbitrage entre les différents biens et dotations monétaire et temporelle des individus. L'exemple classique d'application de cette « nouvelle théorie du consommateur » est celui dans lequel l'arbitrage donne naissance à l'offre de travail : il s'agit de déterminer un temps de loisir et un temps de travail. Une autre application, d'intérêt particulier dans le cadre de ce travail, est celui du temps consacré aux transports.

Tipping (1968) introduit la notion de « consommation intermédiaire »<sup>4</sup> pour appréhender la différence entre le temps consacré à profiter de la consommation d'un bien et le temps rendant cette consommation possible (p. 848) :

*« How much of one's "non-working" time is spent in the maintenance and fuelling of the body in order to keep going as a productive unit (and as a living person)? How should one distinguish between, and measure, the time which is spent in enjoying the consumption of goods and services ("final consumption") and the time which is spent in making such final consumption possible? Perhaps we need a category of "intermediate consumption" to cover such activities as travelling to the theatre, doing the shopping, reading Which? and so on. But whatever the list for one person, it may be quite different for another. Some may regard the journey to work as a curse, a cost to be reduced if possible. Others may regard it as the best part of the day, to be savoured "at leisure". »*

Le caractère de « consommation intermédiaire » donné au transport fait qu'on parle de la demande le concernant comme d'une demande « dérivée ».

Meyer et Straszheim (1971) relaient que les premiers efforts pour estimer des relations définissant la demande de transport reprenaient le cadre marshallien traditionnel d'équilibre partiel. Les flux de demande de transport étaient déterminés en établissant des liens entre *output* et prix, revenu et d'autres variables, à partir d'un agent (utilisateur) représentatif (McFadden, 2001). Ils notent également que ces approches avaient une structure à une seule équation ce qui, dans la pratique, conduisait à exclure de l'analyse un grand nombre de facteurs fondamentaux du comportement des agents, notamment l'espace.

Dans les années 50–60, une approche alternative d'inspiration macroéconomique, combinant une analyse comptable des flux (de type Leontief, 1941) et des relations fondées sur l'agent représentatif de la théorie du consommateur (et de la nouvelle théorie du consommateur), a émergé (Gaudry, 2007). Il s'agit de la séquence de modélisation classique à quatre étapes, que nous présentons dans ce qui suit.

### 1.1.3 Le modèle classique à quatre étapes

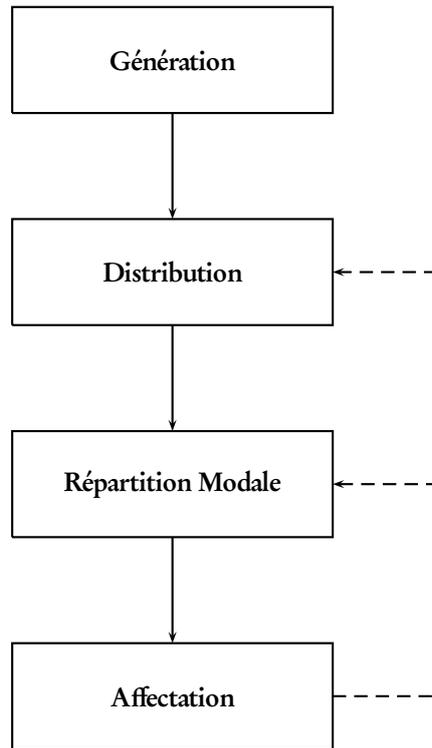
La structure de modélisation en quatre étapes trouve ses origines dans le « fourmillement d'études préparatoires aux développements des réseaux métropolitains et interurbains aux États-Unis durant les années 1950 » (Gaudry, 2007), dont les plus connus sont ceux menés lors du *Chicago Area Transportation Study* (CATS, 1952–1962).

Le modèle à quatre étapes est, à l'origine, un modèle macroéconomique et séquentiel d'analyse des flux de déplacement (Gaudry, 2007). La région étudiée est partitionnée en zones d'analyse de trafic (TAZ, de l'anglais *traffic analysis zones*) au niveau desquelles les flux sont représentés (à l'aide de matrices de flux). Ceci confère au modèle son caractère « macro ». Chaque étape du modèle correspond à l'application d'une relation, relative aux décisions de déplacement (quantité, destination, mode, itinéraire, etc.), aux matrices de flux. Ces relations peuvent avoir été établies

---

4. Il convient de préciser que Becker (1965) notait déjà que son analyse était proche de celle du producteur, à laquelle les « consommations intermédiaires » sont généralement rattachées.

FIGURE 1.2 – Séquence de modélisation classique à quatre étapes



Source : à partir de Ortúzar et Willumsen (2001).

de manière heuristique (l'application du modèle à quatre étapes précède l'apparition de la nouvelle théorie du consommateur), mais elles peuvent également être issues d'une dérivation théorique. Celle-ci prend appui sur la théorie microéconomique standard<sup>5</sup> (on détermine une fonction de demande pour un agent représentatif). Meyer et Straszheim (1971) notent que le format à équations multiples confère au modèle à quatre étapes une flexibilité de représentation des demandes de déplacement sur des réseaux de transport complexes et spatialisés. Elle permet donc des gains de pertinence (dans le sens donné à ce mot par Bonnafous, 1989), par rapport aux premières approches issues du cadre marshallien, au niveau de la représentation des interactions entre offre et demande, évoquées dans la section 1.1.2.1. Elle possède également des avantages en termes de mesurabilité, compte tenu des méthodes de recueil de données pour la modélisation.

La cohérence entre étapes est garantie par les contraintes comptables liées à la représentation matricielle des flux (de manière analogue à celle de l'analyse *input-output*). La figure 1.2 présente la séquence de modélisation traditionnelle. Chacune des étapes est décrite dans ce qui suit.

---

5. Elle peut également prendre appui sur la théorie de l'utilité aléatoire, que nous évoquerons dans la section suivante 1.1.4.

### 1.1.3.1 Génération

La génération est la première étape du modèle à quatre étapes. Elle cherche à quantifier les déplacements réalisés par les résidents de l'aire d'étude. Plus précisément, pour chaque zone de l'aire d'étude, cette étape permet de déterminer le nombre de déplacements — plus généralement, d'unités de mesure des flux — ayant pour origine (on parle d'émissions ou *trip production*, en anglais) et le nombre de déplacements ayant pour destination (on parle d'attractions) la zone considérée. Les émissions et attractions sont déterminées à partir d'un certain nombre de caractéristiques des habitants de chaque zone et des zones elles-mêmes. Dans la séquence des choix individuels, l'aspect « émission » peut être vu comme le choix du nombre de déplacements à réaliser suivant les motifs (Ortúzar et Willumsen, 2001 ; Bonnel, 2004). Formellement, les modèles de génération peuvent être décrits de la manière suivante (Masson, 2000) :

$$T_i = f_o(X_i) \quad (1.3)$$

$$T_j = f_d(X_j) \quad (1.4)$$

où  $T_i$  est le trafic émis par la zone  $i$  et  $T_j$  est le trafic attiré par la zone  $j$ .  $X_i$  et  $X_j$  sont des vecteurs contenant différentes informations relatives aux zones d'origine et de destination  $i$  et  $j$ .  $f_o$  et  $f_d$  sont des fonctions prenant des formes différentes selon les approches de modélisation choisies pour les émissions et attractions. Dans la mesure où les modèles à quatre étapes sont utilisés pour déterminer le volume de déplacements des résidents d'un périmètre, à l'intérieur de ce périmètre, le fait de comptabiliser ce flux à partir des zones d'émission ou à partir des zones d'attraction ne devrait pas le modifier. Il s'en suit qu'on a en général :

$$\sum_{i \in Z} T_i = \sum_{j \in Z} T_j \quad (1.5)$$

où  $Z$  est l'ensemble des zones d'analyse du périmètre. Dit autrement, le total des émissions est égal au total des attractions. Selon la précision permise par les modèles et les données utilisées, il est possible que, dans la pratique, la condition donnée dans l'équation (1.5) ne soit pas respectée. Dans ces cas, il est d'usage de normer les résultats du modèle d'attraction, qui est souvent moins précis, pour que la condition (1.5) soit respectée (on multiplie tous les  $T_j$  par le rapport entre la somme des émissions et la somme des attractions).

### 1.1.3.2 Distribution

La distribution est la deuxième (la troisième dans certains modèles, voir Bonnel, 2004) étape du modèle à quatre étapes. Elle cherche à établir un lien entre les déplacements émis et attirés par les différentes zones de l'aire d'étude. Plus précisément, pour chaque paire de zones de l'aire d'étude, cette étape permet de déterminer le nombre de déplacements — plus généralement, d'unités de mesure des flux — ayant pour origine l'une des zones et pour destination l'autre. On obtient ainsi une matrice dite « origine-destination » (ou matrice O-D). Dans la séquence de choix individuels, elle correspond au choix de la destination (Bonnel, 2004).

Il existe plusieurs familles de modèles permettant de traiter la distribution des déplacements. Une distinction fondamentale avec la montée en puissance des approches dérivées de la théorie de l'utilité aléatoire est celle entre l'approche agrégée, s'appuyant très largement sur des modèles de type gravitaire, et l'approche désagrégée, faisant appel à des modèles de choix discret pour déterminer les choix de destination<sup>6</sup> (Ortúzar et Willumsen, 2001 ; Bonnel, 2004).

Une des principales explications à l'utilisation généralisée de modèles gravitaires dans les exercices de planification peut sans doute être trouvée dans leur apparente simplicité conceptuelle. L'appellation « gravitaire » découle, justement, du fait que le principe de ces modèles peut être compris intuitivement par son analogie avec la loi de la gravitation universelle de Newton, que l'on peut énoncer ainsi : deux corps de masses respectives  $M_A$  et  $M_B$  s'attirent avec une force proportionnelle à chacune de ces masses, et inversement proportionnelle au carré de la distance qui les sépare. Dans la transposition de ce principe au domaine des déplacements urbains, à l'attraction entre deux corps se substitue un autre type d'interaction : le nombre de déplacements émis,  $O$ , et attirés,  $D$ , par les différentes zones de l'aire étudiée. À la place de la distance au carré, on trouve une fonction du temps ou du coût généralisé (caractérisant le niveau de service des transports), qui traduit le degré de « difficulté » à ce qu'il y ait des interactions entre les différentes paires de zones.

Formellement, la relation fondamentale de la distribution peut être définie de la manière suivante :

$$T_{ij} = f_{od}(O_i, D_j, c_{ij}) \quad (1.6)$$

où  $T_{ij}$  est le nombre de déplacements entre une origine  $i$  et une destination  $j$ .  $c_{ij}$  est un vecteur regroupant les caractéristiques de l'offre de transport<sup>7</sup> entre  $i$  et  $j$ , comme le temps ou le coût. En prévision,  $O_i$  et  $D_j$  sont remplacés par les flux modélisés lors de la génération  $T_i$  et  $T_j$ , ce qui correspond à la flèche entre l'étape de génération et celle de distribution de la figure 1.2.

### 1.1.3.3 Répartition modale

La phase de répartition modale a comme objectif de répartir les déplacements entre les différents modes de transport. Lorsqu'elle est placée à la suite de la phase de génération, elle permet d'obtenir des vecteurs d'émission et attraction pour chacun des modes retenus. Lorsqu'elle est placée en troisième position (ce qui est le cas le plus souvent), à la suite de la distribution, elle permet d'obtenir des matrices origine-destination par mode. Dans la séquence des choix individuels, elle correspond au choix du mode de déplacement (Bonnel, 2004).

---

6. Le modèle gravitaire peut être dérivé par la théorie de l'utilité aléatoire (voir section 1.2.2). Cependant, le traitement des origines et des destinations implique une agrégation au niveau zonal et le calage se fait à partir de données zonales (matrices O-D observées). Il convient de noter le développement rapide des approches dérivées de l'intelligence computationnelle — logique floue, réseaux de neurones, entre autres — dans le champ des transports et spécifiquement dans l'étude de la distribution des déplacements (Openshaw, 1998).

7. Si la distribution est placée en deuxième place,  $c_{ij}$  représente les caractéristiques de l'ensemble des modes de transport retenus pour la modélisation. Si la distribution est placée en troisième place, après la répartition modale, il y a un modèle de distribution par mode.

Formellement, si l'on modélise deux modes de transport,  $k$  et  $k'$ , la répartition modale peut être modélisée de la manière suivante :

$$\left. \begin{array}{l} T_{ijk} \\ T_{ijk'} \end{array} \right\} = f_{kk'}(T_{ij}, X, c_{ijk}, c_{ijk'}) \quad (1.7)$$

où  $T_{ijk}$  et  $T_{ijk'}$  sont les déplacements entre  $i$  et  $j$  utilisant les modes  $k$  et  $k'$ , respectivement.  $X$  est un vecteurs contenant différentes caractéristiques de la population (attributs socioéconomiques principalement, à l'échelle zonale ou individuelle). La flèche reliant distribution et répartition modale provient du fait que la matrice O-D tous modes confondus, issue de la distribution, est en entrée de la phase de répartition modale. On note, par ailleurs, que les caractéristiques des modes ont été données de manière séparée, par mode. En effet, dans la distribution, les caractéristiques des modes sont en général combinées au sein d'un même indicateur. La procédure d'agrégation doit idéalement tenir compte du lien entre distribution et répartition modale<sup>8</sup>, pour maintenir une cohérence interne, mais ce n'est pas toujours le cas.

#### 1.1.3.4 Affectation

La phase d'affectation a comme objectif de déterminer le volume des flux de déplacements sur chacun des itinéraires possibles entre une origine et une destination données, pour chaque mode modélisé. Dans la séquence des choix individuels, l'affectation correspond au choix du chemin. Or, comme nous l'avons noté dans le point 1.1.2.1, le fait qu'une quantité de flux passe par un chemin particulier peut avoir un impact sur le niveau de service effectif de ce chemin. En réalité, c'est au niveau de l'étape d'affectation que se fait la confrontation entre caractéristiques de l'offre fixées par l'opérateur et demande, et où l'équilibre du marché des transports est fixé par le biais du niveau de service effectif obtenu.

Formellement, la représentation de cette étape est similaire à celle de la répartition modale. Si, pour le mode  $k$  on peut aller de  $i$  à  $j$  par deux chemins  $\xi$  et  $\xi'$ , alors la procédure d'affectation pour ce même mode peut être définie de la manière suivante :

$$\left. \begin{array}{l} T_{ijk\xi} \\ T_{ijk\xi'} \end{array} \right\} = f_{k\xi\xi'}(T_{ijk}, c_{ijk\xi}, c_{ijk\xi'}) \quad (1.8)$$

Pour prendre en compte l'impact de la demande sur le niveau de service, il est nécessaire de compléter la relation donnée par l'équation (1.8) par la relation suivante :

$$\left. \begin{array}{l} c_{ijk\xi} \\ c_{ijk\xi'} \end{array} \right\} = f_{k\xi\xi'\text{retro}}(T_{ijk\xi}, T_{ijk\xi'}) \quad (1.9)$$

L'affectation permet donc de déterminer la demande qui s'adresse aux différentes infrastructures de transport. Elle permet, simultanément, de déterminer le niveau de service effectif lié à chaque infrastructure. Ce niveau de service effectif est une donnée en entrée des étapes de distribution et de répartition modale, ce qui, dans la figure 1.2, correspond aux flèches reliant l'affectation à ces deux étapes. Notons que les flèches sont en pointillés car la rétroaction n'est pas toujours prise en compte.

---

8. En prenant une forme de type « log-somme ».

Notons également qu'il n'y a pas de flèche reliant l'affectation à la génération, ce qui suppose une indépendance entre ces deux étapes. Il est possible d'établir une relation entre les deux, mais la plupart des travaux empiriques vont plutôt dans le sens d'une corroboration de l'indépendance (Kockelman, 1998).

#### 1.1.4 Utilité aléatoire : clivage entre une modélisation agrégée ou désagrégée

L'évolution des besoins de planification en lien avec l'évolution des politiques de transport a conduit révéler certaines limites d'une modélisation purement agrégée. Des « programmes d'amélioration de la modélisation des transports », plus ou moins officiels (McNally, 2000b), ont vu le jour afin de répondre aux interrogations qui se sont ajoutés aux problèmes classiques d'investissement et de congestion : efficacité globale du système de transport, sécurité, pollution. . . Parmi les raffinements qui en résultèrent, on peut citer le développement des modèles désagrégés de choix discrets (Domencich et McFadden, 1975 ; Ben-Akiva et Lerman, 1985).

McFadden (2001) retrace la genèse des théories de l'utilité aléatoire. Il explique que, la théorie du consommateur est avant tout un outil logique pour explorer les propriétés d'organisations de marché ou de politiques alternatives, d'un point de vue conceptuel. Lorsqu'on essaye de les appliquer à des jeux de données, on est limité à ne reproduire le comportement des agents que de manière très agrégée et on ne tient pas compte de l'hétérogénéité des agents. Par ailleurs, le choix des distributions des erreurs ne peut pas être validé formellement et n'a pas un fondement théorique. Le développement de l'informatique et des bases de données a permis de se focaliser sur les variations de demande entre individus. Il était nécessaire d'expliquer la variabilité des comportements au sein de la théorie du consommateur. Ceci a donné naissance à la théorie de l'utilité aléatoire, sur laquelle les modèles de choix discrets sont fondés.

Ben-Akiva et Lerman (1985) expliquent que la théorie standard et la nouvelle théorie du consommateur font l'hypothèse que l'espace des alternatives de consommation est continu, ce qui permet d'utiliser les outils d'analyse (*calculus* en anglais) traditionnels pour déterminer les fonctions de demande. *A contrario*, comme leur nom l'indique, les modèles de choix discrets décrivent les choix d'une unité de prise de décision — il peut s'agir d'un individu, d'un ménage, d'une entreprise, etc. — parmi certaines alternatives, situées dans un espace discret. Plus précisément, l'ensemble des alternatives, appelé ensemble de choix, doit respecter trois critères (Train, 2009) :

- (i) les alternatives doivent être mutuellement exclusives, du point de vue du décideur. Dit autrement, le choix d'une alternative implique, nécessairement, le non choix des autres alternatives et le décideur ne peut choisir qu'une seule des alternatives.
- (ii) l'ensemble de choix doit être exhaustif, c'est-à-dire que toutes les alternatives possibles y appartiennent. Le décideur choisit nécessairement une des alternatives de l'ensemble de choix.
- (iii) le nombre d'alternatives doit être fini.

Les conditions de choix dans un ensemble tel que défini ci-dessus peuvent être rapprochées d'un cas particulier de l'analyse traditionnelle du consommateur (Ben-Akiva et Lerman, 1985) : les solutions « en coin », où les conditions d'optimalité usuelles

ne sont pas valides en raison de la présence de biens qui peuvent prendre des valeurs nulles. On peut donner un exemple avec 2 biens 1 et 2. La fonction d'utilité est donnée par :

$$U = U(q_1, q_2) \quad (1.10)$$

où  $q_1$  et  $q_2$  sont les quantités de bien 1 et 2, respectivement, et on impose les restrictions suivantes :

$$q_1 = \begin{cases} 1 & \text{si le bien 1 est choisi,} \\ 0 & \text{sinon;} \end{cases} \quad (1.11)$$

$$q_2 = \begin{cases} 1 & \text{si le bien 2 est choisi,} \\ 0 & \text{sinon.} \end{cases} \quad (1.12)$$

On impose également l'exclusivité mutuelle des alternatives :

$$q_1 q_2 = 0. \quad (1.13)$$

Dans ces circonstances, la fonction d'utilité donnée dans l'équation (1.10) ne peut prendre que deux valeurs, correspondant aux deux « coins »,  $U_1 = U(1, 0)$  et  $U_2 = U(0, 1)$ . Le choix optimal résulte donc de la comparaison entre  $U_1$  et  $U_2$ .

Les modèles de choix discrets introduisent une part d'incertitude dans les choix des consommateurs. Les modèles d'utilité aléatoire sont dérivés de la manière suivante (le reste de ce paragraphe emprunte grandement à Train, 2009). On commence par une généralisation de l'exemple à deux biens avec des solutions en coin. Un décideur  $h$  est confronté à un choix parmi  $K$  alternatives. L'alternative  $k \in \{1 \dots K\}$  lui procure l'utilité  $U_{hk}$ . Comme précédemment, le décideur choisit une alternative  $k$  particulière si et seulement si  $U_{hk} > U_{hk'} \forall k' \neq k$  avec  $k, k' \in \{1 \dots K\}$ . Le niveau d'utilité est connu du décideur mais pas de l'analyste. Celui-ci ne peut observer que quelques attributs des alternatives ( $x_{hk}$ ) et quelques attributs du décideur ( $r_h$ ). Il modélise donc l'utilité du consommateur en reliant les attributs qu'il observe au sein d'une fonction d'utilité représentative  $V_{hk} = V(x_{hk}, r_h) \forall k$ . Comme ce modèle n'est qu'une approximation de la réalité,  $U_{hk} \neq V_{hk}$  ou encore  $U_{hk} = V_{hk} + \epsilon_{hk}$  où  $\epsilon_{hk}$  représente les éléments qui ont un impact sur l'utilité du consommateur mais qui ne sont pas appréhendés par  $V_{hk}$ . L'analyste ne connaissant pas  $\epsilon_{hk}$ , il traite ces termes comme étant aléatoires. La densité jointe du vecteur aléatoire  $\epsilon_h \equiv \langle \epsilon_{h1} \dots \epsilon_{hK} \rangle$  est notée  $f(\epsilon_h)$ . Avec celle-ci, l'analyste peut étudier les choix des décideurs en termes probabilistes. Ainsi, la probabilité que le décideur  $h$  choisisse l'alternative  $k$  est :

$$\begin{aligned} P_{hk} &= \Pr(U_{hk} > U_{hk'} \forall k' \neq k) \\ &= \Pr(V_{hk} + \epsilon_{hk} > V_{hk'} + \epsilon_{hk'} \forall k' \neq k) \\ &= \Pr(\epsilon_{hk'} - \epsilon_{hk} < V_{hk} - V_{hk'} \forall k' \neq k). \end{aligned} \quad (1.14)$$

En utilisant la fonction de densité de  $\epsilon_h$ , l'équation (1.14) devient

$$P_{hk} = \int_{\epsilon} I(\epsilon_{hk'} - \epsilon_{hk} < V_{hk} - V_{hk'} \forall k' \neq k) f(\epsilon_h) d\epsilon_h \quad (1.15)$$

où  $I(\cdot)$  est la fonction indicatrice égale à 1 lorsque l'expression entre parenthèses est vraie et 0 autrement. Les différents modèles de choix discrets proviennent des hypothèses faites par rapport à la forme de  $f(\epsilon_h)$ .

## 1.2 Familles de modèles pour la spécification des quatre étapes

Les principes de l'analyse économique de la demande de transport, présentés dans la section précédente, peuvent être déclinés dans plusieurs types de modèles<sup>9</sup>. Notons qu'en théorie, il est possible de se passer d'une structure séquentielle à quatre étapes et de modéliser la demande de transport au sein d'un modèle intégré, s'appuyant sur la théorie de l'utilité aléatoire (Kockelman, 1998). Cependant, dans la plupart des applications pratiques, les modèles de choix discrets sont utilisés pour modéliser une ou plusieurs étapes, sans remettre en cause la structure séquentielle.

Dans ce travail, nous nous limitons à l'étude des modèles à quatre étapes classiques (où la génération et la distribution ne sont pas issues de la théorie de l'utilité aléatoire), parce qu'il s'agit des plus utilisés (voir CERTU, 2002 ; TRB, 2007). Dans cette section, nous présentons les familles des modèles qui seront étudiés dans les chapitres 4 (génération), 5 (distribution) et 6 (répartition modale). Nous évoquons également la phase d'affectation même si ce travail ne couvre pas les aspects temporels de celle-ci en raison du manque de données (voir chapitre 3). La présentation de cette phase est en revanche moins détaillée.

### 1.2.1 Génération

Il existe, une multiplicité d'approches permettant d'obtenir les quantités de déplacements émis et attirés par chaque zone. Le lecteur peut se référer aux différents manuels de modélisation (Ortúzar et Willumsen, 2001 ; Bonnel, 2004) pour une présentation détaillée de ces approches. Dans la pratique actuelle de la modélisation, on a recours le plus souvent aux techniques de classification, et cela malgré le fait qu'on n'a pas de preuve indiscutable de la supériorité de ces méthodes par rapport à celles par régression.

#### 1.2.1.1 Classification croisée

Il s'agit de la méthode de modélisation qu'on accepte généralement comme étant la plus simple et, de ce fait, elle est sans doute la plus répandue, notamment dans le contexte américain (McNally, 2000b ; Badoe et Mwakalonge, 2011). Elle consiste à partitionner l'échantillon de ménages ou de personnes en un certain nombre de catégories à partir d'une ou, le plus souvent, deux variables socioéconomiques. On fait l'hypothèse que les groupes ainsi obtenus sont relativement homogènes à l'intérieur de chaque catégorie et relativement différents d'une catégorie à une autre, en termes de comportements de mobilité<sup>10</sup>. Le modèle, consiste à calculer le nombre de déplacements ou de sorties réalisés en moyenne par les ménages ou personnes de chaque segment de population. Ce nombre de déplacements ou sorties moyen sera appelé dans ce qui suit, de manière indistincte, coefficient ou taux d'émission (de

---

9. Le lecteur souhaitant avoir un bilan comparatif des différentes approches peut consulter Hensher et Button (2005).

10. Qui, bien entendu, correspondent ici à une quantité moyenne de déplacements ou de sorties (selon l'unité de mesure de la mobilité retenue, voir Axhausen, 2000 ; Bonnel, 2004) réalisée au cours de la période étudiée.

l'anglais *trip rates*). Ces coefficients d'émission peuvent ensuite être appliqués aux différents contextes et scénarios envisagés dans l'analyse. Formellement, ces modèles peuvent être décrits de la manière suivante (Doubleday, 1977) :

$$\theta_{kp} = \frac{\sum_{h \in H_k} N_{hkp}}{n(H_k)} \quad (1.16)$$

où  $\theta_{kp}$  est le taux d'émission pour le motif  $p$  des unités décisionnelles  $h$  appartenant au segment de population  $H_k$ .  $N_{hkp}$  est le nombre de déplacements (ou sorties) pour le motif  $p$  réalisés par l'unité décisionnelle  $h$  appartenant au segment de population  $H_k$ .  $n(H_k)$  correspond à l'effectif du segment de population  $H_k$ .

Le caractère non paramétrique de la méthode<sup>11</sup> lui confère une souplesse dans le choix des variables explicatives de par le fait qu'il n'est pas nécessaire d'assumer au préalable une forme particulière pour la relation entre celle-ci et la variable à expliquer.

La limite principale de la méthode se situe au niveau de l'effectif de chaque catégorie. En effet, si l'augmentation du nombre de catégories se traduit par une homogénéité plus importante de celles-ci, elle conduit également, pour un effectif enquêté total donné, à réduire l'effectif de chaque classe et donc à rendre les coefficients calculés moins fiables. La définition des classes peut conduire, en outre, à avoir des problèmes d'effectif insuffisant seulement pour certaines d'entre elles, ce qui peut conduire à des progressions contre-intuitives dans les coefficients des différentes classes, pour une variable déterminée (Mwakalonge, 2010). Le choix du nombre des variables et des modalités de celles-ci répond donc plus ou moins à un arbitrage entre nombre de classes, homogénéité et effectif de chaque classe.

### 1.2.1.2 Régression linéaire

L'approche de modélisation de la génération par régression linéaire, à l'échelle zonale ou de l'unité de prise de décision (ménage, individu), consiste à définir une relation entre le nombre total de déplacements ou de sorties (selon l'unité de mesure de la mobilité retenue), qui est la variable à expliquer, et une ou plusieurs variables explicatives (démographiques, socioéconomiques, caractérisant l'utilisation du sol, etc.). Les variables explicatives doivent être mises en forme de sorte que la relation entre celles-ci et la variable à expliquer puisse être envisagée comme étant plausiblement linéaire.

Les modèles de régression linéaire étant parmi les techniques les plus utilisées et les mieux connues en économétrie, nous ne détaillons pas leurs caractéristiques ici<sup>12</sup>. Nous nous limitons à rappeler que la spécification d'un modèle de ce type est rendue délicate en présence de corrélation entre les variables explicatives retenues et par la présence possible d'hétéroscédasticité (notamment lorsque le modèle est spécifié à l'échelle zonale, puisqu'il y a un effet « taille de la zone » qui peut jouer).

---

11. Elle n'assume aucune relation de distribution probabiliste entre la variable à expliquer et les variables explicatives.

12. Nous renvoyons le lecteur aux nombreux manuels d'économétrie disponibles (par exemple Guitton, 1963 ; Greene, 2011).

### 1.2.1.3 Les arbres de segmentation (classification ou régression)

Strambi et van de Bilt (1998) remarquent que la spécification des modèles de génération peut être une affaire d'une difficulté croissante si elle n'est pas abordée par une forme de processus systématique. Ils distinguent trois éléments à prendre en compte : (i) la sélection des variables pertinentes, (ii) les choix de catégorisation des variables et (iii) le traitement des interactions potentielles entre variables explicatives. Les deux méthodes de modélisation présentées précédemment ont des limites à des niveaux différents de la procédure de spécification. En effet, selon Strambi et van de Bilt, il ne semble pas y avoir dans la littérature de procédure optimale de sélection et de catégorisation des variables pour les modèles de classification, malgré quelques recherches dans ce sens (Stopher et McDonald, 1983 ; Supernak *et al.*, 1983, par exemple). En revanche, implicitement dans ces modèles, toutes les interactions entre variables explicatives sont traitées. Pour les modèles de régression, on peut faire appel aux différentes approches de type *stepwise* pour la sélection des variables, mais l'inclusion d'effets d'interaction entre les différentes variables explicatives est difficile.

Les procédures de type arbres de classification ou régression permettent de concilier, d'une certaine manière, les avantages des deux mondes. Le principe de ces méthodes est basé sur une approche non paramétrique et consiste en un fractionnement successif de la population selon une structure arborescente. Il existe plusieurs méthodes de ce type, mais celle retenue pour cette étude est l'algorithme CART (*Classification And Regression Trees*, Breiman *et al.*, 1984). À chaque étape de l'algorithme, l'objectif est de sélectionner, au sein d'un groupe de variables potentiellement explicatives des comportements de mobilité, celles qui sont les plus discriminantes possibles selon un certain critère. Le critère dépend de la forme de la variable à expliquer. L'algorithme CART produit des arbres binaires<sup>13</sup> : il évalue l'ensemble des divisions (binaires) de la population possibles compte tenu des variables explicatives retenues et choisit celle dont le pouvoir discriminant est le plus élevé. Dans le cas où la variable à expliquer est une variable numérique, le critère de choix est le coefficient de détermination  $R^2$  (et donne un arbre de régression). À chaque étape de l'algorithme, la division retenue sera celle qui maximise le  $R^2$ . Le fonctionnement est donc très proche de ce qui peut être fait dans une régression de type *stepwise*. Dans le cas où la variable à expliquer est qualitative, la méthode CART utilise l'indice de Gini (et donne un arbre de classification). Prenons l'exemple d'une variable à expliquer qui répartit les individus en  $M$  modalités. Si on note  $H_k$  un segment de la population,  $n(H_k)$  l'effectif du segment et  $n_m(H_k)$  l'effectif pour lequel la variable à expliquer prend la modalité  $m$ , de sorte que  $n(H_k) = \sum_{m=1}^M n_m(H_k)$ , alors la mesure d'impureté est donnée par :

$$I(H_k) = \sum_{m=1}^M \frac{n_m(H_k)}{n(H_k)} \left( 1 - \frac{n_m(H_k)}{n(H_k)} \right) \quad (1.17)$$

La fonction  $I(H_k)$  donné par l'équation (1.17) est l'indice de diversité de Gini. Un segment est dit « pur » s'il contient des individus d'une seule classe. Dans un tel cas,

---

13. D'autres méthodes comme l'algorithme CHAID (*CHI-squared Automatic Interaction Detector*, Kass, 1980), utilisé dans le cadre de modèles de génération par Strambi et van de Bilt (1998) par exemple, produisent des divisions plus nombreuses à chaque étape.

$I(H_k) = 0$ . Plus le mélange des classes dans le segment  $H_k$  est important est plus  $I(H_k)$  est élevé. Ainsi, la sélection d'une division doit être telle que les segments descendants soient plus « purs » dans leur composition en termes de variable à expliquer, que le nœud (ou segment) parent. Autrement dit, il faut que le mélange des classes soit moins important dans les segments descendants que dans le nœud parent. Cette méthode fournit *a priori* un résultat proche de celui fourni par une classification ascendante hiérarchique. Cependant, elle est plus proche des méthodes de régression et de classification classiques que des méthodes non supervisées de classification. En effet, il ne s'agit pas de faire apparaître des classes homogènes dans l'absolu mais de chercher des groupes d'individus les plus explicatifs des modalités d'une variable qualitative (cas de la variable nominale à expliquer). Le principe est de chercher à chaque étape la dichotomie la plus liée à la variable privilégiée.

### 1.2.2 Distribution

Dans le cadre du présent travail, nous avons pris le parti d'utiliser une approche agrégée pour la modélisation des flux par origine-destination. En réalité, la modélisation du choix de la destination est formellement très proche de celle du choix du mode et l'analyse de sa transférabilité risque aussi de l'être. Nous avons donc décidé de ne considérer l'approche désagrégée que dans le cadre de l'étude du choix du mode et de nous intéresser aux modèles gravitaires pour la modélisation des flux par origine-destination.

Il existe plusieurs approches proposant des fondements théoriques au modèle gravitaire — comme l'approche par l'état le plus probable (Wilson, 1967), connue encore sous le nom de maximisation de l'entropie, les approches qui assument une loi de probabilité, multinomiale ou, plus récemment, de poisson (Sen et Smith, 1995), pour les différentes O-D, associées à un principe d'efficience, les approches de minimisation des coûts de transport (Erlander, 1977), les approches de maximisation de l'utilité aléatoire (Cochrane, 1975 ; Anas, 1983) ou encore les approches par minimisation de l'information de discrimination (Snickars et Weibull, 1977), parmi d'autres. Une présentation détaillée de l'ensemble de ces approches dépasserait les ambitions de cette étude<sup>14</sup>. Nous nous limitons ici à faire la présentation issue de la maximisation de l'utilité aléatoire dans le cadre d'un modèle multinomial, en empruntant à Cochrane (1975) et Bonnel (2004). Ce choix vient du fait que, comme le notent Anas (1983), Erlander et Stewart (1990) ou encore Ortúzar et Willumsen (2001), même si les différentes approches de dérivation sont souvent équivalentes d'un point de vue purement formel (dans le sens de formule mathématique), le cadre théorique peut permettre une interprétation plus fine des résultats obtenus. Par ailleurs, l'inscription disciplinaire du présent travail conduit également au choix de l'approche par maximisation de l'utilité aléatoire<sup>15</sup>.

---

14. Nous référons le lecteur intéressé par ces questions au remarquable manuel d'Erlander et Stewart (1990), entièrement consacré au modèle gravitaire dans la modélisation transport, où il trouvera une présentation complète des différentes approches théoriques, ainsi qu'une très instructive « fiction épistémologique » permettant de mettre en évidence les avantages, limites et contradictions internes de chacune des approches.

15. La dérivation classique par le principe de maximisation de l'entropie est donnée dans l'annexe A.1 (page 209).

Dans un premier temps, Cochrane s'intéresse au modèle gravitaire simplement contraint<sup>16</sup>, c'est-à-dire un modèle où le nombre de déplacements émis par chaque zone est fixé au préalable (les modèles simplement contraints sont en général utilisés pour modéliser les déplacements pour le motif achat-services). Pour les zones de destination, on suppose qu'il y a un facteur qui détermine le nombre relatif de possibilités de déplacement offertes par chacune d'entre elles. Toutes ces possibilités de déplacement sont ouvertes à tous les voyageurs. On suppose que les différentes possibilités de déplacements vers les zones d'attraction font partie de la même population, ce qui en termes mathématiques permet de supposer que la probabilité qu'une possibilité de déplacement offre un niveau d'utilité, en termes monétaires, compris dans un certain intervalle est donnée par une seule fonction de densité. On suppose, par ailleurs, qu'il est possible d'estimer l'équivalent monétaire de la désutilité de chaque possibilité de déplacement (sous la forme d'un coût généralisé). Il est donc possible d'estimer la probabilité avec laquelle le surplus (utilité monétaire nette des coûts de déplacement) soit compris entre deux valeurs. La probabilité que le déplacement optimal soit vers une zone particulière s'accroît avec le nombre d'opportunités de cette zone et décroît avec la distance (qui implique un coût supérieur).

Formellement, en suivant la présentation faite par Bonnel (2004), on peut écrire l'utilité d'une possibilité de déplacement de la zone  $i$  vers la destination potentielle  $j_k$ , pour un individu  $h$ , de la manière suivante :

$$U_{hj_k} = V - c_{ij} + \epsilon_{hj_k} \quad (1.18)$$

où  $V$  est une quantité positive,  $c_{ij}$  est le coût généralisé nécessaire à  $h$  pour se rendre de  $i$  à  $j$  et  $\epsilon_{hj_k}$  est un terme aléatoire. Les origines et destinations appartiennent à l'ensemble  $Z$ .

Pour résoudre le problème, on postule que celui-ci est indépendant et identiquement distribué pour l'ensemble de la population et qu'il suit une loi de Gumbel de paramètres  $(0, \mu)$ , de telle sorte que le maximum de  $U_{hj_k}$  au sein de la zone  $j$ , noté  $U_{hj}$ , suit lui aussi une loi de Gumbel de paramètres

$$\left( \frac{1}{\mu} \left( \sum_k e^{\mu(V - c_{ij})} \right), \mu \right). \quad (1.19)$$

Le terme  $\mu(V - c_{ij})$  est constant pour une zone donnée. Le nombre de destinations potentielles dans la zone  $j$  est fonction des opportunités de cette zone et égal à l'attraction de  $j$ ,  $D_j$ <sup>17</sup>, si l'on a utilisé un modèle de génération. L'équation (1.19) devient :

$$\left( \frac{1}{\mu} (\log(D_j) + \mu(V - c_{ij})), \mu \right). \quad (1.20)$$

---

16. Les contraintes aux marges du modèle gravitaire imposent des restrictions sur les déplacements de la matrice origine-destination modélisée. Dans le cas du modèle simplement contraint, les sommes par ligne de la matrice O-D modélisée doivent être égales aux émissions observées (il s'agit d'un modèle contraint à l'origine, les modèles contraints à la destination étant plus rares dans la modélisation des déplacements de personnes). Dans le cas du modèle doublement contraint, la contrainte s'applique à l'origine et à la destination.

17. Notons que, dans la section précédente  $T_j$  correspond aux attractions modélisées alors que, ici,  $D_j$  correspond aux attractions observées.

Si on pose

$$V_j = \frac{1}{\mu} (\log(D_j) + \mu(V - c_{ij})) \quad (1.21)$$

on peut écrire

$$U_{hj} = V_j + \epsilon_{hj} \quad (1.22)$$

où  $\epsilon_{hj}$  est un terme aléatoire suivant une loi de Gumbel de paramètres  $(0, \mu)$ .

La probabilité que  $j$  soit la destination optimale  $P_j$  devient alors (voir la dérivation de cette probabilité dans la section suivante sur la répartition modale) :

$$\begin{aligned} P_j &= \frac{e^{\mu V_j}}{\sum_{j' \in Z} e^{\mu V_{j'}}} = \frac{e^{\mu \left( \frac{1}{\mu} (\log(D_j) + \mu(V - c_{ij})) \right)}}{\sum_{j' \in Z} e^{\mu \left( \frac{1}{\mu} (\log(D_{j'}) + \mu(V - c_{ij'})) \right)}} \\ &= \frac{D_j e^{\mu(V - c_{ij})}}{\sum_{j' \in Z} D_{j'} e^{\mu(V - c_{ij'})}} = \frac{D_j e^{-\mu c_{ij}}}{\sum_{j' \in Z} D_{j'} e^{-\mu c_{ij'}}}. \end{aligned} \quad (1.23)$$

Le nombre de déplacements allant de la zone  $i$  vers la zone  $j$  est donc donnée par :

$$T_{ij} = O_i \cdot P_j = \frac{O_i D_j e^{-\mu c_{ij}}}{\sum_{j' \in Z} D_{j'} e^{-\mu c_{ij'}}}. \quad (1.24)$$

L'équation (1.24) est celle correspondant au modèle simplement contraint. Pour passer du modèle gravitaire simplement contraint au modèle gravitaire doublement contraint, qui reflète mieux le cas où les opportunités à la destination sont exclusives (un commerce peut être fréquenté par plusieurs personnes alors qu'un emploi ne peut être occupé que par une personne), on corrige l'utilité par un facteur additif  $\lambda_j$ . L'équation (1.21) devient alors :

$$V_j = \frac{1}{\mu} (\log(D_j) + \mu(V - c_{ij} - \lambda_j)) \quad (1.25)$$

et l'équation (1.24) devient :

$$T_{ij} = \frac{O_i D_j e^{-\mu(c_{ij} + \lambda_j)}}{\sum_{j' \in Z} D_{j'} e^{-\mu(c_{ij'} + \lambda_{j'})}} = \frac{O_i D_j e^{-\mu \lambda_j} e^{-\mu c_{ij}}}{\sum_{j' \in Z} D_{j'} e^{-\mu \lambda_{j'}} e^{-\mu c_{ij'}}}. \quad (1.26)$$

En appliquant un changement de variable, on obtient la formulation plus habituelle du modèle gravitaire doublement contraint donnée par :

$$T_{ij} = A_i O_i B_j D_j \exp(-\mu c_{ij}) \quad (1.27)$$

$$A_i = \left( \sum_j B_j D_j \exp(-\mu c_{ij}) \right)^{-1} \quad (1.28)$$

$$B_j = \left( \sum_i A_i O_i \exp(-\mu c_{ij}) \right)^{-1}. \quad (1.29)$$

Le changement de variable appliqué est  $e^{-\mu \lambda} = B_j$ .

Il convient de noter que la forme de la désutilité induite par les coûts de déplacement donne par la suite la forme de la fonction de résistance (ici exponentielle négative de paramètre  $\mu$ ). Par ailleurs, même si la dérivation du modèle emprunte à la théorie des choix discrets, son estimation se fait à partir de données agrégées à l'échelle zonale (à partir de matrices origine-destination observées).

### 1.2.3 Répartition modale à partir d'une modélisation désagrégée

Pour le présent travail, nous avons retenu le modèle de choix discrets de type logit multinomial (McFadden, 1974), en raison de son usage répandu<sup>18</sup>. Ce choix permet également d'explorer la capacité prédictive de ces modèles dans un cadre de choix simple en raison du nombre d'alternatives limité (par rapport à un modèle de choix de destination, par exemple).

Pour dériver le logit multinomial, on fait l'hypothèse que l'utilité non observée des unités de décision est indépendante et identiquement distribuée (hypothèse iid) et que sa distribution suit une loi de Gumbel. Ainsi, la densité pour chaque élément de l'utilité non observée s'exprime de la manière suivante (cette présentation des modèles mlogit multinomiaux emprunte à Train, 2009) :

$$f(\epsilon_{hk}) = e^{-\epsilon_{hk}} \exp(-e^{-\epsilon_{hk}}), \quad (1.30)$$

où  $h$  est l'unité de prise de décision et  $k$  est une des alternatives — soit un des modes de transport — parmi lesquelles le choix est réalisé. La distribution cumulative correspondante est

$$F(\epsilon_{hk}) = \exp(-e^{-\epsilon_{hk}}). \quad (1.31)$$

La distribution de la différence de deux variables distribuées selon la loi de Gumbel est logistique. Dit autrement, si l'on note  $\epsilon_{hkk'}^* = \epsilon_{hk} - \epsilon_{hk'}$  alors

$$F(\epsilon_{hkk'}^*) = \frac{e^{\epsilon_{hkk'}^*}}{1 + e^{\epsilon_{hkk'}^*}}. \quad (1.32)$$

L'expression générale de la probabilité de choix d'une alternative  $k$ , donnée dans l'équation (1.14), peut être réorganisée de la manière suivante

$$P_{hk} = \Pr(\epsilon_{hk'} < \epsilon_{hk} + V_{hk} - V_{hk'} \quad \forall k' \neq k). \quad (1.33)$$

Pour un  $\epsilon_{hk}$  donné, en utilisant l'hypothèse iid et l'expression de la densité cumulative de Gumbel, donnée dans l'équation (1.31), on peut déterminer l'expression de la probabilité de choix de  $k$  par  $h$  conditionnellement à la valeur de  $\epsilon_{hk}$  :

$$P_{hk} \mid \epsilon_{hk} = \prod_{k' \neq k} \exp(-e^{-(\epsilon_{hk} + V_{hk} - V_{hk'})}). \quad (1.34)$$

---

18. Il convient de noter que l'étape de répartition modale est étroitement liée au développement des modèles de choix discrets qui ont valu à McFadden l'attribution du prix en mémoire d'Alfred Nobel en économie (voir McFadden, 2001).

La probabilité de choix non conditionnelle est l'espérance de la probabilité conditionnelle, équation (1.34), par rapport à  $\epsilon_{hk}$ . Elle est donnée dans l'équation (1.35)

$$\begin{aligned} P_{hk} &= \int_{-\infty}^{+\infty} (P_{hk} | \epsilon_{hk}) e^{-\epsilon_{hk}} \exp(-e^{-\epsilon_{hk}}) d\epsilon_{hk} \\ &= \int_{-\infty}^{+\infty} \left( \prod_{k' \neq k} \exp(-e^{-(\epsilon_{hk} + V_{hk} - V_{hk'})}) \right) e^{-\epsilon_{hk}} \exp(-e^{-\epsilon_{hk}}) d\epsilon_{hk}. \end{aligned} \quad (1.35)$$

On peut réécrire l'équation (1.35) en fonction de **toutes** les alternatives, incluant l'alternative  $k$ , de la manière suivante

$$\begin{aligned} P_{hk} &= \int_{-\infty}^{+\infty} \left( \prod_{k'} \exp(-e^{-(\epsilon_{hk} + V_{hk} - V_{hk'})}) \right) e^{-\epsilon_{hk}} d\epsilon_{hk} \\ &= \int_{-\infty}^{+\infty} \exp\left(-\sum_{k'} e^{-(\epsilon_{hk} + V_{hk} - V_{hk'})}\right) e^{-\epsilon_{hk}} d\epsilon_{hk} \\ &= \int_{-\infty}^{+\infty} \exp\left(-e^{-\epsilon_{hk}} \sum_{k'} e^{-(V_{hk} - V_{hk'})}\right) e^{-\epsilon_{hk}} d\epsilon_{hk}. \end{aligned} \quad (1.36)$$

En utilisant le changement de variable suivant

$$\delta = e^{-\epsilon_{hk}} \Rightarrow d\delta = -e^{-\epsilon_{hk}} d\epsilon_{hk},$$

la probabilité non conditionnelle de choisir  $k$  devient

$$P_{hk} = \int_0^{+\infty} \exp\left(-\delta \sum_{k'} e^{-(V_{hk} - V_{hk'})}\right) d\delta. \quad (1.37)$$

L'intégrale du terme de droite de l'équation (1.37) a une forme fermée donnée par

$$P_{hk} = \left[ -\frac{\exp\left(-\delta \sum_{k'} e^{-(V_{hk} - V_{hk'})}\right)}{\sum_{k'} e^{-(V_{hk} - V_{hk'})}} \right]_0^{+\infty} = \frac{e^{V_{hk}}}{\sum_{k'} e^{V_{hk'}}} \quad (1.38)$$

On retrouve dans l'équation (1.38) la forme traditionnelle de la probabilité de choix d'une alternative. En général, on fait l'hypothèse que l'utilité observée est linéaire. McFadden (1974) a montré que, sous cette hypothèse, la fonction de log-vraisemblance des probabilités de choix est concave en relation aux paramètres de la fonction d'utilité observée.

Une des propriétés caractéristiques du modèle logit multinomial est qu'il est en accord avec l'axiome de Luce (1959)<sup>19</sup>, encore connu sous le nom de propriété d'indépendance des alternatives non pertinentes (axiome IIA, de l'anglais *Independence of Irrelevant Alternatives*). Pour deux alternatives quelconques  $k$  et  $l$ , le ratio des probabilités est donné par

$$\frac{P_{hk}}{P_{hl}} = \frac{e^{V_{hk}} / \sum_{k'} e^{V_{hk'}}}{e^{V_{hl}} / \sum_{k'} e^{V_{hk'}}} = \frac{e^{V_{hk}}}{e^{V_{hl}}} = e^{V_{hk} - V_{hl}}. \quad (1.39)$$

19. Dans le cas de Luce, il s'agit d'un axiome à partir duquel il dérive le modèle logit.

Ce ratio ne dépend que des alternatives  $k$  et  $l$ , et, en particulier, de l'utilité observée de ces alternatives. Dit autrement, les probabilités relatives de choix de ces deux alternatives restent les mêmes quelles que soient les autres alternatives ou leurs attributs. Pour Luce, cette propriété peut être vue comme un équivalent probabiliste de l'hypothèse de transitivité des préférences. Elle découle directement de l'hypothèse d'indépendance des utilités non observées, c'est-à-dire de la spécification de l'utilité observée. Elle a deux avantages principaux :

- dans le cas où les alternatives sont trop nombreuses, comme dans les modèles de choix de mode, les paramètres de la fonction d'utilité peuvent être estimés de manière consistante à partir d'un échantillon des alternatives ;
- un autre avantage peut être trouvé si on n'est intéressé que par un sous-ensemble d'alternatives : si on considère que la propriété IIA s'applique, on peut estimer un modèle uniquement sur le sous-ensemble d'intérêt.

### 1.2.4 Affectation

La phase d'affectation fait appel à une représentation des réseaux issue de la théorie des graphes. Les infrastructures sont représentées par des arcs et des nœuds auxquels on associe certaines caractéristiques permettant de calculer les temps de parcours. Pour les réseaux viaires, on introduit des éléments permettant de caractériser la relation entre utilisation des infrastructures et temps de parcours. Pour les réseaux de transports en commun, on introduit des informations relatives aux itinéraires des lignes et aux horaires de passage aux arrêts. La congestion n'est généralement pas prise en compte pour les transports en commun.

Les équations (1.8) et (1.9) suggèrent que la détermination des charges sur les infrastructures et des niveaux de service doit se faire par des procédures itératives. Les procédures d'affectation retenues sont de type statique (elles prennent en compte des conditions de circulation moyennes sur une heure de pointe).

L'affectation sur le réseau routier se fait par un algorithme d'équilibre, respectant le premier principe de Wardrop. Pour chaque O-D, les temps de trajet<sup>20</sup> sur tous les itinéraires utilisés sont égaux et inférieurs à ceux qui seraient expérimentés par un véhicule sur un itinéraire pas encore utilisé. Chaque conducteur cherche à minimiser son coût de transport de manière non-coopérative. Les flux de trafic qui répondent à ce principe sont généralement qualifiés d'équilibre de l'utilisateur, puisque chaque utilisateur choisit unilatéralement son itinéraire optimal. Un équilibre est atteint lorsqu'aucun utilisateur ne peut abaisser son coût de transport par une action unilatérale.

L'affectation sur le réseau de transport en commun se fait par une approche de type multi-chemin, en fonction des horaires, en raison du maillage relativement fin sur l'agglomération lyonnaise.

## 1.3 Synthèse et remarques conclusives

Ce chapitre a permis de donner un aperçu de la diversité des approches possibles pour modéliser la demande de transport dans une perspective économique.

---

20. Il s'agit de temps généralisés. Voir chapitre 3.

Les comportements de déplacement peuvent être modélisés à une échelle agrégée, où les flux entre les origines et destinations sont représentés dans des matrices à l'échelle zonale qui, comme le remarque Gaudry (2007), sont proches, dans leur logique comptable et de circuit fermé, des tableaux de l'analyse *input-output* de Leontief (1941). La constitution de ces matrices O-D est souvent réalisée de manière séquentielle par des modèles dits à quatre étapes.

Les comportements de déplacement peuvent également être analysés à l'échelle des unités de prise de décision par l'intermédiaire des modèles de choix discrets. Kockelman (1998) propose un modèle intégré des décisions relatives à la mobilité des individus en matière de mobilité. Cependant, comme le souligne Train (2009), le fait de pouvoir théoriquement modéliser par une approche désagrégée, ne veut pas dire qu'il faut le faire. Ainsi, dans la pratique les modèles de choix discrets sont intégrés dans une ou plusieurs étapes des modèles à quatre étapes, sans remettre en cause la structure séquentielle de ceux-ci.

Pour choisir une stratégie de modélisation, il est nécessaire de les évaluer en fonction de leur opérationnalité. La stratégie à retenir est celle qui présente le meilleur compromis entre cohérence, pertinence et mesurabilité.

L'objectif du présent travail étant de tester la capacité prospective des modèles de prévision, il apparaît convenable de retenir des modèles proches de la pratique. Ainsi, nous testons des modèles à quatre étapes où la génération et la distribution sont de type agrégé, classique. Pour la répartition modale, nous retenons une approche désagrégée, ce qui permet d'étudier les propriétés de ces modèles dans une application correspondant au cadre dans lequel ils ont été développés.

---

# Transférabilité temporelle des modèles de demande de transport : état des savoirs

Ce chapitre présente un état des savoirs sur la robustesse de l'hypothèse de stabilité temporelle des modèles de prévision de la demande de transport.

Dans un premier temps, il établit un lexique du travail empirique de modélisation et il étudie la nécessité de valider un modèle de prévision à l'aide de données sur plusieurs dates. Il introduit, dans ce contexte, le concept de transférabilité temporelle.

Dans un deuxième temps, il présente une revue de la littérature relative à la validation temporelle des modèles de prévision. Cette deuxième section s'appuie en partie sur le travail de Fox et Hess (2010)

## 2.1 Validation des modèles prospectifs ou étude de leur transférabilité temporelle

Cette section présente les principaux concepts relatifs au travail empirique. Elle met en lien la procédure de validation d'un modèle prospectif avec l'analyse de sa transférabilité temporelle.

### 2.1.1 Vocabulaire de l'approche empirique

Le caractère résolument empirique de ce travail rend nécessaire la définition d'un certain nombre de termes associés aux différentes étapes présentes dans la mise en place d'outils de modélisation dans la planification des transports. Il semble intéressant de prendre appui sur un manuel récent portant sur la validation de ces modèles (Cambridge Systematics, Inc, 2010), réalisé dans le cadre du Programme d'Amélioration de la Modélisation des Déplacements (TMIP)<sup>1</sup> de l'Administration Fédérale des Routes (FWHA)<sup>2</sup> aux États-Unis. La proposition des définitions de ce manuel trouve son origine dans le constat qu'une des difficultés rencontrées lors

---

1. *Travel Model Improvement Program*

2. *Federal Highway Administration*

de la validation des modèles est l'absence de définitions claires des termes relatifs aux différentes procédures de calcul. Afin de combler cette carence, les définitions suivantes y sont proposées :

- Estimation/Calibration : utilisation des techniques d'analyse statistique et des données d'observation afin de déterminer les valeurs des paramètres d'un modèle. L'estimation concerne essentiellement des données désagrégées (sans biais ni facteurs de correction) mais elle peut, dans certains cas, porter sur des données plus agrégées.
- Assertion : déclaration *a priori* de la forme ou des paramètres d'un modèle sans avoir recours à l'analyse statistique des données d'observation.
- Validation : application des modèles estimés ou calibrés et comparaison avec les données observées. Idéalement, les données observées retenues pour cette comparaison ne sont pas celles qui ont servi à l'estimation ou la calibration (ce qui n'est, en pratique, pas toujours une alternative réalisable). La validation peut employer des données additionnelles recueillies en même temps que celles utilisées pour l'estimation ou la calibration. Elle peut aussi avoir recours aux estimations (ou calibrations) faites sur un autre jeu de données (correspondant, par exemple, à une autre année, pour un modèle statique, ou une autre ville). Elle devrait également inclure des tests de sensibilité.
- Tests de sensibilité : application des modèles individuellement et dans leur ensemble (dans le cas où on aurait, par exemple, une séquence de modélisation) en utilisant des hypothèses ou des données en entrée différentes de celles utilisées pour leur estimation ou calibration. Les tests de sensibilité des modèles individuels peuvent inclure des estimations des matrices variance-covariance des coefficients. Le test de sensibilité du modèle dans son ensemble peut se faire par son application en utilisant un jeu d'hypothèses différentes concernant les inputs démographiques, socioéconomiques ou de niveau de service des transports, ce qui permet de juger du caractère plausible et raisonnable des résultats.

Il convient de noter que ces définitions s'appliquent à un modèle préalablement spécifié. Bonnel (2004) définit la spécification (ou formalisation) comme « la mise en équation des relations entre variables ».

### 2.1.2 La démarche de prévision : validation temporelle et étude de la transférabilité

La séquence usuelle de modélisation des déplacements urbains repose sur trois hypothèses fondamentales dans son utilisation en prévision (Smith et Cleveland, 1976) :

- (i) les variables exogènes peuvent être prévues de manière précise ;
- (ii) les modèles fournissent une bonne simulation, juste du point de vue des comportements, des déplacements observés pendant la période de référence ;
- (iii) la structure des modèles en matière de variables explicatives ainsi que les paramètres sont stables au cours du temps.

On peut reformuler la procédure de prévision en utilisant le vocabulaire présenté dans le point précédent. On commence par spécifier un modèle comportant des variables

exogènes plus faciles à prédire que les variables endogènes, dans un souci de cohérence d'objectifs (hypothèse (i)). On estime/calibre les coefficients du modèle avec des données correspondant à une période de référence (données en coupe instantanée qui limitent également la structure du modèle qui doit être statique). Dans certains cas, où, par exemple, on n'aurait pas assez de données, les paramètres des modèles peuvent être assertés par l'analyste (en fonction de son expérience ou de sa connaissance du « terrain »). Une fois que le modèle a été estimé, il est souvent validé, de manière sommaire, en regardant des indicateurs de qualité des calages, c'est-à-dire que l'on évalue la capacité du modèle à reproduire les observations qui ont servi à l'estimer (hypothèse (ii)). Pour la prévision, on introduit des modifications sur les variables exogènes (populations, emplois, caractéristiques des réseaux en termes de capacités...), de telle sorte que celles-ci correspondent à ce que l'on croit savoir de l'état des choses à l'horizon de prévision (hypothèse (i)), et on applique le modèle avec les variables modifiées (hypothèse (iii))<sup>3</sup>.

On peut argumenter qu'une procédure de validation du modèle plus appropriée consisterait à comparer les prévisions du modèle avec les observations à l'horizon de prévision, en distinguant les écarts des variables endogènes prédites et observées selon que ceux-ci proviennent d'erreurs au niveau des prévisions des valeurs des variables exogènes (éventuellement modélisées de manière exogène) ou d'écarts dans les paramètres du modèle de prévision auquel on s'intéresse (il s'agit là d'un test de validité de l'hypothèse (iii)). Cette question de la validité temporelle des modèles de prévision de la demande est également connue sous le nom de « transférabilité temporelle ». La transférabilité d'un modèle peut être définie de la manière suivante (Koppelman et Wilmot, 1982) : le transfert est l'application d'un modèle, information ou théorie sur le comportement, développé pour un contexte particulier, dans l'objectif de décrire le comportement correspondant dans un autre contexte. La transférabilité est la valeur, l'utilité, l'utilisabilité<sup>4</sup> du modèle, information ou théorie transférée, dans le nouveau contexte. Hansen (1981), cité par Galbraith et Hensher (1982) et Karasmaa (2008), propose une hiérarchie à quatre niveaux pour l'analyse de la transférabilité :

- (i) la transférabilité de postulats généraux sur les comportements (comme, par exemple, la maximisation de l'utilité) ;
- (ii) la transférabilité d'une classe de modèles mathématiques ;
- (iii) la transférabilité de spécifications de modèles ;
- (iv) la transférabilité des coefficients d'un modèle.

La reformulation de la procédure de prévision que nous avons proposée dans le paragraphe précédent concerne plus particulièrement le quatrième type de transférabilité de cette catégorisation — et éventuellement le troisième, puisque l'estimation et la spécification sont liées (significativité des coefficients). Il en est de même pour la très grande majorité des études sur la transférabilité des modèles de prévision de la demande de transport.

Généralement, on distingue deux types de transférabilité : temporelle, qui correspond au cas dans lequel on s'intéresse à un territoire à des moments différents dans

---

3. On peut considérer que l'on fait une assertion sur la valeur des paramètres à l'horizon de prévision.

4. Anglicisme issu du mot *usability*.

le temps, et spatiale, qui correspond au cas où l'on s'intéresse à plusieurs territoires au même moment dans le temps. Karasmaa (2008) distingue deux types d'approches de la transférabilité. La première étudie l'hypothèse d'égalité des paramètres par des tests statistiques (biais de transfert). Karasmaa note que cette approche a l'inconvénient d'être trop restrictive dans la mesure où aucun modèle n'est parfaitement spécifié et qu'il y a toujours une certaine dépendance au contexte (l'hypothèse de stabilité serait quasi-systématiquement rejetée). La deuxième approche consiste à reconnaître dès le départ qu'il y a eu des changements et à étudier les moyens par lesquels les écarts et les biais qui en résultent peuvent être pris en compte sans avoir à ré-estimer un nouveau modèle en entier ni recueillir une grande quantité de données nouvelles. On peut noter que l'approche de la transférabilité faite par Karasmaa est moins celle d'une définition que celle de la recherche de méthodes pour limiter le biais de transfert, ce qui suppose une nouvelle procédure de validation pour ces méthodes. Cette recherche se limite à l'étude de la validité des modèles de prévision usuels et propose donc que, pour valider ces modèles, il est non seulement nécessaire de s'intéresser à la stabilité paramétrique, mais également à la qualité des prévisions réalisées en faisant en sorte que l'erreur sur les variables exogènes soit nulle.

On indique, dans ce qui suit, les principales contributions portant sur la transférabilité temporelle des modèles de prévision de la demande de transport.

## 2.2 Évidence sur la transférabilité temporelle des modèles de prévision de la demande de transport

Cette section fait état de la littérature portant sur la transférabilité temporelle des trois premières étapes des modèles à quatre étapes.

### 2.2.1 Génération

Les modèles de génération agrégés — c'est-à-dire les régressions multiples à l'échelle zonale — ont été étudiés principalement dans les années soixante et au début des années soixante-dix. À cette période, on commence à refaire des enquêtes déplacements dont certaines sont comparables, de par leur méthode, aux premières enquêtes réalisées pendant les années cinquante. La disponibilité de ces données rend ces études possibles.

Dès 1966, Dodd proposa une étude sur les modèles de génération (à la journée et à l'heure de pointe) à l'échelle zonale, à partir de données d'enquêtes ménages de 1956 et 1964 à Toronto. Pendant cette période, l'aire métropolitaine de Toronto a connu des changements importants en ce qui concerne les déterminants usuels de la quantité de déplacements. La population a augmenté de 30 % et la motorisation s'est accrue de 45 %. Malgré ces changements importants, les modèles étudiés ont montré une bonne performance prospective : les déplacements par personne sont restés relativement constants.

Ashford et Holloway (1972) analysent des modèles zonaux (régression linéaire et classification) de prévision des émissions pour cinq motifs de déplacement<sup>5</sup>. Les données dont ils se servent proviennent d'enquêtes ménages réalisées dans l'aire métropolitaine de Pittsburg en 1958 et 1967. Les coefficients des relations (régressions) ont connu des changements significatifs. Malgré cela, les deux types de modèles (régression et classification) ont permis de reproduire sans trop d'erreurs les déplacements pour le motif travail, mais pour les autres motifs étudiés des taux d'erreur relativement forts ont été trouvés. Les auteurs concluent en disant que, sur la période normale de prévision (20 ans) ces modèles ne peuvent servir qu'à donner des ordres de grandeur pour les relations étudiées.

Kannel et Heathington (1973) étudient des modèles de génération de type régression linéaire à l'échelle des ménages en les comparant aux modèles de régression linéaire à l'échelle zonale. Ils utilisent des données d'enquêtes sur l'État de l'Indiana pour 1964 et 1971. L'enquête de 1964 a un taux de sondage d'environ 5 %, avec plus de 10 000 ménages enquêtés. Parmi ces ménages, l'étude ne porte que sur ceux constitués par une seule famille, propriétaire de son lieu d'habitation et disposant de réponses complètes concernant les déplacements réalisés, la taille du ménage, le revenu et l'accès à la voiture. 4 300 ménages ont été ainsi retenus. Les données de 1971 correspondent à 357 ménages, parmi ceux ayant été retenus en 1964, enquêtés à nouveau, en prenant soin de garder le même jour d'enquête. Ceci a permis aux auteurs de tester la stabilité des relations de « production » de déplacements à l'échelle des ménages. Ils trouvent que les modèles estimés en 1971 connaissent une évolution plus importante du coefficient correspondant à la motorisation que celui correspondant à la taille du ménage. Cette variation n'est pas suffisamment forte pour induire une mauvaise prévision (rétrospective, en raison des données) des déplacements réalisés en 1964 avec les coefficients de 1971.

Downes et Gyenes (1976) utilisent des données de 1962 et 1971 sur la ville de Reading pour étudier la transférabilité temporelle de trois modèles de génération : un modèle de régression linéaire zonal, un modèle de régression linéaire au niveau des ménages et un modèle de classification. Pour les régressions zonales, les variables explicatives retenues sont la population et le nombre de véhicules possédés. Pour les régressions à l'échelle des ménages, les variables explicatives retenues sont la taille des ménages et des *dummy*-variables correspondant aux différents niveaux de motorisation. Ils montrent que les résultats des 3 modèles sont stables dans le temps, donnant des erreurs de prévision à l'échelle zonale entre 10 et 15 %, et suggèrent que la meilleure approche est la régression linéaire à l'échelle du ménage.

Yunker (1976) étudie la stabilité temporelle de la chaîne de modélisation des déplacements de la partie Sud-Est de l'État de Wisconsin, en utilisant des données d'enquêtes de 1963 et 1972. Il évalue les modèles de planification utilisés par la Southeastern Wisconsin Regional Planning Commission pour les motifs travail et achats (déplacements basés au domicile) : un modèle de génération à partir de régressions zonales ; un modèle de répartition modale pré-distribution mettant en lien la part des déplacements TC avec la motorisation et l'accessibilité relative aux transports en commun et aux autoroutes, et un modèle de distribution gravitaire avec une

---

5. Quatre motifs basés au domicile — travail, achats, enseignement et autres — et un motif regroupant les déplacements n'étant pas basés au domicile.

fonction de résistance exponentielle. Il trouve que ceux-ci prédisent raisonnablement les observations faites en 1972. Il faut noter ici que les prévisions sont de bonne qualité alors que le nombre de déplacements a augmenté de 25 %.

Smith et Cleveland (1976) s'intéressent aux phases de génération et de répartition modale (pré-distribution). Ils utilisent des données d'enquêtes auprès des ménages de Détroit, réalisées en 1953 et 1965, pour étudier des modèles de génération de type régression linéaire à l'échelle du ménage ainsi que des modèles de classification. Ces derniers ne semblent pas stables dans la mesure où les auteurs obtiennent des différences d'environ 20 % dans les valeurs des ratios d'émissions. Au niveau des régressions, les paramètres ne sont pas stables (test de Chow) à moins d'enlever les ménages ne se déplaçant pas de l'échantillon. Il faut noter que, même si les tests de stabilité au niveau paramétrique n'ont pas été satisfaisants, les équations calibrées pour 1953 produisent des prévisions acceptables des émissions observées en 1965. Mais, la perspective générale de cette étude sur la transférabilité est plutôt négative.

Doubleday (1977) se sert des mêmes données que Downes et Gyenes (1976) pour évaluer la transférabilité des modèles de classification appliqués à l'échelle des individus (et non des ménages). Il suggère que les ratios d'émission dépendent du type de catégorisation adopté et qu'ils ne sont, en général, pas stables au cours du temps, à l'exception de ceux relatifs aux hommes ayant un emploi. Les ratios d'émission des retraités et des femmes ne sont pas stables et semblent dépendre de l'accessibilité.

Kollo et Purvis (1984) étudient la stabilité de moyen-long terme de modèles de classification à deux variables : possession d'un véhicule et revenu du ménage. Ils utilisent des données d'enquêtes sur l'aire métropolitaine de San Francisco pour 1965 et 1981. Ils trouvent que des changements significatifs des taux d'émission pour une catégorie de la population sont contrebalancés par la modification de la structure de la population. Il en résulte que le taux d'émission moyen ne change pas significativement.

Walker et Peng (1991) s'intéressent à la question de la stabilité de long terme des relations de génération. À partir des données d'enquêtes de 1960 et 1987-88 sur la région de la Delaware Valley (aire métropolitaine de Philadelphia), ils évaluent la stabilité des ratios d'émission obtenus pour quelques méthodes de classification. Ils montrent que la sélection des variables a un impact sur la stabilité temporelle des ratios d'émissions et proposent un classement des méthodes de classification par niveau de stabilité. Ainsi, les approches par taille du ménage semblent être les moins stables, les modèles reposant sur le niveau de revenu ont un niveau de stabilité un peu meilleur et les classifications faisant appel au niveau de motorisation présentent la plus grande stabilité.

Badoe et Steuart (1997) remarquent que la plupart des études qui se sont intéressées à la stabilité temporelle des modèles de génération ont été réalisées sur des aires d'étude qui ont connu peu ou pas de changements (démographique, en termes d'infrastructure) ou elles ont été réalisées sur des périodes relativement courtes. De ce fait, les conclusions qu'elles pouvaient avancer en matière de stabilité des paramètres perdent de leur portée. Ils établissent alors une étude assez détaillée des changements connus par l'agglomération de Toronto (*Great Toronto Area*) entre 1964<sup>6</sup> et 1986<sup>7</sup>.

---

6. Données de l'enquête *Metropolitan Toronto and Regions Transport Study* : 21 823 ménages enquêtés en face-à-face, ce qui correspond à un taux de sondage d'environ 3,3 %. Il s'agit des mêmes données que celles utilisées par Dodd (1966).

7. *Transport for Tomorrow Survey*, enquête téléphonique auprès de 61 452 ménages, ce qui

Sur cette période, celle-ci a connu des changements importants parmi lesquels une augmentation de la population de 48 % (passant de 2,1 à 4,1 millions d'habitants) alors que le nombre de ménages est passé de 0,71 à 1,47 million, soit un accroissement de 106 %. Il y a également eu un doublement de la population résidant en périphérie. Le nombre de déplacements par personne est passé de 1,44 à 2,11, alors que le nombre de déplacements par ménage est passé de 5,5 à 5,85. Dans ce contexte, Badoe et Steuart, estiment des modèles de régression à l'échelle des ménages, pour les déplacements basés au domicile (pour l'ensemble des déplacements et pour les déplacements segmentés suivant les motifs travail, achats, loisir, et privés (*personal business*)). Les variables explicatives retenues sont le nombre de personnes du ménage, le nombre de personnes ayant un emploi, le nombre de permis et le nombre de véhicules. Ils étudient la capacité des modèles calibrés en 1964 à prévoir les déplacements de 1986 au niveau des ménages, des zones et des 6 communes de l'agglomération. Ils montrent que les modèles relatifs à l'ensemble des déplacements et au motif travail reproduisent assez bien les déplacements observés à l'échelle zonale, lors de l'estimation. En prévision, le modèle pour l'ensemble des déplacements réalise de bonnes prévisions (erreur autour de 10 %) alors que le modèle pour le motif travail surestime les déplacements d'environ 20 %. Les modèles pour les autres motifs ont des capacités descriptive et prédictive assez pauvres.

Cotrus *et al.* (2005) étudient les transférabilités temporelle et spatiale des modèles de génération par régression linéaire et Tobit au niveau de l'individu, sur les villes de Tel-Aviv et Haïfa, à partir de données d'enquêtes pour 1984 et 1996–97. Ils trouvent que les paramètres des deux types de modèles ne sont pas statistiquement stables au niveau temporel. Cependant, les prévisions faites avec les modèles conduisent à faire des erreurs de l'ordre de 7 % pour les modèles par régression et seulement 3 % pour les modèles Tobit, qui semblent mieux reproduire les comportements des individus qui ne se déplacent pas.

Badoe et Chen (2004a) s'intéressent à l'utilisation de données de plusieurs enquêtes dans l'estimation de modèles de régression à l'échelle des ménages, à des fins de prévision. Ils notent que les études sur les méthodes usuelles de transfert (voir la section 2.2.3) sont souvent réalisées avec seulement deux jeux de données. L'objectif est ainsi d'utiliser une partie des données du second jeu pour appliquer les méthodes de transfert à un modèle estimé seulement avec le premier jeu de données et de juger de la capacité du modèle transféré à reproduire l'ensemble des données du second jeu. Pour introduire de manière plus convenable l'analyse de la performance des modèles en prévision, Badoe et Chen utilisent 3 jeux de données d'enquête sur l'agglomération de Toronto, pour les années 1986, 1991 et 1996. Ils estiment plusieurs modèles correspondant à différentes manières de combiner (*pool*) les données de 1986 et 1991 pour prévoir les déplacements (tous motifs confondus) en 1996, dans le cadre d'estimations jointes. Ils trouvent que les modèles donnant les meilleurs résultats, tant en estimation qu'en prévision sont ceux prenant en compte l'écart des variances des erreurs en 1986 et en 1991. Ils évaluent par ailleurs la taille de l'échantillon nécessaire pour réaliser l'estimation conjointe, en notant qu'il est nécessaire d'avoir des échantillons de taille similaire pour cet exercice. Ils concluent qu'au-delà de 8 000 observations dans chaque échantillon le gain est très faible. Mwakalonge et Badoe

---

correspond à un taux de sondage d'environ 4 %.

(2012) complètent le travail de Badoe et Chen (2004a), en réalisant l'analyse sur une période plus longue (1986–2006). Ils s'intéressent à deux manières de combiner les données de plusieurs enquêtes : l'estimation jointe (utilisant les données combinées pour estimer un modèle) et la combinaison des paramètres des modèles estimés indépendamment avec les données d'une seule enquête. Ils observent que les modèles d'estimation jointe (avec les données de 1986, 1996 et 2001) où l'on admet des écarts de variance entre les différentes enquêtes permettent de réaliser les meilleures prévisions pour 2006.

Badoe et Mwakalonge (2011) comparent plusieurs méthodes pour traiter le problème des « cases vides » (segments pour lesquels on n'a pas d'observation) dans les modèles de classification croisée. Les données utilisées sont celles pour l'agglomération de Toronto (1986, 1996, 2001 et 2006). La comparaison est menée dans une perspective de capacité prédictive. Ils montrent que les méthodes classiques (*Multiple Classification Analysis* ou MCA, Stopher et McDonald, 1983), se fondant sur un recalcul de toutes les cases (y compris celles pour lesquelles on dispose d'observations) ont une performance prédictive nettement moins bonne que celles qui se limitent à traiter les cases vides. Parmi ces dernières, les plus performantes sont la méthode « ligne-colonne »<sup>8</sup> et la méthode de Mandel (1995) qui combine des moyennes des lignes et des colonnes.

## 2.2.2 Distribution et modèles gravitaires

Une revue rapide de la littérature consacrée au modèle de distribution gravitaire révèle une abondance d'études portant sur des considérations d'ordre théorique, comme les fondements théoriques du modèle (voir la section 1.2.2). Les contributions méthodologiques, comme les techniques de calibration (Deming et Stephan, 1940 ; Furness, 1965 ; Hyman, 1969 ; Evans, 1971 ; Batty et Mackie, 1972 ; Kirby, 1974 ; Beardwood et Kirby, 1975 ; Williams, 1976) ou la forme de la fonction de résistance (Tanner, 1961 ; Evans et Kirby, 1974), sont au moins aussi nombreuses. Ces recherches ont été conduites pendant les années 1970, le sujet étant moins d'actualité aujourd'hui. À l'opposé, la littérature empirique sur le modèle gravitaire est beaucoup plus rare, particulièrement en ce qui concerne la validité de celui-ci à plusieurs dates différentes. Openshaw est la référence incontournable en ce qui concerne les études empiriques en rapport avec le modèle gravitaire. En 1976, il proposa une première étude sur la capacité de neuf variantes du modèle gravitaire à reproduire les interactions spatiales observées. Il conclut que les modèles fondés sur la théorie, comme ceux ayant une origine empirique, ont du mal à atteindre une performance descriptive satisfaisante. Plus tard, en 1979, il s'intéressa à la performance du modèle gravitaire en prévision en étudiant la région de Durham, au nord de l'Angleterre, pour laquelle des temps-réseaux précis sont disponibles pour deux dates, 1966 et 1971. Si la performance descriptive du modèle est moyenne, celle en prévision est pratiquement la même que celle obtenue lors de la calibration, à condition que des estimations correctes des variables d'entrée soient disponibles. Les erreurs liées à l'hypothèse de stabilité des

---

8. Celle-ci estime deux modèles linéaires (un pour les lignes et l'autre pour les colonnes) à partir des cases non vides. Pour chaque case vide, le modèle donne deux estimations : une en ligne et une en colonne. Les deux estimateurs sont ensuite combinées en trouvant des pondérations qui minimisent la variance.

paramètres de calage sont relativement moins importantes que celles liées aux erreurs de prévision des variables exogènes. L'amélioration de la performance descriptive est possible, mais au détriment de la qualité des prévisions, en raison d'une sensibilité accrue aux changements de paramètres.

Southworth (1979) étudie l'impact de la structure spatiale (segmentation de la population par niveau de revenu et lieu de résidence) sur les paramètres du modèle gravitaire en calibrant plusieurs modèles contraints à la destination, en faisant une distinction par motif de déplacement, à partir de données de la London traffic survey de 1962. Il observe une grande variabilité des paramètres, et une amélioration de la qualité des calages apportée par la segmentation. Dans une deuxième étude (Southworth, 1983), il poursuit son travail sur les modèles segmentés en introduisant la composante temporelle, à partir de données du Grand Londres pour 1962 et 1971. Il cherche à comparer l'impact des changements dans les temps de parcours et ceux de désagréments non-temporelles sur les paramètres du modèle gravitaire. Ses résultats confirment ce qui avait déjà été avancé par Openshaw : la segmentation n'améliore pas la qualité des prévisions. Les prévisions sont plus sensibles aux écarts de temps de parcours qu'aux changements paramétriques. Compte tenu de ces résultats, tant Openshaw que Southworth recommandent de limiter l'usage du modèle gravitaire à des études descriptives et d'éviter son utilisation en prospective.

Volet et Hutchinson (1986) ont testé le pouvoir descriptif d'un modèle gravitaire comparé à celui d'un modèle de type « facteurs de croissance » (Fratat) sur les flux interzonaux domicile-travail de la ville de Toronto, pour 1971 et 1981. Ils ont également testé la capacité du modèle gravitaire calibré en 1971 à prévoir les flux observés en 1981. Leurs résultats suggèrent que le modèle très simple de type « facteurs de croissance » donne de meilleurs résultats en termes de prévision que le modèle gravitaire. Plus généralement, les deux types de modèles ont des difficultés à imiter les modifications dans la structure des trajets domicile-travail qui sont dues à des changements des modèles de vie et de travail des différentes catégories de la population.

Duffus *et al.* (1987) ont étudié la performance prospective d'un modèle gravitaire à fonction de résistance « par paliers »<sup>9</sup> et l'impact de l'inclusion de K-facteurs (correctifs spécifiques à certaines O-D) en utilisant des données d'enquête déplacements sur la ville de Winnipeg pour 1962, 1971, 1976 et 1981. Ils observent que l'allongement de l'horizon de prévision se traduit par une baisse de la qualité des prévisions réalisées par le modèle. L'inclusion des K-facteurs permet d'améliorer le pouvoir descriptif des modèles, mais se traduit également par une baisse de leur pouvoir prospectif. Ceci est sans doute lié à l'absence de fondements comportementaux permettant de déterminer les K-facteurs.

Plus récemment, Elmi *et al.* (1999) ont entrepris une étude sur la transférabilité temporelle du modèle gravitaire pour les déplacements liés au motif travail. Cette étude est très proche de celle de Volet et Hutchinson (1986) et la complète sur plusieurs aspects. L'aire d'étude est identique : Toronto, bien que Elmi *et al.* (1999) utilisent des données de trois enquêtes déplacements (la Metropolitan Toronto and Region

---

9. Dans la fonction de résistance par paliers, on spécifie des intervalles de temps ou de distance et on associe un niveau de « résistance » à chacun de ces intervalles. Cette fonction a donc autant de paramètres que l'on spécifie de segments de temps ou de distance (voir Evans et Kirby, 1974 ; Ortúzar et Willumsen, 2001).

Transport Study de 1964 et les Transportation Tomorrow Survey de 1986 et 1996) sur une période de 30 ans. Le modèle testé est le même, c'est-à-dire le modèle gravitaire doublement contraint avec une fonction de résistance exponentielle. Elmi *et al.* (1999) enrichissent leur étude de quatre segmentations du modèle : en fonction de la zone principale d'activité (relation centre-périphérie), du sexe, de la disponibilité d'un véhicule particulier et du type d'occupation des individus. Les résultats de l'étude montrent une variabilité du paramètre de la fonction de résistance dans le temps, qui suggère une évolution des comportements et une inclinaison à consacrer plus de temps aux déplacements. Globalement, les modèles segmentés ont une meilleure performance descriptive, notamment le modèle qui utilise la segmentation par type d'occupation. La performance prospective des modèles segmentés est également meilleure, ce qui semble aller à l'encontre des études que nous avons présentées jusqu'ici. Néanmoins, les auteurs tempèrent leurs propos en suggérant que, d'un point de vue pratique (outil d'aide à la décision), le gain apporté par la segmentation ne semble pas suffisant pour recommander son introduction dans les études de terrain. Ainsi, les auteurs concluent que le modèle gravitaire doublement contraint usuel apparaît comme étant satisfaisant pour l'usage en prévision.

Dans un cas un peu à part, Mikkonen et Luoma (1999) s'intéressent à l'évolution des paramètres d'un modèle d'interaction spatiale par paliers (à ne pas confondre avec une fonction de résistance par paliers) sur les villes de Vaasa, Seinäjoki, Kokkola et Pietarsaari en 1971, 1980-81 et 1991. Leur étude est un peu différente des autres études présentées ici car le modèle utilise la distance et non le temps de déplacement comme variable en entrée. Ils trouvent une évolution du paramètre de la fonction de résistance mais cette évolution est non monotone (on se serait attendu normalement à une réduction continue du paramètre). Les auteurs trouvent une explication à ce phénomène dans la localisation des gros centres commerciaux.

Les quelques études de caractère empirique sur la transférabilité du modèle gravitaire semblent converger sur l'idée que les affinements nécessaires pour reproduire une situation de calage de manière plus satisfaisante rendent hasardeuse l'utilisation du modèle en prospective. La recherche d'une alternative s'est révélée difficile. Les deux dernières décennies ont vu le développement de l'application des techniques du champ de l'intelligence computationnelle (logique floue, réseaux de neurones...) aux problèmes d'optimisation dans le domaine des transports. Ainsi, à partir du milieu des années 1990, on a assisté à l'apparition d'une série d'études sur l'utilisation d'une architecture de type « réseaux de neurones » pour l'étape de distribution (voir notamment Black, 1995 ; Openshaw, 1998). Elles suggèrent que cette approche serait plus performante que les modèles de type gravitaire. Cependant, Mozolin *et al.* (2000) avancent que la supposée meilleure performance des modèles en réseaux de neurones est uniquement descriptive et non prédictive. Ils montrent, en utilisant des données de recensement relatives aux déplacements domicile-travail de 1980 et 1990 sur l'aire métropolitaine d'Atlanta, qu'un modèle gravitaire simple, avec une fonction de résistance exponentielle, prédit mieux une matrice O-D future qu'un modèle de type réseaux de neurones. On retrouve donc un cadre de figure similaire à celui dépeint par Openshaw et Southworth : les nouvelles approches sont plus performantes en matière descriptive — elles permettent de gérer significativement mieux les situations où l'information est rare — mais elles sont moins performantes que l'approche classique en prévision.

### 2.2.3 Partage ou choix modal

Les études de transférabilité des modèles de répartition modale, et plus généralement, des modèles de choix discret appliqués aux transports — dont les modèles intégrés de choix de la destination et choix du mode méritent une mention spéciale dans le contexte de ce travail —, sont indiscutablement les plus nombreuses, et les plus élaborées en matière de tests statistiques permettant d'évaluer la stabilité des paramètres. Bien que ces études se soient penchées surtout sur l'aspect spatial — la possibilité d'adapter un modèle à un autre contexte spatial à moindre coût étant une incitation d'une certaine force — l'aspect temporel est loin d'avoir été négligé. Il convient de noter ici que le développement foisonnant des modèles désagrégés de choix discret dans la modélisation des transports est dû, en partie, à la croyance, fondée ou pas, qu'une meilleure connaissance des comportements à l'échelle individuelle se traduirait par une meilleure performance prospective des efforts de modélisation, essentiellement parce que cette meilleure connaissance des comportements se traduirait par une stabilité des coefficients associés à la modélisation de ces comportements. Plus précisément, comme le signalent Badoe et Miller (1995a) : une fois qu'un modèle est bien spécifié, qu'il capture convenablement le processus de prise de décision, dans un contexte, on devrait pouvoir l'appliquer à d'autres contextes aussi longtemps que le processus de prise de décision demeure inchangé, autrement-dit que les goûts et préférences des agents demeurent inchangés. Ce qui suit emprunte à Karasmaa (2008) et Fox et Hess (2010).

Ben-Akiva et Atherton (1977) s'intéressent au développement d'une méthodologie pour prévoir des changements de comportements sur les déplacements courts, notamment ceux qui résultent d'incitations au covoiturage. L'aspect temporel de la modélisation est testé à partir de données relatives à la « Shirley Highway » de Santa Monica, sur la période 1970–1974, où la mise en place d'une voie séparée pour les voitures occupées par 3 personnes ou plus et d'un service de bus a été menée. L'étude a conclu à la transférabilité temporelle des modèles logit incrémentaux.

Parody (1977) étudie le pouvoir prédictif à court terme (environ 2 ans) de plusieurs spécifications (modèles allant de 3 à 7 variables explicatives) d'un modèle logit à deux modalités : bus et voiture particulière. Il prend appui sur des données d'une expérience menée par la *University of Massachusetts* à Amherst sur la mise en place d'un service de bus gratuit associée à une augmentation du coût du stationnement sur le campus universitaire. L'expérience a eu lieu entre 1972 et 1974 et elle prend donc aussi en compte l'augmentation des prix des carburants à la suite du choc pétrolier de 1973. Les enseignants et les étudiants participent à quatre séries d'enquêtes téléphoniques entre 1972 et 1974. Les mêmes personnes sont interviewées, et les modèles sont calibrés en enlevant de l'échantillon les captifs d'un mode. L'analyse du pouvoir prédictif de chacune des spécifications testées indique que l'ensemble des modèles, et notamment ceux incluant des données sur le niveau de service ainsi que socioéconomiques (à la différence des modèles incluant uniquement des variables de niveau de service), permettent de faire une bonne prévision des parts modales.

Talvitie et Kirshner (1978) étudient la transférabilité temporelle des paramètres d'un modèle de répartition modale en utilisant des données sur la ville de San Francisco en 1973 et 1975, c'est-à-dire avant et après la mise en place du *Bay Area Rapid Transit* (BART), qui est un système de trains rapides desservant San Francisco

et les villes proches. En termes de spécification des modèles, ils constatent que l'utilisation séparée du temps de marche et du temps d'attente est préférable à l'utilisation de leur somme, ou « temps hors véhicule ». En termes de transférabilité, l'hypothèse de stabilité des paramètres, testée à partir d'une hypothèse d'égalité de l'ensemble des paramètres (constantes incluses), ne trouve pas de corroboration. Le modèle n'est pas transférable dans le temps.

Ces résultats ont été, en quelque sorte, corroborés par Train (1978), qui, en utilisant le même jeu de données que Talvitie et Kirshner, trouve que les prévisions faites par le modèle calibré en 1973 surestime l'attractivité du BART, même si les paramètres de niveau de service sont relativement stables. L'écart est lié à la difficulté d'avoir des valeurs justes — réseaux modélisés correspondant à la réalité — pour le temps de marche à pied. Une étude complémentaire utilisant les mêmes données (Train, 1979) permet de mettre en avant une relative meilleure performance en prévision des modèles plus élaborés (et décrivant mieux la situation de référence) que des modèles simples. Ce résultat est cependant tempéré par le fait que tous les types de modèles testés laissent à désirer en matière prévisionnelle.

McCarthy (1982) reprend le même jeu de données, mais les retravaille en enlevant la population captive d'un seul mode et les individus dont les réponses changent trop d'une enquête à l'autre. D'autre part, un troisième jeu de données, un « vrai » pré-BART, a été développé pour palier au fait que des délais opérationnels ont fait que l'enquête de 1973 a eu lieu alors que certains itinéraires du BART étaient déjà disponibles. Les résultats obtenus suggèrent que les modèles estimés produisent de bons résultats en matière de prévision. Les tests statistiques sur l'hypothèse de stabilité des paramètres ne permettent pas d'écarter l'hypothèse de stabilité. La transférabilité temporelle à court terme est donc vérifiée.

Silman (1981) analyse la stabilité temporelle d'un modèle de choix modal, pour les déplacements pour le motif travail, désagrégé, portant sur 3 modes (VP conducteur, VP passager et Bus). Il utilise des données d'enquêtes réalisées sur l'aire métropolitaine de Tel-Aviv, en 1972 (recueillies avant le choc pétrolier) et 1976. Dans un premier temps, il effectue des essais sur les deux jeux de données à la recherche de la spécification la plus adaptée à la prévision. Le modèle retenu est caractérisé par un terme générique de coût (traduisant le fait que la désutilité liée à une dépense devrait être la même, indépendamment du type de dépense). Les modèles calibrés à partir des données de 1972 permettent de prédire de manière satisfaisante les observations de 1976, malgré l'existence de changements structurels profonds entre les deux dates. Les coefficients du modèle sont stables au cours du temps à l'exception du coefficient pour le nombre de véhicules. Il faut noter que cet écart est sans doute dû au fait que l'information sur l'accès à la voiture était un peu différente dans les deux jeux de données dans la mesure où le deuxième ne comportait pas d'information sur la possession du permis de conduire. Une remarque importante faite par Silman est que la meilleure performance des modèles incluant des variables socioéconomiques peut être due moins à l'importance en soi de ces variables dans le processus de choix mais au fait que leur inclusion permet d'accroître, indirectement, le pouvoir explicatif des variables de niveau de service.

Badoe et Miller (1995a) étudient la transférabilité temporelle de long-terme des modèles logit de choix modal, de manière relativement complète. Ils utilisent des données des enquêtes ménages déplacements réalisées sur l'aire métropolitaine de

Toronto (*Great Toronto Area*) en 1964 (*Metropolitan Toronto and Regions Transport Study*) et 1986 (*Transport for Tomorrow Survey*). Les deux enquêtes ne contiennent pas des données identiques : l'enquête de 1964 posait des questions relatives à la catégorie socioprofessionnelle des ménages et à leur revenu alors que celle de 1986 ne dispose pas de ces informations. Certains compléments ont été tirés des données de recensement. Les données relatives au niveau de service proviennent de réseaux modélisés afin de reproduire la situation observée aux années d'enquête. Les modèles estimés étudient trois alternatives en termes de modes : voiture particulière en tant que conducteur, transports en commun (à l'exception du rail) et marche à pied. Sept spécifications ont été testées, allant de la plus simple, un modèle de part modale, au modèle le plus complexe, incluant des variables de niveau de service et des caractéristiques socioéconomiques individuelles, avec une segmentation heuristique par types d'individus aux comportements similaires (10 segments de marché). En ce qui concerne la transférabilité temporelle des modèles, les tests statistiques conduisent à rejeter l'hypothèse d'égalité des paramètres dans le temps, ce qui indique que les modèles ne sont pas stables dans le temps. D'un point de vue théorique, la transférabilité temporelle est donc rejetée. Néanmoins, d'un point de vue pratique, le modèle « transféré » comporte des informations utiles dans le contexte d'application. Ainsi, par exemple, les valeurs de l'indice de transférabilité (indicateur similaire à un ratio de vraisemblance) pour certains modèles étaient de 0,9. Dit autrement, le modèle de 1964 a 90 % de la capacité prédictive du modèle de 1986. La spécification pour laquelle le choix modal est conditionné par le fait de posséder une voiture particulière, est celle qui a donné les meilleurs résultats en termes de transférabilité. Les modèles simples faisant appel seulement aux niveaux de service apparaissent comme étant suffisamment robustes dans les conditions normales d'application des modèles en planification. Les spécifications les plus complexes, fortement segmentées, ont donné de mauvais résultats en prévision. Pour ces derniers, les auteurs suggèrent qu'ils peuvent être sur-spécifiés par rapport au jeu de données de calibration. Ce dernier résultat fait écho à certains travaux vus précédemment : il semblerait qu'il y a un niveau optimal de complexité dans la spécification des modèles en vue de leur utilisation en prévision.

Le travail de Badoe et Miller comporte un complément intéressant en matière d'amélioration des pratiques de modélisation. Il s'agit d'une étude (Badoe et Miller, 1995b) sur la performance de quatre méthodes de mise à jour des modèles désagrégés de choix modal, à savoir : (i) la mise à jour bayésienne, (ii) normalisation des paramètres (*transfer scaling*), (iii) estimation de transfert combiné (*combined transfer estimation*) et (iv) estimation de contexte jointe (*joint context estimation*). Pour les quatre procédures, l'hypothèse est faite qu'un « petit » jeu de données est disponible pour faire l'actualisation. Les résultats de l'étude suggèrent que les trois dernières procédures apportent des gains en prévision. Cependant, si le « petit » échantillon d'actualisation a plus de 400 ou 500 observations, l'estimation d'un modèle à partir de ces données resta la meilleure stratégie.

Karasmaa et Pursula (1997) étudient la transférabilité temporelle d'un modèle combiné de choix de destination et choix de mode pour les déplacements domicile-travail, sur l'aire métropolitaine de Helsinki. L'étude est faite à partir de données d'enquêtes pour 1981 et 1988 et de données de niveau de service des transports provenant de réseaux modélisés à ces dates. L'objectif principal était de tester les

différentes méthodes de mise à jour des modèles disponibles (à savoir, les quatre méthodes évoquées ci-dessus) en utilisant des tailles différentes de sous-échantillons des données de 1988. Leurs résultats corroborent ceux de Badoe et Miller (1995b) en ce qui concerne l'intérêt des méthodes d'actualisation, bien qu'ils révèlent aussi une grande variance des coefficients. Les résultats produits en termes de transférabilité des coefficients de 1981 sont plutôt contraires à l'hypothèse de stabilité des paramètres ; Karasmaa et Pursula trouvent des différences significatives.

## 2.3 Conclusion

Il est important de noter qu'il n'existe pas de consensus sur la transférabilité temporelle des modèles. Pour toutes les catégories de modèles, certaines études trouvent une stabilité des paramètres et d'autres trouvent une absence de stabilité. Il en va de même pour ce qui est des performances prospectives.

Un élément qu'il semble important de dégager est le risque de surdétermination des modèles par rapport aux données de calibration. Il semblerait que, sous certaines conditions, il est possible de trouver un équilibre entre un bon niveau de spécification et une bonne performance prospective.

Au niveau de la génération, on observe que les modèles par classification sont plus utilisés actuellement, mais la comparaison en termes de performance prospective de ces modèles par rapport à celle des modèles par régression linéaire à l'échelle du ménage ou de l'individu ne semble pas concluante.

Au niveau de la distribution, le modèle gravitaire modérément désagrégé, notamment par rapport aux types d'emplois, semble l'approche la plus performante en matière de prévision. Les études ayant surtout porté sur le motif travail, le choix des variables à partir desquelles une segmentation permettrait d'améliorer la qualité des calages est ouvert.

Au niveau de la répartition modale, les approches désagrégées semblent se dégager comme étant les plus pertinentes grâce au fait qu'elles permettent d'inclure plus de variables explicatives. L'inclusion de certaines variables socioéconomiques semble améliorer considérablement le pouvoir prédictif des modèles, de manière un peu indirecte, en augmentant la significativité des paramètres relatifs au niveau de service des réseaux de transport.

Finalement, pour l'ensemble de la chaîne, il est capital de souligner l'importance du rôle joué par les traitements d'homogénéisation des données pour assurer leur comparabilité ainsi que celle des modèles qu'elles alimentent.

---

## Terrain d'étude et données pour la modélisation

*Presque à mi-chemin entre Londres et Florence, Lyon marie avec harmonie un culte intense du négoce et de la haute tradition intellectuelle, un maintien rigide et une bonhomie de bon aloi, des fleuves puissants et de douces collines, et surtout, selon les saisons et les jours, passe sans transition des lourdes brumes industrielles des Midlands au ciel immense, profond et ocre de Toscane.[...] Accrochée aux rives du Rhône et de la Saône, cette ville est née des transports et sa vocation fut avant tout d'être un carrefour.*

— Jean Arrivetz, *Histoire des Transports à Lyon, Avant-Propos*

La contrainte en matière de disponibilité des données est une des raisons — peut-être même la raison principale — qui font que, malgré son ancienneté et les critiques qui lui sont adressées<sup>1</sup>, le modèle à quatre étapes demeure incontournable dans la pratique de la planification des transports en urbain. Cela ne veut pas dire que les besoins en matière de données pour son utilisation ne sont pas considérables. Ce chapitre fait état des données relatives à l'agglomération lyonnaise, utilisées dans la présente étude. Il permet de mettre en évidence certains éléments méthodologiques à considérer dans leur utilisation — tout comme dans leur production — dans le contexte d'une étude diachronique et les choix qui ont été réalisés à cet égard dans ce travail. Il structure leur présentation en fonction de la place qu'elles occupent lors de la mise en place d'une modélisation de la demande transports telle que présentée dans la section 1.1. Une première section présente l'intérêt de l'agglomération lyonnaise comme terrain d'étude, en mettant en évidence certaines de ses spécificités. Une deuxième section présente les données relatives à la demande de transport, en insistant sur leur source et les retraitements nécessaires afin d'assurer leur comparabilité dans le temps. Une dernière section présente les données relatives à l'offre de transport. Elle décrit les principes adoptés dans ce qui a été l'une des tâches les plus difficiles pour l'accomplissement de la présente étude : la codification de l'historique des réseaux de transport.

---

1. Sur les critiques adressées au modèle à quatre étapes voir McNally (2000b), McNally (2000a) ou Boyce (2002), entre autres.

### 3.1 L'agglomération lyonnaise : reflet des mutations dans les politiques d'aménagement des transports en France

Montès (2003) met en évidence certaines caractéristiques de l'agglomération lyonnaise qui la rendent un terrain d'étude intéressant dans le domaine de l'aménagement urbain et, en particulier, de la planification des transports. Il s'agit de la deuxième agglomération française, après Paris, et sa population dépasse le million d'habitants<sup>2</sup>. En termes d'infrastructures de transport, (pp. 10–11) :

« La communauté urbaine de Lyon comprend environ 2 800 km de voirie dont une soixantaine d'autoroutes, sur lesquelles circulent plus de 400 000 voitures particulières [par jour] auxquelles il faut ajouter le transit.[...] Le réseau de transports en commun est de bonne qualité et comprend, au premier septembre 2000, 98 lignes de bus et trolley (1 173 km de lignes cumulées), 4 de métro (27 km), et 2 de funiculaire (1,2 km), qui assuraient 163 millions de déplacements [par an] en 2001 [...]. Il faut y ajouter[...] deux lignes de *tramway*<sup>3</sup> ouvertes en janvier 2001 (19 km). »

Bieber (1985) remarque que « Lyon a connu l'ensemble des transformations économiques et sociales qui ont marqué la société française, depuis la guerre, sans que des spécificités trop fortes se fassent sentir ». Cette représentativité de l'agglomération lyonnaise peut également être observée en ce qui concerne les mutations dans les politiques d'aménagement urbain en France (Montès, 2003, p. 5) :

« De fait, n'y trouve-t-on pas à la fois un “maire bétonneur”, Louis Pradel, chantre d'une modernisation fondée sur le primat de la route (à l'origine de la plus célèbre autoroute intra-urbaine française), primat tempéré par le début du réseau de métro lié à la desserte du centre commerçant ; la création d'une intercommunalité, forcée au départ, volontaire ensuite ; la réflexion sur la nécessité d'un réel réseau interconnecté de transports qui tire parti de tous les modes de transport. »

Montès note que les limites chronologiques des politiques d'aménagement sur l'agglomération lyonnaise, depuis la fin des années 1950 sont bien identifiables. Une première période, allant de 1958 à 1973, où « la fin de la reconstruction et la prospérité économique sont à l'origine de l'apparition d'une idéologie optimiste et volontariste » donnant lieu à un aménagement des transports impulsé par une planification centralisée au niveau de l'État, avec une visée globale, alors que les agglomérations sont « peu organisées sur le plan institutionnel ». Les transports sont organisés en fonction de l'industrie, et de manière à « préserver la liberté qu'exprime la mobilité »<sup>4</sup>. Une deuxième période, allant de 1973 à 1984, où l'État se désengage progressivement du

---

2. Ce seuil semble avoir une importance dans la définition des politiques d'aménagement des transports, notamment en ce qui concerne les investissements massifs en systèmes de transports collectifs en site propre (spécialement le métro, voir Bieber, 1985).

3. En italique dans le texte d'origine.

4. Les aménagements transports suivent une logique d'augmentation de la capacité, en quelque sorte.

financement des projets d'aménagements locaux et où l'on entame vraiment le processus de décentralisation. Il s'agit d'une période de transition et d'attente. Finalement, la troisième période, allant de 1984 au début des années 2000 (époque à laquelle paraît l'ouvrage de Montès), caractérisée par un affermissement du local. La création du SYTRAL (**SY**ndicat Mixte des **TR**ansports de Lyon et de l'**Ag**glomération Lyonnaise) en 1985 est un exemple du renforcement du local<sup>5</sup>.

On peut rajouter, aux trois phases décrites par Montès, une dernière période d'aménagement dont le début se situe vers l'année 2000 : celle du recul de l'automobile, à la suite des aménagements réalisés en 2001, visant à réintroduire et développer les transports en commun de surface sur site propre (trams et couloirs de bus). Pour la première fois, on observe une baisse de la part modale de l'automobile (constat, voire résultat phare, de l'enquête ménages déplacements de 2006) avec 15 % de déplacements en moins qu'en 1995 sur le Grand Lyon (SYTRAL, 2007a,b).

Il convient de noter que notre étude, compte tenu des contraintes en matière de données, ne correspond qu'aux comportements de mobilité sur la troisième et la quatrième phase. En revanche, les infrastructures de transport sont en partie le résultat de décisions prises dans la première et la deuxième phase.

Un dernier élément à évoquer est la topographie. Lyon a un relief dissymétrique (Montès, 2003) : l'Ouest se caractérise par des « hauteurs d'accès abrupt (les Monts du Lyonnais et les Monts d'Or) ; le Nord, par « un plateau terminé par la costière de la Dombes » ; et l'Est est constitué par une grande plaine « formée des terrasses du Rhône, en partie inondable du fait du caractère sauvage de ce fleuve ; il en est de même pour la Saône, qui se jette dans le Rhône à Lyon ». Le relief est un élément important puisqu'il permet d'expliquer, en conjonction avec les préoccupations d'aménagement du territoire de chaque époque, la forme actuelle des infrastructures de transport lyonnaises. Bieber (1985) remarque que le « site très particulier de la ville ancienne a induit une anisotropie marquée », le cœur historique de la ville s'étant étendu à partir du site médiéval aux pieds de Fourvière, dans un premier temps, à la Presqu'île, située entre la Saône et le Rhône puis, dans un second temps, aux plaines de l'Est, constituant des zones naturelles de développement. Les collines du Nord et de l'Ouest ont constitué des barrières efficaces limitant le développement.

---

5. Sur la page de présentation du SYTRAL, dans son site internet (<http://www.sytral.fr/2-sytral-presentation.htm>), on peut lire : « Le SYTRAL est un syndicat mixte dont l'instance décisionnaire est composé d'élus issus du Grand Lyon et du département du Rhône. Les décisions sont prises au sein d'un Comité syndical composé de 26 élus. Ces élus définissent la politique des transports urbains dans l'agglomération et contrôlent sa mise en œuvre. Le SYTRAL exerce de nombreuses missions dont l'objectif majeur est l'amélioration des déplacements et de la qualité de vie dans la communauté urbaine de Lyon. Il délègue l'exploitation du réseau TCL dans le cadre d'une délégation de service public. Le territoire de compétences du SYTRAL correspond au Périmètre des Transports Urbains (PTU) de l'agglomération. En collaboration avec les collectivités locales, le SYTRAL élabore le Plan de Déplacements Urbains (PDU). Un des outils principaux, nécessaire à l'élaboration du PDU est l'Enquête Ménages Déplacements (EMD) réalisée régulièrement auprès des habitants du Grand Lyon et des communes du PTU.

« La politique d'aménagement des transports en commun du SYTRAL, s'exprime à travers le Plan de Mandat qui établit les investissements de la collectivité pour les projets qui seront réalisés sur une période définie (actuellement 2008-2014). »

## 3.2 Les données relatives à la demande de transport

Les données relatives à la demande de transport sont celles qui permettent d'établir les comportements de mobilité observés, de référence. Il s'agit donc, dans l'approche agrégée, des données permettant d'établir les matrices O-D par mode<sup>6</sup>. Dans l'approche désagrégée, il s'agit de données caractérisant des unités décisionnelles cohérentes (ménages ou individus) ainsi que les décisions prises en matière de mobilité (nombre de déplacements pour un certain motif, destination, mode, etc.).

Les données de type enquête ménages déplacements (CERTU, 2008) sont la source privilégiée pour les deux types d'approches<sup>7</sup>.

### 3.2.1 Les enquêtes ménages déplacements

Nous avons vu dans le chapitre précédent que, pour caractériser les émissions, on a besoin de faire le lien entre une quantité de déplacements (ou sorties) — étant la variable à expliquer — et les caractéristiques socioéconomiques des individus ou les caractéristiques de zones, soit la répartition spatiale des activités. Les sources de données permettant d'établir au mieux ce lien sont les enquêtes ménages déplacements. Pour cette étude, nous avons utilisé les trois dernières enquêtes ménages déplacements (EMD) réalisées sur l'agglomération lyonnaise (1985, 1995 et 2006), que nous présentons à continuation.

#### 3.2.1.1 Présentation générale des enquêtes ménages déplacements

Les EMD sont l'un des instruments essentiels pour l'étude de la mobilité et de son évolution en France. Nous reprenons ici les éléments décrits dans CERTU (2008). La méthode « standard Certu » de réalisation des EMD repose sur les principes suivants :

- les EMD visent à reconstituer la mobilité des habitants pour un jour ouvrable « moyen » de semaine, hors vacances scolaires ;
- elles sont réalisées au domicile des personnes, par des enquêteurs spécialement formés à ce type de recueil. Toutes les personnes de cinq ans et plus habitant dans le logement sont interrogées en personne ;
- tous les déplacements réalisés la veille du jour d'enquête par chaque personne enquêtée sont recensés. Les caractéristiques de chaque déplacement — motif, mode, origine, destination, heure de départ et d'arrivée — sont recueillies ;
- les EMD portent sur un échantillon représentatif des ménages de l'aire enquêtée. Le plan de sondage est stratifié par secteur de résidence. Dans chaque secteur,

---

6. Il convient de noter que si l'on dispose des matrices O-D observées par mode (ce que la répartition modale cherche à reproduire), alors on dispose automatiquement des matrices O-D tous modes (distribution) et des marges de celle-ci (émissions et attractions de la génération).

7. La modélisation agrégée peut se passer des données d'enquête dans la mesure où il est possible de déterminer des matrices O-D par d'autres méthodes comme celles faisant appel aux données de comptage. En revanche, quelle que soit la méthode retenue pour la détermination de la matrice O-D observée, la quantité de données requise est très importante (pour avoir une fiabilité suffisante sur chaque case de la matrice). L'approche désagrégée fait exclusivement appel à des données d'enquête, même si le développement des nouvelles technologies de communication semblent ouvrir d'autres possibilités.

les ménages sont tirés aléatoirement à partir d'un fichier de logements. Le logement désigné doit constituer la résidence principale du ménage. La taille de l'échantillon est déterminée de manière à assurer une fiabilité minimum des résultats permettant une analyse sectorielle. Le taux d'échantillonnage global est d'environ 1 %.

#### 3.2.1.2 Comparabilité dans le temps des EMD lyonnaises

Une analyse du type de celle entreprise dans la présente étude est inévitablement confrontée à la possibilité de biais en lien avec la qualité des données utilisées. En principe, le respect de la méthode d'enquête « standard Certu » pour les EMD de 1985, 1995 et 2006<sup>8</sup> assure une comparabilité des données dans le temps. Dans la pratique, il y a toujours un certain nombre d'éléments, en lien avec la méthode de collecte de données, qui peut produire des perturbations et, en conséquence, avoir un impact sur leur comparabilité temporelle. Il convient donc d'analyser à quelles conditions, la comparabilité des EMD lyonnaises peut être acceptée comme hypothèse de travail.

**3.2.1.2.1 Perturbations difficilement mesurables** Certaines perturbations affectent l'ensemble des travaux employant des données d'enquête tout en étant difficilement mesurables et donc peu traitables. Richardson (2003) distingue trois sources potentielles de biais systématiques dans les données provenant d'enquêtes : la réponse erronée (ou inconsistante, de l'anglais *inaccurate reporting*), le codage incomplet (non-réponse partielle ou, dit autrement, les personnes ne terminant pas le questionnaire ou ne souhaitant pas répondre à toutes les questions) et la non-réponse totale (les individus ne voulant ou ne pouvant pas répondre aux enquêtes). Nous discutons ici principalement la première et la troisième. La deuxième a pour effet de conditionner les variables socioéconomiques utilisables, notamment le revenu.

La réponse erronée concerne les cas où l'analyste remarque que certaines des réponses sont probablement erronées ou incomplètes. Dans le cas des EMD, on peut penser que la méthode en face à face peut réduire l'impact de cette source d'erreur dans la mesure où les questions posées ne sont pas de nature très sensible pour les enquêtés (outre peut-être le cas du niveau de revenu et de certains « déplacements privés » que les enquêtés souhaiteraient garder secrets vis-à-vis de l'enquêteur ou des autres membres du ménage) et que la présence de l'enquêteur permet d'expliquer le sens de certaines questions aux enquêtés. L'application de la méthode, dans des contextes différents (trois périodes différentes) peut cependant conduire à produire des différences dans ce qui est retenu comme étant une réponse erronée pour chaque enquête. La qualité des enquêteurs, de leur recrutement, de leur formation et de leur suivi peut par exemple changer en fonction du prestataire chargé de réaliser l'enquête. Ceci peut, entre autres, conduire à des changements dans ce qui est effectivement codé comme étant un déplacement (sur ce point voir également la section 4.1.1.1). Pour les EMD de l'agglomération lyonnaise, la partie « conception » de l'enquête est

---

8. La méthode « standard Certu » a également été utilisée pour l'enquête de 1976.

restée à peu près inchangée, alors que la réalisation de l'enquête a été assurée par différents prestataires (publics et privés)<sup>9</sup>.

La non-réponse totale peut conditionner la représentativité des enquêtes (Bonnell et Le Nir, 1998 ; Richardson, 2003). Celles-ci sont censées être réalisées sur un échantillon représentatif de la population. Idéalement, cette représentativité de la population doit permettre d'obtenir une représentativité des comportements ciblés par l'enquête, traduits par les variables d'intérêt retenues (réponses aux questions posées). En effet, les caractéristiques socioéconomiques des personnes qui acceptent d'être enquêtées ne sont pas toujours représentatives de celles de l'ensemble de la population. Dillman et Christian (2005) montrent, par exemple, que les non répondants sont moins éduqués, plus jeunes et appartenant davantage au sexe masculin que les répondants. De surcroît, il n'est pas toujours possible de postuler une indépendance entre la non-réponse et les comportements de mobilité (Bonnell, 2004). Ampt (1997), Sammer (1997) ainsi que Bonnell (2003) ont mis en évidence une relation décroissante entre le nombre de déplacements quotidiens et le nombre de relances nécessaires pour obtenir un retour, dans le cas d'une enquête postale. Dans les enquêtes téléphoniques, on observe plutôt une relation inverse (croissante, Bonnell, 2003). En analysant les causes de la non-réponse, Ampt (1997) montre que la non réponse peut être liée à la mobilité de la personne enquêtée. En particulier, deux cas de figure sont envisagés : dans le premier, le répondant potentiel considère que sa mobilité, faible, n'est pas intéressante pour les organismes conduisant l'enquête. Dans le second, il peut craindre la longueur de l'enquête, incompatible avec son emploi du temps, de par son « hyper-mobilité ». La sous-représentation des populations peu mobiles et très mobiles est également mise en évidence par van Evert et Moritz (2000) et Richardson (2003). Dans la même veine, Kalfs et van Evert (2003) montrent que les personnes ayant des schémas d'activités très simples ou très complexes ont tendance à moins répondre aux enquêtes déplacements. Ils montrent également que les utilisateurs des transports en commun ont tendance à moins répondre aux enquêtes. Dans le cas d'une étude à partir de plusieurs enquêtes, comme celle qui est développée dans le présent travail, le biais de non-réponse peut être amplifié par l'augmentation de celle-ci au cours du temps. En effet, on observe une baisse généralisée des taux de réponse aux enquêtes (Atrostic *et al.*, 2001). Pour les EMD, Bayart (2009) note que le taux de non-réponse était de 47 % en 2006 alors qu'il s'élevait à 35 % en 1995 et que ce taux est supérieur à celui de 1985<sup>10</sup>. Elle remarque que cette baisse peut être attribuable à une certaine lassitude (abondance des démarches de sondage et d'enquêtes commerciales avec des

---

9. Pour les trois dates, la maîtrise d'ouvrage de l'enquête a été assurée par le SYTRAL. En 1985, la Direction Départementale de l'Équipement (DDE) du Rhône assurait la conduite d'opération et le Centre d'Études Techniques de l'Équipement (CETE) de Lyon la réalisation de l'enquête, en collaboration avec l'Institut National de la Statistique et des Études Économiques (INSEE) (CETE de Lyon, 1986). En 1995, la DDE du Rhône, le CETE de Lyon et l'Agence d'urbanisme de la Communauté Urbaine de Lyon assuraient l'assistance à la maîtrise d'ouvrage. L'INSEE assisté par le CETE de Lyon était chargé de la conduite de l'enquête (INSEE *et al.*, 1995). En 2006, l'assistance à la maîtrise d'ouvrage a été assurée par le CETE de Lyon, avec la participation de la DDE du Rhône et de l'Agence d'urbanisme de Lyon. La réalisation des enquêtes sur le terrain a été assurée par la société SOFRECO (CETE de Lyon, 2006).

10. Pour ce dernier on dispose simplement d'un taux de refus, plus restrictif que le taux de non réponse, de 13 %.

questionnaires qui se complexifient pour recueillir davantage de données<sup>11</sup>, pouvant se mesurer en termes de coûts et perte de temps pour les individus, d'après Bradley, 2006) et à l'appréhension de révéler des informations personnelles. Bonnel (2004) remarque qu'il est possible de limiter l'impact de la non-réponse et de son lien avec les comportements de mobilité, lors du redressement des enquêtes, en prenant en compte les variables « explicatives » de la non-réponse. Cela nécessite de faire l'hypothèse que les comportements des personnes ayant répondu et n'ayant pas répondu sont les mêmes, au sein d'un même groupe (caractérisé par les variables retenues). Dans le cas français, Armoogum et Madre (1997) ont utilisé l'algorithme CALMAR qui permet de faire cette correction de la non-réponse pour les EMD.

Au-delà des biais potentiels inhérents à la méthode d'enquête décrits ci-dessus — que l'on pourrait qualifier de structurels et qui sont « prévisibles » (dans le sens où on s'attend à les retrouver même si la méthode d'enquête cherche à les minimiser) — les données peuvent également être affectées par la conjoncture. Lors de la réalisation des EMD, on prévoit usuellement deux semaines supplémentaires d'enquête pour palier aux éventualités climatiques (jours de neige, par exemple), de perturbation des transports (grève TCL), etc. En 1985 et 1995, les aléas ont eu un impact faible. Les rapports sur le déroulement des enquêtes se limitent à signaler que les conditions météorologiques ont été relativement clémentes. En 1995, une grève des transports en commun a fait qu'un jour d'enquête a dû être reporté (CETE de Lyon, 1986 ; INSEE *et al.*, 1995). En revanche, en 2006 la conjoncture a joué un rôle nettement plus important : le début de l'enquête a coïncidé avec une période de mouvements sociaux nombreux suivis d'incidents graves (émeutes) qui imposèrent, dans un premier temps, la suspension de l'enquête dans certains secteurs, puis, rapidement, l'interruption sur l'ensemble du périmètre. Ceci a occasionné un retard dans la collecte, surtout sur les secteurs les plus sensibles de l'agglomération lyonnaise. Une autre conséquence des incidents a été un *turn-over* très important des enquêteurs (il était prévu de recruter environ 200 enquêteurs et il en a finalement fallu près de 500), ce qui a conduit à la réalisation de sessions de formation dans des conditions « d'urgence » (CETE de Lyon, 2006).

N'ayant aucune emprise sur ces facteurs, ni de possibilité d'évaluer leur impact sur les données des EMD, nous nous sommes limités à signaler leur existence potentielle. Les conclusions qui seront tirées en utilisant les EMD dans les prochains chapitres (4, 5 et 6) sont donc établies sous réserve que les biais issus des facteurs exposés ici sont négligeables.

**3.2.1.2.2 Homogénéité des données disponibles** Le respect de la méthode d'enquête « standard Certu » permet à chaque enquête de suivre un cahier des charges identique en ce qui concerne la méthode de collecte des données. En revanche, il est possible que des éléments, comme le périmètre couvert par l'enquête ou certaines variables recueillies, changent d'une année à une autre. Il est donc nécessaire, pour les comparer, de procéder à une homogénéisation.

---

11. La complexification des questionnaires des enquêtes ménages peut trouver des explications à la fois dans les nouveaux besoins de données en matière de modélisation (modélisation par activités, par exemple), mais surtout à la participation de nouveaux acteurs (avec des besoins en données spécifiques) faisant partie du montage institutionnel pour assurer le financement des enquêtes.

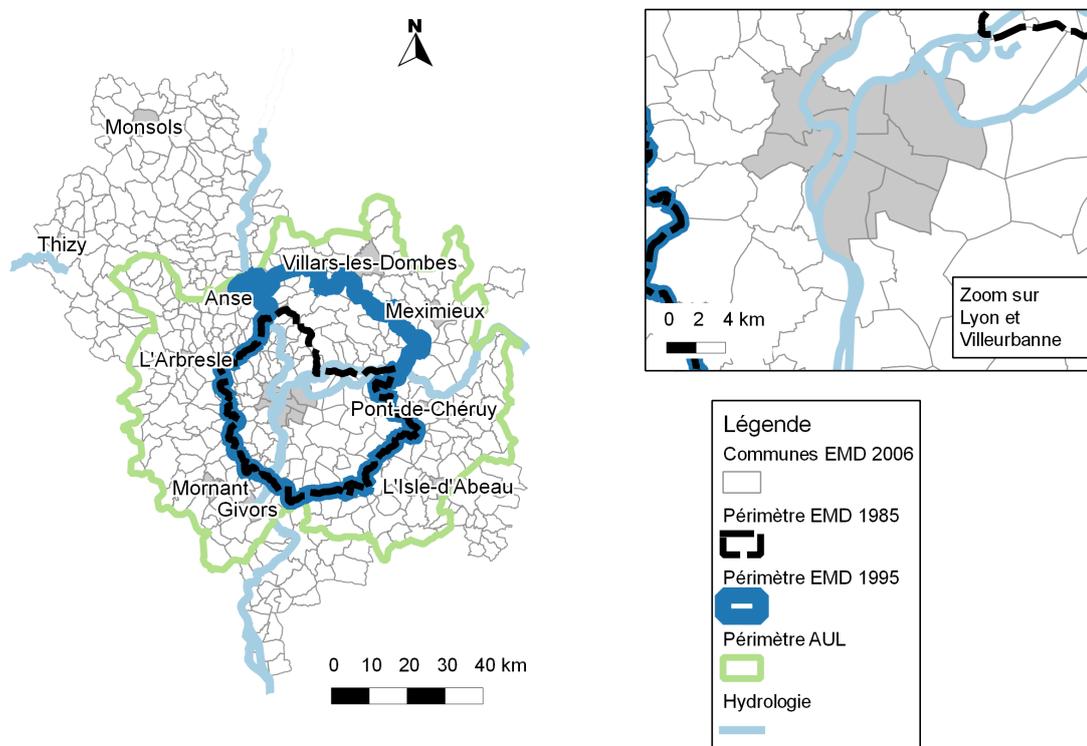
Le premier aspect à prendre en compte est l'étendue géographique des enquêtes. Le périmètre des EMD sur l'agglomération lyonnaise a été considérablement élargi au cours du temps. En 1985, 1995 et 2006 l'enquête portait sur 71, 99 et 453 communes, respectivement. La figure 3.1 donne un aperçu des communes appartenant aux différents périmètres des EMD. Le périmètre de l'enquête de 1985 est proche de celui du Grand Lyon, notamment au nord. Il est un peu plus étendu au sud-ouest et surtout au sud-est. Le périmètre de 1995 contient entièrement le périmètre de 1985 auquel se rajoute une partie de l'Ain, au nord-est. Le périmètre de 2006 est beaucoup plus étendu que celui de 1995 en incluant notamment une grande partie du Beaujolais, au nord-ouest. Dans la carte, nous avons également inclus le périmètre de l'aire urbaine lyonnaise (296 communes) qui, pour des raisons d'ordre pratique que nous développerons un peu plus loin (section 3.3 consacrée aux données relatives aux réseaux de transport), correspond à la limite géographique de la présente étude. Nous travaillons donc sur 3 périmètres : celui de l'EMD de 1985, celui de l'EMD de 1995 et celui de l'aire urbaine lyonnaise. Les analyses des chapitres 4 et 5 feront donc référence, en matière de données à une année d'enquête et à un périmètre<sup>12</sup> (par exemple, « EMD 1995 Périm. 1985 » indique qu'un modèle a été construit à partir des données de l'EMD de 1995 mais en ne retenant que celles correspondant au périmètre de l'EMD de 1985). Notons qu'avec l'élargissement du périmètre d'enquête, il y a également eu une augmentation du nombre de ménages enquêtés. En 1985, l'échantillon contient 4 691 ménages (11 449 personnes). En 1995, on dispose de 5 851 ménages enquêtés sur le périmètre de 1985 et 6 001 sur le périmètre de 1995 (soit 13 579 et 13 997 personnes, respectivement). En 2006, on a 6 169 ménages sur le périmètre de 1985, 6 544 sur le périmètre de 1995, 8 857 sur l'aire urbaine et 11 229 sur le périmètre de 2006 (soit encore 13 586, 14 523, 20 302 et 25 656 personnes, respectivement).

Le deuxième aspect à prendre en compte est la question du zonage. Dans les EMD, les adresses exactes des ménages enquêtés ne sont pas géocodées. La distribution spatiale des résidences et des lieux de déplacement est définie par l'intermédiaire d'un découpage du périmètre d'enquête selon différents zonages. Ceux-ci changent pour chaque nouvelle enquête (les zonages fins sont très différents alors que les secteurs de tirage sont relativement proches, sans être toujours identiques). En réalité, pour comparer des modèles calibrés à plusieurs dates, il est nécessaire de disposer d'un zonage commun pour la définition des différents éléments nécessitant la référence à une origine, à une destination, ou les deux (matrices O-D, par exemple). Autrement, il est possible que les différences observées entre deux calages soient en partie liées aux différences de zonage. Un cas extrême est étudié par Openshaw (1977) qui montre qu'il est possible de calibrer un modèle gravitaire en choisissant des valeurs paramétriques *a priori* et en modifiant le zonage pour faire la correspondance. Cette question est notamment abordée dans la littérature géographique sous le nom de « problème de l'unité surfacique modifiable » (*modifiable areal unit problem*, ou encore MAUP, voir Openshaw, 1984 ; Miller, 1999 ; Briant *et al.*, 2010, pour plus de détail)<sup>13</sup>. Dans le

12. Pour le chapitre 6, l'analyse portera sur un périmètre commun : celui du Grand Lyon. Ceci est lié au fait que seulement ce territoire est desservi par les transports en commun lyonnais (TCL).

13. La présence des effets de la MAUP dans le contexte des modèles de régression linéaire a été mise en évidence par Clark et Avery (1976), Fotheringham et Wong (1991) et Amrhein (1995), entre autres. Le traitement de la question dans le cas des modèles gravitaires, à partir de leur formulation

FIGURE 3.1 – Communes de l'agglomération lyonnaise incluses dans le périmètre des différentes EMD ainsi que de l'aire urbaine lyonnaise (AUL).



contexte spécifique de la définition des zones pour une étude classique de la demande de transports (*Traffic Analysis Zones*), Viegas *et al.* (2009) analysent l'impact de la définition du zonage sur un certain nombre de variables rentrant dans la construction des modèles, à partir de *grid-cells* de différentes tailles. Ils montrent que la définition d'une taille optimale est un arbitrage entre plusieurs facteurs, notamment la part des déplacements intra-zones et le nombre de cases sans déplacements. Dans le cadre de la présente étude, nous avons retenu deux zonages pour l'analyse : le premier donne priorité à un nombre suffisant de ménages par zone, afin d'étudier les modèles de génération et de répartition modale<sup>14</sup>. Il est constitué à partir des découpages en secteurs de tirages des trois EMD. Le second est un découpage à l'IRIS (voir, section 3.3.1.1.1), dont l'objectif est de minimiser les intra-zones et de s'appuyer sur des bases socio-démographiques de l'INSEE.

Le troisième aspect à prendre en considération est la nécessité de vérifier l'homogénéité des différentes variables socioéconomiques, tant en ce qui concerne la présence des mêmes variables sur toutes les enquêtes que sur la compatibilité des modalités pour une variable déterminée. En fonction des partenaires institutionnels (contraintes

issue de la maximisation de l'entropie, avec une prise en compte des tailles des échantillons, a été fait par Batty et Sikdar (1982a,b,c,d).

14. Ce zonage, ainsi que le périmètre du Grand Lyon, est donné dans la figure A.1 de l'annexe A.2, page 213.

de financement des enquêtes), et bien que disposant d'un noyau de variables commun, les EMD peuvent être amenées à inclure des variables reflétant une préoccupation spécifique à une date, qui peut ne pas trouver de correspondance à d'autres dates. On pourrait dire, non sans un certain schématisme, que le nombre de variables augmente avec le temps (un exemple est le gain en détail sur l'accès aux transports en commun dans l'EMD de 2006). Un choix important à évoquer concerne le revenu des ménages, qui est une variable explicative usuellement retenue dans les études de planification. Dans les données des EMD, il n'est disponible qu'en 1995 et 2006. Par ailleurs, en 2006 la part des ménages ayant refusé de répondre sur ce point est assez élevée (plus qu'en 1995). Ces raisons font que l'on ne peut pas le prendre en compte comme variable explicative pour l'analyse diachronique. Lorsque les variables existent sur toutes les enquêtes, il est nécessaire de vérifier que les unités, pour les variables continues (ou du moins quantitatives), et les modalités des variables qualitatives sont compatibles d'une enquête à une autre (unités de mesure, découpage en modalités ainsi que nomenclature de celles-ci). Nous avons réalisé une homogénéisation des modalités de variables, comme l'occupation des personnes dans le ménage et leur catégorie socio-professionnelle. Cette dernière présentait des changements, notamment pour la caractérisation des personnes réalisant des études (tous niveaux). Les variables relatives aux étudiants sont un cas un peu à part en raison de l'échantillonnage. En effet, les résidences universitaires ont été prises en compte dans l'EMD de 1995 mais pas dans celle de 2006, ce qui crée des différences fortes entre les deux années. La comparabilité des données relatives aux étudiants est donc limitée pour des questions de représentativité (différences dans la population visée par l'enquête).

### 3.2.1.3 Calculs de précision

Dans le contexte du présent travail, nous sommes amenés à comparer différents indicateurs calculés à partir des données des EMD lyonnaises. Hormis les biais potentiels décrits dans le point précédent, la précision de ces indicateurs dépendra de la variabilité de données recueillies mais également du plan d'échantillonnage. Idéalement, une étude à partir de données d'enquête devrait prendre en compte ces éléments et fournir des indicateurs de précision. Bonnel (2004) remarque que cette opération est rarement accomplie, que ce soit dans les études (générales) sur la mobilité ou dans les études de modélisation. Dans la présente étude, nous ne traitons que partiellement la question, en fonction des modèles étudiés (chapitre 4). Le calcul des écarts-types prenant en compte le plan de sondage est possible pour les modèles de segmentation ainsi que de régression linéaire par l'intermédiaire du logiciel R (R Development Core Team, 2011) et de sa librairie `survey` (Lumley, 2010).

Dans d'autres cas, le calcul de la variabilité est plus complexe (modèle de distribution, par exemple). Nous nous limiterons alors à utiliser des pondérations normalisées (la somme totale des observations pondérées étant alors égale au nombre d'observations) pour le calcul d'indicateurs. Cette méthode permet de conserver des ordres de grandeur à peu près corrects pour des indicateurs tels que le  $\chi^2$  (voir le chapitre 5) tout en intégrant, dans une certaine mesure, la représentativité de chaque observation.

### 3.2.2 Dynamiques urbaines constatées à partir des EMD

Nous nous limitons dans cette section à présenter des évolutions démographiques et socioéconomiques observables à partir des enquêtes ménages déplacements. Il est nécessaire, bien entendu, de les mettre en relation avec les évolutions de la mobilité. Nous avons opté pour faire la présentation des évolutions dans les comportements de mobilité au début des chapitres 4, 5 et 6, pour pouvoir faire le lien, de manière plus directe, entre les évolutions des comportements de mobilité et la capacité prédictive des modèles étant censés représenter le comportement en question.

La table 3.1 donne les évolutions démographiques et socioéconomiques — pour quelques variables choisies<sup>15</sup> — constatées sur la période d'étude, en prenant en compte les élargissements de périmètre (figure 3.1), à partir des données des EMD de l'agglomération lyonnaise.

Sur le plan démographique, on constate que le nombre de ménages a augmenté d'environ 26 % sur le périmètre de l'EMD de 1985. Le nombre de personnes s'est également accru mais de manière moins importante (un peu plus de 13 %). La taille des ménages a évolué à la baisse puisque la part des ménages composés d'une ou deux personnes s'est accrue, alors que la part des ménages composés de trois ou quatre personnes s'est fortement réduite. On peut voir, dans les EMD de 1995 et 2006, que l'élargissement du périmètre d'analyse se traduit, dans la plupart des cas, par une augmentation de la part des ménages à trois ou quatre personnes et par une diminution de celle des ménages à une ou deux personnes. Au niveau du cycle de vie des ménages, on observe que la part des ménages constitués par une personne (ou des personnes) jeune(s), vivant sans lien conjugal et sans enfant (stade 1) augmente d'environ 44 % entre 1985 et 1995 puis diminue d'environ 11 % entre 1995 et 2006 (à périmètre constant). L'élargissement du périmètre d'analyse se traduit par une forte baisse de la part de cette catégorie, la rapprochant de la valeur constatée en 1985. La part des ménages incluant un couple jeune sans enfant (stade 2) a un peu augmenté entre 1985 et 2006 (augmentation relative d'environ 8 %), sur le périmètre de l'EMD de 1985. Comme pour les ménages au stade 1, ceux au stade 2 sont plutôt localisés à proximité du centre. La part des ménages ayant des enfants (stades 3, 4 et 5 selon l'âge du benjamin) décroît avec le temps, spécialement dans le cas des ménages dont l'enfant le plus jeune a au moins 15 ans (baisse relative d'environ 30 % entre 1985 et 2006 dans le périmètre de l'EMD de 1985), et nonobstant le fait que la prise en compte des périmètres élargis tende à augmenter légèrement leur part (en 2006, la part des ménages au stade 5 connaît une augmentation relative d'un peu moins de 7 % lorsque l'on intègre le reste de l'aire urbaine). La part des ménages au stade 6 (couples plutôt âgés, sans enfant) diminue puis augmente (baisse relative de 9 % puis augmentation de 16 %) à périmètre constant. En 2006, la part de ces ménages est plus importante sur le périmètre de l'aire urbaine que sur celui de

15. Les variables retenues sont celles habituellement retenues dans les études de la demande de transport. Le revenu des ménages n'est pas pris en compte car l'information n'était pas disponible en 1985 (à ce propos, cette variable n'est que rarement prise en compte en même temps que la motorisation, dans les études de la demande de transport, en raison de la forte corrélation entre les deux). La variable « cycle de vie » n'apparaît pas explicitement dans les EMD mais elle a été reconstituée par nos soins à partir de l'âge de la personne de référence du ménage, de son statut marital et de la présence d'enfants dans le ménage (selon les mêmes principes que Kitamura et Kostyniuk, 1986, décrits en annexe B du présent document).

TABLE 3.1 – Évolutions démographiques et socioéconomiques constatées dans l'agglomération lyonnaise à partir des données redressées des EMD, en prenant en compte les différents périmètres d'enquête (figure 3.1).

Caractéristiques des ménages et des personnes	EMD 1985	EMD 1995		EMD 2006		
		P1985	P1995	P1985	P1995	AUL
<b>Ménages</b>						
Total (milliers)	431,2	515,8	536,3	541,7	567,1	698,3
Nb. de personnes (%)						
1	27,1	32,6	32,1	34,8	34,1	31,0
2	30,4	31,1	31,0	32,7	32,6	33,5
3	17,8	15,1	14,9	13,1	13,2	13,8
4	16,7	13,5	14,0	11,8	12,4	13,6
5+	7,9	7,7	8,0	7,6	7,7	8,1
Nb. de pers. trav. (%)						
0	31,5	35,9	35,4	38,4	37,7	36,5
1	36,8	34,5	34,6	33,1	32,8	31,6
2+	31,7	29,6	30,0	28,6	29,5	31,8
Cycle de vie (%)						
1	11,7	16,9	16,5	15,1	14,7	12,7
2	6,1	6,4	6,4	6,6	6,4	6,1
3	12,9	12,4	12,4	11,9	12,2	12,8
4	8,8	8,5	8,8	7,6	7,8	8,5
5	23,1	18,9	19,3	16,2	16,6	17,3
6	19,9	18,0	18,0	20,9	20,9	22,4
7	17,5	18,8	18,6	21,6	21,4	20,2
Nb. de véhicules dispo. (%)						
0	24,7	22,9	22,4	21,7	20,9	18,3
1	49,9	46,5	46,0	47,5	46,9	43,8
2	22,0	25,9	26,8	26,4	27,4	31,6
3+	3,5	4,6	4,9	4,4	4,8	6,3
<b>Personnes</b>						
Total (milliers)	1 088,1	1 220,0	1 280,8	1 231,4	1 301,1	1 654,4
<b>Personnes enquêtées (5 ans et plus)</b>						
Total (milliers)	1 017,9	1 138,2	1 195,2	1 146,5	1 210,0	1 537,6
Âge (%)						
5-14	13,1	13,7	14,0	13,0	13,2	14,0
15-17	5,3	4,4	4,5	4,4	4,4	4,5
18-44	42,9	43,6	43,7	39,5	39,5	38,6
45-64	27,0	25,3	24,9	25,8	25,9	26,6
65+	11,7	13,0	12,9	17,3	17,0	16,3
Occupation (%)						
Autre	1,6	1,4	1,3	1,5	1,5	1,5
Actif employé	44,0	43,5	43,5	43,6	44,0	44,5
Fait des études	24,2	26,4	26,7	24,3	24,4	24,6
Chômeur	3,9	5,0	4,9	4,1	4,0	3,8
Retraité	15,1	16,1	16,0	22,3	22,0	21,6
Reste au foyer	11,1	7,6	7,7	4,1	4,1	4,0
<b>Personnes majeures (18 ans et plus)</b>						
Total (milliers)	830,1	932,3	974,4	947,1	996,8	1 252,7
Sexe (%)						
Hommes	47,1	46,6	46,9	46,6	46,7	47,0
Femmes	52,9	53,4	53,1	53,4	53,3	53,0
Permis de conduire (%)						
Titulaire du permis	72,0	78,8	79,2	84,2	84,7	86,2
Pas de permis	28,0	21,2	20,8	15,8	15,3	13,8

Note de lecture : le détail de construction de la variable Cycle de vie est donné dans l'annexe B (page 214).

Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006.

l'EMD de 1985. Cette configuration correspond vraisemblablement à celle de ménages dont le domicile est situé en périphérie en raison de la présence d'enfants (comme pour les stades 3 à 5), ces derniers ayant quitté le domicile familial dans un passé relativement proche. La part de ménages se trouvant au septième stade du cycle de vie (adultes plutôt âgés sans lien conjugal) augmente sur la période étudiée de plus de 23 %. Ils sont localisés plutôt dans le périmètre de l'EMD 1985, ce qui peut s'expliquer principalement par l'étalement urbain (ils ont acheté leur lieu de résidence à une époque où la ville était moins étendue) et éventuellement par le changement de résidence. Cette augmentation reflète un vieillissement de la population, qui peut être observé, de manière plus directe, sur les évolutions des parts des classes d'âge où l'on peut constater une très forte augmentation de la part des personnes âgées de 65 ans ou plus (augmentation relative de 48 % entre 1985 et 2006, à périmètre constant). Cette augmentation se fait bien entendu au détriment des autres classes d'âge, avec une répartition plutôt équilibrée (nous entendons par là que toutes les classes sont touchées). Finalement, la répartition des adultes par sexe n'a presque pas changé sur la période analysée.

Sur le plan socioéconomique, le vieillissement de la population, observé dans de qui précède, se traduit directement par une forte augmentation de la part des retraités (augmentation relative d'environ 48 % entre 1985 et 2006, sur le périmètre de l'EMD de 1985). On remarque que l'augmentation a surtout eu lieu entre 1995 et 2006. En 2006, la part des retraités décroît très légèrement lorsque l'on prend en compte les données correspondant au périmètre élargi. L'augmentation de la part des retraités semble avoir un impact relativement faible sur le nombre d'actifs employés. La part de cette catégorie reste relativement stable sur l'ensemble de la période (autour de 44 %). La même remarque peut être faite pour les chômeurs. En revanche, la part des personnes au foyer diminue considérablement entre 1985 et 2006 (baisse relative d'environ 63 %). L'augmentation des départs à la retraite a donc été compensée, au niveau de la part de la population active, par la diminution des personnes restant au foyer ; autrement dit, par la généralisation du travail féminin. Au niveau de la composition des ménages en termes de nombre de personnes actives ayant un emploi, on remarque une augmentation de la part des ménages n'ayant pas de personnes actives. La part des ménages ayant une ou deux personnes actives décroît, notamment celle des ménages à une personne active. Pour les ménages ayant deux personnes ou plus activement employées, on aurait pu s'attendre à voir leur part augmenter compte tenu de la baisse de la part des personnes au foyer. En réalité, l'augmentation de la participation des femmes au marché du travail combinée avec la réduction de la taille des ménages, fait que l'on observe une baisse (légère) de la part des ménages à deux travailleurs ou plus.

Au niveau de l'accès à la voiture, on observe une augmentation de l'équipement des ménages en véhicules disponibles. En effet, on observe une diminution relative de la part des ménages non motorisés d'un peu plus de 12 % entre 1985 et 1995. L'élargissement du périmètre d'analyse, en 1995 et en 2006 se traduit par une réduction encore plus importante de la part des ménages non motorisés. Alors que la part des ménages n'ayant à disposition qu'un seul véhicule diminue très légèrement entre 1985 et 2006, la part des ménages ayant deux ou plus véhicules disponibles connaît une augmentation relative d'environ 20 %, sur le périmètre de l'EMD de 1985. Lorsque l'on compare les résultats pour l'ensemble des données en 1985 et en 2006,

on observe que la part des ménages n'ayant pas de véhicule disponible et celle des ménages ayant deux véhicules ou plus sont à peu près similaires. Elles sont, chacune, égales à environ la moitié de la part des ménages ayant un seul véhicule disponible. En 2006, sur le périmètre de l'aire urbaine lyonnaise, les parts des ménages avec un véhicule à disposition et celle des ménages avec deux véhicules ou plus sont proches (la part des ménages « mono-motorisés » est un peu plus élevée) et elles sont égales, chacune, à plus du double de la part des ménages non motorisés. L'augmentation de l'équipement en matière de véhicules s'accompagne d'une réduction de la part des personnes majeures n'ayant pas de permis de conduire (baisse relative d'un peu plus de 43 %, entre 1985 et 2006 sur le périmètre de 1985).

Les changements constatés ci-dessus, pour la période étudiée, correspondent bien à ceux qui ont pu être observés dans d'autres villes des pays développés. On retrouve l'inexorable vieillissement de la population qui se traduit par l'accroissement du nombre de retraités alors que, parallèlement, le travail féminin se généralise. L'évolution de la part de la population active résulte de l'importance relative de ces deux phénomènes : dans le cas de l'agglomération lyonnaise, entre 1985 et 2006, l'augmentation de la part des retraités a totalement été compensée par l'augmentation de la participation des femmes dans le marché du travail (donnant lieu à la forte baisse de la part des personnes restant au foyer). En même temps, la taille des ménages décroît avec l'augmentation des ménages uni-personnels et la baisse du nombre d'enfants. La croissance de la population urbaine, liée à l'attractivité des centres urbains, se traduit par l'étalement de la ville, en concomitance avec un accès plus facile à l'automobile (augmentation des véhicules disponibles et de la part des personnes majeures ayant le permis), permettant l'accès à des zones plus éloignées et moins denses (moins chères)<sup>16</sup>.

### 3.3 Données relatives à l'offre de transport

Nous avons établi dans la section 1.1 que l'offre de transport a des caractéristiques multiples et spécialisées, qui correspondent à l'« entité » à transporter. Une fois que l'on admet que le moyen de transport analysé peut effectivement être utilisé, l'élément principal intégré dans les modèles à quatre étapes est le temps généralisé, qui combine le temps de parcours ainsi que d'autres éléments caractéristiques de la qualité du service<sup>17</sup>. Pour la présente étude, il est nécessaire de caractériser l'offre de transport correspondant aux différentes dates pour lesquelles les données de demande sont disponibles. Il est également nécessaire de vérifier que les mesures de temps utilisées peuvent être comparées entre deux dates. En l'absence de données de temps

---

16. L'expression « en concomitance » a été choisie ici car elle ne présuppose pas un sens de causalité entre l'étalement urbain et l'accès à la voiture. Ces deux éléments font l'objet d'une interaction complexe, en présence d'effets de rétroaction, dont l'analyse complète va au-delà de la portée de la présente étude.

17. Les concepts de temps généralisé et de coût généralisé sont deux facettes de l'idée selon laquelle on peut combiner les différents éléments caractérisant le service de transport au sein d'un seul indicateur. Celui-ci résulte de la « transformation » des différentes caractéristiques de l'offre en équivalents temporels ou monétaires, qui peuvent ensuite être additionnées les uns aux autres (Bonnell, 2004 ; Crozet, 2005). On retrouve donc derrière ces indicateurs la notion de substituabilité, bien connue des économistes.

comparables pour chaque date d'enquête, les variations des valeurs des paramètres des différents modules de la séquence de modélisation les prenant en entrée peuvent ne pas traduire seulement des changements dans les comportements de mobilité des agents de l'aire étudiée. Elles peuvent, tout simplement, être le reflet des inconsistances entre les mesures de temps utilisées pour caractériser l'offre de transport aux différentes dates. Ainsi, par exemple, pour le cas du module de distribution, Southworth (1979, 1983) a montré que les paramètres du modèle gravitaire sont très sensibles à la mesure de séparation — temps ou distance — utilisée lors de la calibration. Au niveau du module de répartition modale, Train (1978) a mis en évidence l'impact de la mauvaise estimation des temps de marche à pied sur la stabilité des paramètres d'un modèle désagrégé.

Une première alternative pour caractériser les temps de parcours aurait été de s'appuyer sur les données contenues dans les EMD. Cette solution a très vite été écartée. En effet, les temps déclarés provenant des différentes enquêtes peuvent ne pas être comparables entre eux, notamment lorsque les besoins de désagrégation des données sont importants<sup>18</sup>. La deuxième alternative pour caractériser l'offre de transport est celle qui est généralement adoptée dans la pratique : on modélise les réseaux de transport à partir de graphes, dont les différents éléments constitutifs (arcs, nœuds. . .) permettent de représenter des aspects particuliers de l'offre (capacité, vitesse de circulation, sens de circulation, mouvements prioritaires aux carrefours. . .). Nous avons vu dans le chapitre 2 que les études sur la transférabilité temporelle des modèles de prévision des déplacements se sont souvent appuyées sur des temps produits à partir de réseaux codifiés existants au préalable (issus d'études de trafic ou patrimoine des unités de recherche des universités ou des agences de planification des villes concernées). Il est apparu qu'une telle donnée était indisponible en France<sup>19</sup>. Une grande partie de ce travail a donc été consacrée à la production de données « consistantes » pour caractériser les réseaux de transport (réseau routier et transport en commun urbains) sur la période étudiée. La propriété de consistance fait ici référence à l'utilisation d'un même outil pour mesurer l'évolution d'une variable entre deux dates.

La codification de réseaux étant un travail lourd et extrêmement chronophage, il semblait raisonnable de prendre appui sur des efforts de codification existants, plutôt que de se lancer dans la codification de réseaux *ex nihilo*. Ainsi, nous avons décidé de prendre comme point de départ — par la suite l'effort de codification s'est peu à peu mutualisé — les réseaux de référence codifiés pour le projet SIMBAD (Nicolas *et al.*, 2009 ; Nicolas, 2011) pour ensuite y intégrer les différents changements constatés entre les dates des EMD. L'utilisation des réseaux de référence du projet

---

18. Si, les temps déclarés sont consistants au niveau agrégé, ils le sont beaucoup moins à un niveau plus désagrégé. Le faible nombre de personnes enquêtées par O-D rend leur utilisation délicate en modélisation. Ceci s'explique par les différences de perception des personnes enquêtées et par la précision avec laquelle ils sont habitués à mesurer et parler de leurs déplacements (à la minute près, à la demi-heure près). On parle d'effet d'arrondis.

19. Si les différents organismes chargés de la gestion des transports particuliers et collectifs disposent de réseaux codifiés, ceux-ci correspondent à l'état actuel du réseau sans qu'il y ait vraiment une procédure d'archivage définie. Il faut également signaler qu'en France, chaque organisme gère le réseau sur lequel il est compétent, sans qu'il y ait vraiment de coopération entre les différents organismes. De ce fait, la comparabilité entre des réseaux produits par des organismes différents semble douteuse.

SIMBAD apportait deux avantages majeurs liés à la méthodologie appliquée lors de la codification : le premier étant d’explorer les potentialités offertes par les bases de données géographiques existantes — que ce soient par exemple des bases comme Navteq© ou Géoroute© pour la reconstitution du réseau routier, ou les données mises publiquement à disposition par les gestionnaires en ce qui concerne le réseau de transports collectifs — et le deuxième étant d’automatiser le plus possible la construction du réseau. Ces deux avantages permettent d’assurer une reproductibilité de la démarche de codification dans le temps. Il est nécessaire de signaler ici que le choix des réseaux codifiés pour le projet SIMBAD comme réseaux de référence impose deux contraintes pour ce travail. La première contrainte tient au fait que les différents outils logiciels (systèmes d’information géographique spécialisés dans les transports ou SIG-T) sont commerciaux et que, de ce fait, il est difficile d’utiliser des réseaux codifiés avec un logiciel sur un autre logiciel. Les réseaux du projet SIMBAD ont été codifiés en utilisant le logiciel DAVISUM<sup>20</sup>. La codification de l’historique des réseaux est héritière de ce support. La deuxième contrainte correspond au choix du périmètre et du niveau d’agrégation pour l’analyse.

Cette section présente la méthodologie retenue pour construire les réseaux de référence et leurs évolutions dans le temps. Nous commençons par aborder la codification de l’historique du réseau routier. Nous décrivons ensuite celle du réseau de transports collectifs. Les descriptions des codifications des deux réseaux sont bâties sur le même schéma, avec un exposé des principes généraux de codification et leur traduction au sein des réseaux de référence de SIMBAD (principes d’automatisation et résultats de calage) suivi d’une présentation de la codification de l’historique à partir des réseaux de référence. Cette présentation reprend en grande partie des éléments de Godinot et Bonnel (2008), et de Nicolas *et al.* (2009) concernant la codification des réseaux de référence.

### 3.3.1 Codification du réseau routier : vers une codification automatique ?

Cette section présente, dans un premier point, les principes utilisés pour la codification du réseau routier de référence. Un second point est consacré à l’application de ces principes pour codifier les réseaux correspondant aux années des EMD.

#### 3.3.1.1 Méthodologie de la codification des réseaux routiers de référence

##### 3.3.1.1.1 Niveau d’agrégation

**Principe** Nous avons vu dans la section 1.1.2 que la modélisation des réseaux répond au besoin de prendre en compte l’espace dans l’analyse des transports. Cette prise en compte passe par la définition de zones d’analyse, permettant de traduire les localisations des différents lieux d’activités. Le réseau modélisé reflète les possibilités de se rendre d’une zone à une autre. Les choix de taille des zones et de finesse de

---

20. Une présentation des principaux logiciels de modélisation, SIG-T, utilisés en France peut être trouvée dans Du Crest et Clément (1999) et dans Bhouri (2002).

représentation des réseaux doivent donc être spécifiés par l'analyste en fonction des objectifs de la modélisation, en veillant à conserver une cohérence entre les deux (Bonnell, 2004). En général, on peut dire que la représentation du réseau est aussi fine que les données — ainsi que les autres contraintes relatives à l'exercice de prévision — le permettent (Ismart, 1990), sachant que celles-ci doivent permettre, dans le cas de réseaux codés avec beaucoup de détail, de simuler ce qui se passe à une telle échelle (mouvements tournants, feux, effets de remontée de queues de bouchons...).

**Choix du projet SIMBAD** Dans le projet SIMBAD, le choix a été fait de retenir un zonage et une codification du réseau routier très fins malgré « un objectif de simulation agrégée de long terme, où il n'est pas question de simuler la charge ou le fonctionnement d'une infrastructure particulière, mais de produire des indicateurs permettant d'évaluer des politiques de transports à l'aune du développement durable à l'échelle de l'agglomération » (Nicolas *et al.*, 2009).

En réalité, la contrainte temporelle a été déterminante. La création d'un réseau sur une étendue aussi vaste<sup>21</sup> représente un temps de travail trop important au regard du temps disponible sur le projet. L'objectif de SIMBAD sur ce point a été de tester la capacité des bases de données existantes à représenter les réseaux, tant pour le réseau routier que pour celui de transport collectif. Cette volonté s'est traduite par un traitement des réseaux, tels que définis dans les bases de données existantes, de la manière la plus automatique possible. Or, les agrégations de liens aux caractéristiques différentes nécessitent toujours un traitement particulier dont l'automatisme ne peut qu'être limitée. En conservant la totalité du réseau, il était possible de décrire chacun des liens à partir des informations disponibles sur les infrastructures correspondantes.

Afin de conserver une certaine cohérence entre le niveau de détail du réseau et la taille des zones d'analyse, le découpage retenu est le plus fin disponible, à savoir le découpage en IRIS<sup>22</sup>. Il s'agit d'un découpage en 777 zones pour l'aire urbaine lyonnaise (voir figure 3.2).

Toutefois, si la représentation est très détaillée, au sens où l'ensemble des voiries sont représentées, y compris les plus fines disponibles dans les bases de données de réseau, elle ne l'est pas du tout au niveau de la mesure des attributs objets permettant de représenter ces infrastructures. De ce fait, la précision du réseau routier du projet SIMBAD est largement illusoire. L'analyse des résultats doit donc se limiter à une analyse macro à l'échelle de l'agglomération. Il est possible d'utiliser le réseau modélisé pour obtenir des temps de parcours ou des temps généralisés entre secteurs de l'agglomération. En revanche, il ne serait pas raisonnable d'utiliser des résultats de charge sur une infrastructure ou sur un projet de transport particulier.

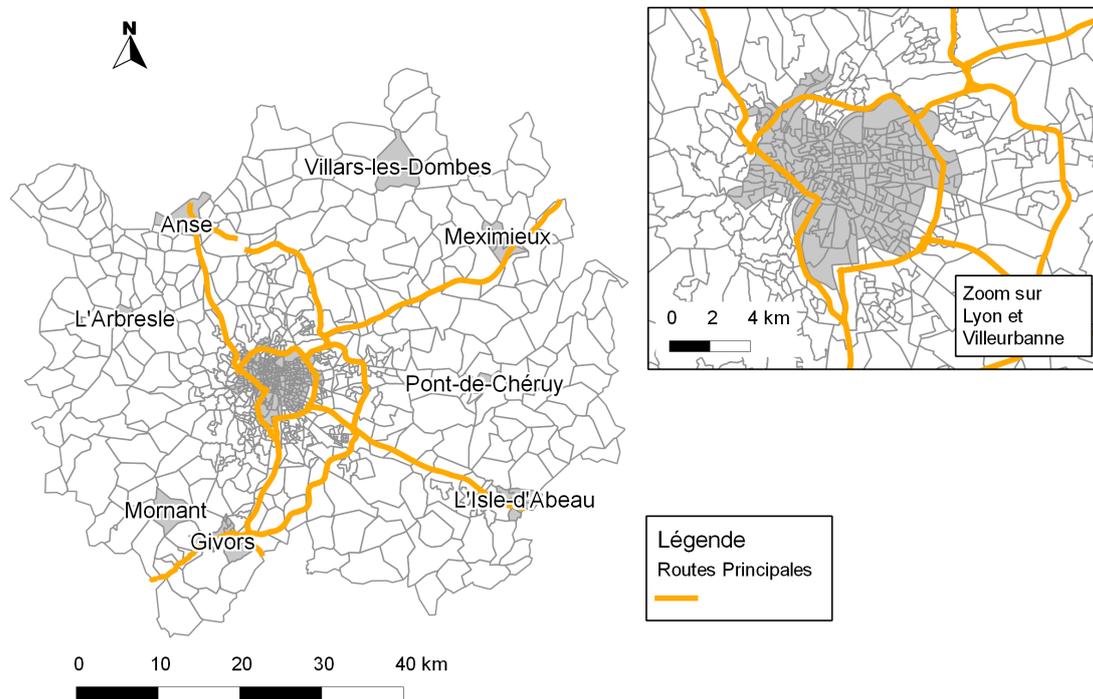
**Impact pour notre étude** Le périmètre d'étude du projet SIMBAD couvre l'ensemble des périmètres des EMD de 1985 et de 1995. En revanche, il est limité par rapport à celui de l'EMD de 2006. Notre utilisation de ces réseaux devra donc se limiter au périmètre de l'AUL lorsque l'analyse portera sur les données de 2006.

---

21. Le choix retenu dans le projet SIMBAD est de travailler sur le périmètre d'une aire urbaine (celle de Lyon, voir figure 3.2). Ceci permet d'envisager de manière convenable les questions d'étalement urbain et leurs conséquences sur les consommations d'énergie et les émissions de gaz à effet de serre liées aux mobilités qui s'exercent sur un tel territoire (Nicolas *et al.*, 2009).

22. Niveau le plus fin de divulgation des données socio-démographiques par l'INSEE.

FIGURE 3.2 – Périmètre et zonage retenus pour le projet SIMBAD : 777 IRIS de l'Aire urbaine lyonnaise en 1999.



Les modèles que nous étudions ici, notamment ceux de distribution et de répartition modale utilisent en entrée des données sur les temps généralisés. Ces données sont issues des réseaux codifiés avec la méthodologie retenue pour SIMBAD. Leur utilisation est donc possible dans le cadre du présent travail. En revanche, comme nous l'avons signalé lors de la description des EMD et de leur comparabilité temporelle, la prise en compte des différences de zonage est importante. Il est nécessaire d'établir une méthode permettant de rendre compatibles des temps établis à l'échelle de l'IRIS avec d'autres échelles potentielles d'analyse. Dans la littérature relative aux modèles de distribution ou de choix modal, ces détails techniques sont rarement discutés. En général, deux types de méthodes sont retenus. Le premier consiste à faire une moyenne des différents temps. Cette méthode pose bien entendu le problème des pondérations à retenir pour la moyenne. La deuxième méthode consiste à utiliser tels quels les temps entre les IRIS dont les centroïdes sont les plus proches de ceux des zones d'analyse (Openshaw, 1977). Cette dernière méthode sera retenue.

**3.3.1.1.2 Représenter les infrastructures par des arcs et des nœuds** La description d'un réseau pour les besoins de modélisation nécessite la description de sa structure (spécification de ses composantes comme, par exemple, les sections de voies, les intersections...), des propriétés de ses composantes (longueur, vitesse...) et des relations entre ces éléments et les flux de déplacements (Lamb et Havers, 1970). Elle s'appuie sur la théorie des graphes. Les réseaux sont ainsi composés d'arcs (liens) et de nœuds qui peuvent être de différents types et posséder un certain nombre

d'attributs. On distingue principalement deux types de liens : les tronçons ou liens de réseau et les connecteurs de centroïde. Il y a également deux types de nœuds : les nœuds de réseau et les centroïdes de zone. Cette section se concentre sur les tronçons et les nœuds de réseau.

Les liens de réseau permettent de représenter les infrastructures de transport. Dans le cas d'une description fine, ils peuvent représenter les différentes voiries existantes. Dans le cas d'un réseau plus agrégé, ils traduisent plus les fonctionnalités d'un groupe d'infrastructures que ses caractéristiques physiques. Le lien est doté d'un certain nombre d'attributs qui permettent de calculer, en particulier, le temps généralisé nécessaire pour le parcourir. Généralement, on trouve les attributs suivants :

- la longueur ;
- la vitesse à vide : c'est-à-dire la vitesse en l'absence de toute circulation sur le lien. Cette information est parfois remplacée par le temps de parcours à vide, puisque la vitesse sert à calculer le temps de parcours à partir de la longueur du lien ;
- la capacité : elle est généralement fournie en nombre d'unités de voitures particulières par heure<sup>23</sup>. Il s'agit de la capacité maximale pouvant être absorbée par le lien ;
- la tarification : elle permet de prendre en compte un éventuel péage. Elle est à distinguer du coût d'usage qui est généralement kilométrique et n'est donc pas défini au niveau du lien, même s'il requière sa longueur.

Le nœud de réseau correspond à une intersection entre des liens de réseau. Il permet d'assurer la connexité du graphe. Il correspond d'une certaine manière au carrefour du réseau viaire, même s'il n'a pas forcément une traduction physique directe dans un réseau agrégé. On peut lui associer les caractéristiques suivantes :

- les mouvements possibles entre les liens connectés au nœud : cette information permet par exemple de représenter des interdictions de tourner à droite ou à gauche ;
- la nature du carrefour : carrefour à feux, priorité à droite, stop... Dans ce cas, l'information est généralement complétée par un certain nombre d'attributs permettant de calculer le temps nécessaire pour traverser le carrefour, par exemple durée de cycle du feux et temps de vert disponible, lois d'écoulement pour un stop...

**Une utilisation des bases de données existantes** Pour le réseau routier, les deux principales bases existantes ont été acquises et examinées à savoir les bases GéoRoute©<sup>24</sup> et Navteq©<sup>25</sup>. Après analyse, l'utilisation de la base Navteq© a été privilégiée, tout en intégrant, chaque fois que cela semblait utile, les informations contenues dans la base GéoRoute©. En réalité, cette dernière ne répond pas suffisamment aux besoins de modélisation d'un réseau routier pour deux raisons :

---

23. L'Unité de Voiture Particulière ou UVP est un indicateur permettant d'agréger différents types de véhicules pour tenir compte de leur occupation différenciée de l'espace de voirie. Par exemple, un poids lourds équivaut souvent à 2 ou à 3 UVP selon sa taille.

24. Base contenant les informations sur les réseaux routiers, produite par l'IGN.

25. Navteq est une société, subsidiaire de Nokia, qui fournit des cartes numériques et d'autres contenus dynamiques pour la navigation GPS et les services de localisation.

- si l'ensemble des voiries est bien représenté avec une classification hiérarchique, aucune information sur les capacités et les vitesses à vide, ni sur les mouvements tournants aux nœuds n'est fournie ;
- le maintien de cette base dans le futur par l'IGN semble compromis alors qu'il faut assurer l'évolution du réseau à partir de l'évolution du contenu de ces bases.

La base Navteq© contient la totalité du réseau viaire de l'aire urbaine y compris les voies de desserte locale. Elle contient un peu plus de 113 000 liens et près de 45 000 nœuds. Elle contient les coordonnées des extrémités des liens, le nombre de voies de circulation, le sens de circulation, la capacité journalière, la vitesse à vide et les mouvements possibles aux nœuds (extrémités du lien). Elle permet donc de disposer, en théorie, de toutes les informations nécessaires à l'estimation du temps de parcours du lien en fonction de la congestion, à l'exception de la courbe débit-temps. Elle fournit également une typologie des liens en 5 niveaux hiérarchiques, allant des voies autoroutières les plus structurantes aux plus petites voies de desserte locale. Celle-ci présente cependant certaines limites — notamment un mauvais classement de certaines voies structurantes et une continuité des types mal assurée — qui ont rendu nécessaire un retraitement.

**Typologie des liens** La codification d'un réseau passe par la définition d'une typologie des liens qui répond notamment à deux objectifs :

- regrouper les liens les plus proches en termes fonctionnels afin de leur affecter des attributs communs (dont les courbes débit-temps) ;
- permettre un regroupement des liens selon certaines caractéristiques afin de faciliter l'analyse des résultats de l'affectation des matrices de demande.

La typologie de la base Navteq© n'étant pas tout à fait adaptée aux besoins de la modélisation, elle a été partiellement reprise dans le cadre de SIMBAD. La codification des trois niveaux hiérarchiques les plus élevés de la typologie Navteq© a été vérifiée, pour s'assurer de la cohérence de la codification afin que les liens composant une voie autoroutière ou routière appartiennent tous au même type tant que les caractéristiques de la voie restent inchangées.

La validation de la typologie a également conduit à retravailler la distinction opérée entre liens « en ville » et « hors ville ». Pour cela, la définition de la tache urbaine contenue dans la base Géoroute© a été utilisée. Tous les liens dont une extrémité est incluse dans la tache urbaine sont codifiés « en ville », ceux dont les deux extrémités sont hors de la tache urbaine sont codifiés « hors ville ». Cette distinction est utilisée dans la définition des attributs des liens et surtout des nœuds. Elle a conduit à une réduction des vitesses pour les tronçons recodés « en ville ». La table 3.2 présente les résultats de la typologie (Godinot et Bonnel, 2008).

**Choix des courbes débit-temps et des capacités en heure de pointe** La base Navteq© contient certains attributs des liens, mais pas tous ceux nécessaires à la modélisation. Il manque la courbe débit-temps qui permet de tenir compte de la congestion dans le calcul des temps de parcours.

La capacité disponible dans la base Navteq© est une capacité journalière. Or, afin de pouvoir estimer correctement la congestion en heure de pointe, une capacité

### 3. Terrain d'étude et données pour la modélisation

TABLE 3.2 – Typologie des liens de réseau

N° de Type		Niveau hiérarchique	Descriptif	Nombre de liens	
Base initiale	Base finale			Base Navteq©	Base SIMBAD
0 à 9	0 à 9		Voie fermée	13 047	13 001
de 10 à 17	de 10 à 17	1	Hors ville, Autoroutes et VRU	840	876
de 18 à 25	de 18 à 25	1	En ville, Autoroutes et VRU	0	389
de 20 à 29	de 27 à 31	2	Hors ville	1 299	6 325
de 30 à 39	de 30 à 39	2	En ville	17 127	8 923
de 40 à 49	de 40 à 49		Échangeurs, bretelles	684	685
répartis sur 9 types	de 50 à 53		Giratoires, Rond-point	0	1 672
de 54 à 57 et 90	de 54 à 57	3	Hors ville, Autres routes classifiées	7 942	7 723
de 65 à 66	de 64 à 67	3	En ville, Autres routes classifiées traversant la localité	396	2 179
79	79	4	Hors ville Routes connexions communes (souvent limitation poids)	395	179
de 70 à 78	de 74 à 75	4	En ville, Autres routes principales	10 974	9 343
	82	5	Hors ville, routes connexions principales, Zone 30, rues résidentielles	0	5 530
83	82 et 90	5	En ville, routes connexions principales, Zone 30, rues résidentielles	60 836	56 696
	95		Liens Train		
	96		Liens Métro		
	97		Liens Tramway		6
	98		Liens Funiculaire		
	99		Liens Bus et Navette		202
29 types + 0 type TC	42 types + 5 types TC		Total	113 540	113 546

Source : d'après Godinot et Bonnel (2008)

horaire maximale est nécessaire. La capacité journalière a donc été divisée par 10, ce qui est une pratique habituelle des bureaux d'études. Cette capacité horaire a été ajustée, pour les axes principaux, en utilisant les données de comptage produites par les gestionnaires des différents réseaux routiers<sup>26</sup>.

Faute de données pour caler les courbes débit-temps, le parti pris dans SIMBAD a été de s'appuyer sur les typologies retenues par la société PTV qui distribue et a utilisé DAVISUM dans plusieurs études récentes en France. Les courbes retenues sont de type *BPR* (*Bureau of Public Roads*, développées aux États-Unis dans les années soixante), et plus précisément, les courbes *BPR2* qui permettent de distinguer la

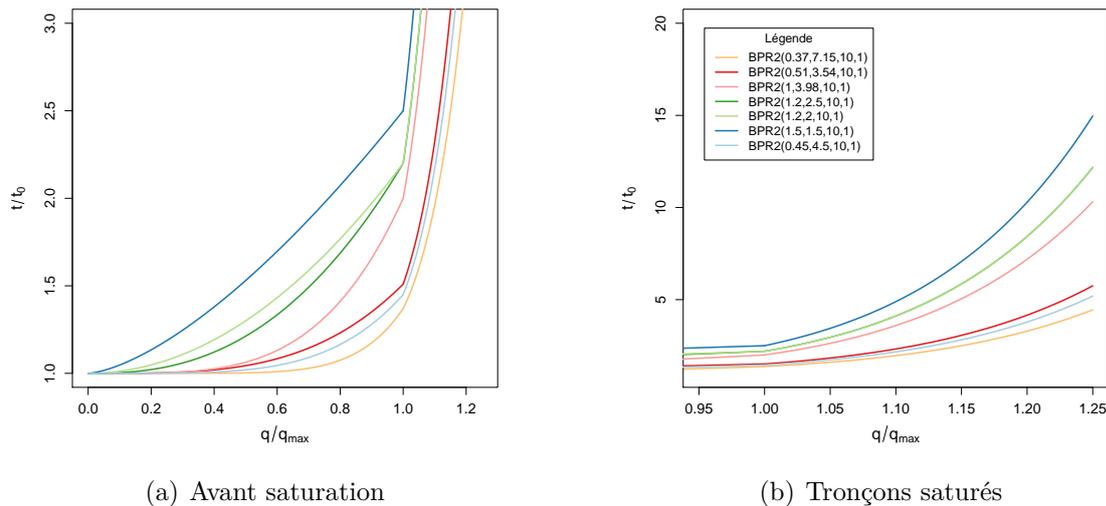
26. Ont été corrigées ainsi certaines capacités qui s'avéraient manifestement trop faibles. Le traitement des sur-capacités au regard des données de trafic s'avère plus délicat car cela peut correspondre à la « réalité », même en présence de congestion. La congestion peut ne pas être due à une capacité insuffisante au droit du comptage, mais peut correspondre à une contrainte en aval.

situation non saturée de la situation saturée (Godinot et Bonnel, 2008) :

$$t = BPR2(a, b, b', c) = \begin{cases} t_0 \left( 1 + a \left( \frac{q}{q_{max} \cdot c} \right)^b \right) & \text{pour } \frac{q}{q_{max} \cdot c} \leq 1 \\ t_0 \left( 1 + a \left( \frac{q}{q_{max} \cdot c} \right)^{b'} \right) & \text{pour } \frac{q}{q_{max} \cdot c} > 1 \end{cases}$$

où  $t$  est le temps de parcours sur un lien ;  $t_0$  est le temps de parcours à vide ;  $q$  le volume (charge) sur le lien ;  $q_{max}$  le volume maximum, c'est-à-dire la capacité maximale du lien.  $a$ ,  $b$ ,  $b'$  et  $c$  sont des paramètres de calage. La figure 3.3 donne l'allure des courbes utilisées — en distinguant la partie « avant saturation » 3.3(a) de la partie « après saturation » 3.3(b) — et la figure 3.4 donne la correspondance entre les courbes retenues et les types de tronçons. On peut observer que, pour certaines voies, les coefficients retenus conduisent à une augmentation brutale des temps lorsque le seuil de saturation ( $q/q_{max} = 1$ ) est atteint.

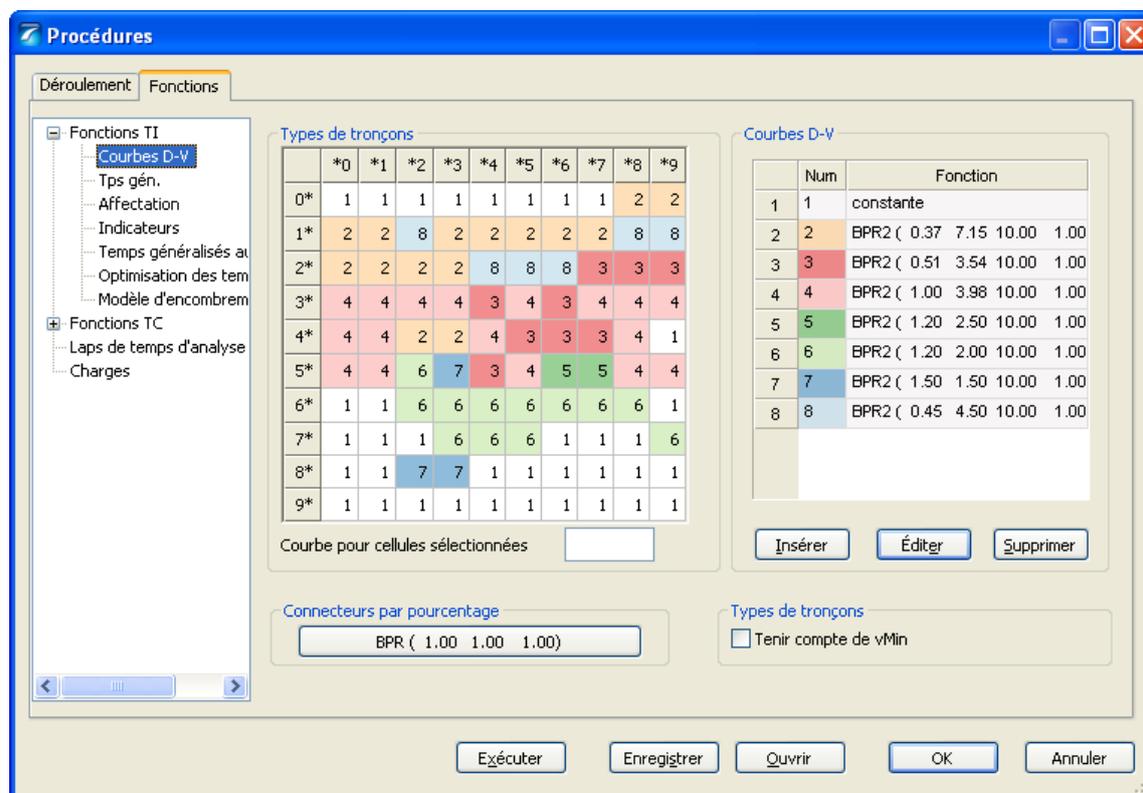
FIGURE 3.3 – Courbes débit-temps retenues dans SIMBAD en distinguant la partie « avant saturation » et la partie « après saturation ».



**Typologie des nœuds et mouvements tournants aux carrefours** Les nœuds du réseau initial Navteq© ne sont pas typés. Rien ne distingue un nœud constitutif d'un carrefour d'un nœud constitutif d'une simple jonction entre deux liens et rien ne distingue un type de carrefour d'un autre. Cette information est pourtant nécessaire afin d'appliquer un temps de traversée du nœud qui dépende de la nature du mouvement à ce nœud (mouvement tout-droit ou mouvement tournant, circulation dans le flux principal du carrefour ou dans un flux secondaire, nature du carrefour à feux, stop, giratoire...).

Les nœuds correspondant à de simples points graphiques sont classés comme tels (type 1) et les nœuds situés en fin de voirie comme nœuds « terminaux » (type 10). Ensuite le recours à la base de données GéoRoute© a permis de dégager des types de nœuds particuliers comme les nœuds constitutifs de rond-point, d'accès autoroutier ou de bifurcations sur ce réseau.

FIGURE 3.4 – Courbes débit-temps associées aux différents types de liens (capture d'écran DAVISUM).



Un plan de positionnement des feux et des stops sur l'aire urbaine de Lyon n'étant pas disponible, différentes hypothèses ont été faites afin de typer les autres nœuds du réseau. Un traitement différencié a été appliqué aux intersections situées dans et hors de la tache urbaine. La figure 3.5 présente les principes de codification retenus (Godinot et Bonnel, 2008).

Au sein de DAVISUM, deux éléments doivent compléter cette typologie afin de déterminer la pénalité de temps au passage du nœud. Le premier concerne la définition du flux principal au niveau du carrefour. Ce flux principal correspond généralement au mouvement le plus important au sein du carrefour. Il conduit à classer les mouvements au carrefour en 4 types :

- P–P : mouvement correspondant au flux principal ;
- P–S : mouvement du lien inclus dans le flux principal vers un lien n'appartenant pas au flux principal ;
- S–P : mouvement d'un lien n'appartenant pas au flux principal vers un lien inclus dans le flux principal ;
- S–S : mouvement entre deux liens qui ne sont pas inclus dans le flux principal.

Cet attribut est présent au sein de la base Navteq©. Toutefois, les incohérences sont très fréquentes. La correction de la base est nécessaire pour assurer la continuité des axes principaux en tant que mouvement principal. Toutefois, cette correction ne peut être réalisée qu'au niveau de chacun des carrefours. Elle s'est limitée aux voies de niveaux hiérarchiques les plus élevés (1 à 3, Godinot et Bonnel, 2008). Le second élément concerne la définition des mouvements tournants. DAVISUM propose 4 types

FIGURE 3.5 – Récapitulatif de l’affectation des nœuds aux types de carrefours.

		Type	Nb. noeuds						
Nœuds graphiques n°1		n°1 :	9 565						
Nœuds terminaux n°10		n°10 :	7 825						
Niveau hiérarchique 1	→	{							
				Entrée/Sortie n°15	316				
		Bifurcation n°20	141						
Rond Point n°14		n°14 :	1 597						
niveaux 2 à 5	{	Hors Tache Urbaine	→	Cédez le Passage n°11	n°11 :	17 717			
		Dans la Tache Urbaine	{	niveaux 4 et 5	{	3 branches	Cédez le Passage n°11	n°11 :	17 717
				4 branches		Mixte n°12	n°12 :	2 637	
		niveaux 2 et 3	→	Carrefour urbain n°13	n°13 :	4 894			

*Source : d’après Godinot et Bonnel (2008)*

TABLE 3.3 – Pénalités aux mouvements tournants dans SIMBAD (en secondes).

Type de nœud	Carrefour	S-S				S-P				P-S				P-P			
		d	td	g	dt												
1	Nœud graphique	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	Cédez le passage, priorité à droite	5	10	15	15	10	20	25	25	5	10	15	15	5	10	15	15
12	Mixte (feux stop)	7	17	22	22	7	17	20	22	7	7	15	20	5	5	7	7
13	Carrefours urbains	5	15	20	20	5	10	15	20	5	10	10	20	5	5	10	10
14	Giratoires	0	0	0	0	10	10	10	10	0	5	5	0	0	0	0	0
15	Entrées / Sorties autoroute	0	0	0	0	30	30	30	30	30	30	30	30	0	0	0	0
20	Carrefours particuliers	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

*Source : d’après Godinot et Bonnel (2008).*

de mouvements tournants :

- d : tourne à droite ;
- td : tout droit ;
- g : tourne à gauche ;
- dt : demi-tour.

La définition du mouvement tout droit est effectuée de manière automatique à partir de la géométrie du carrefour. Les autres mouvements sont ensuite définis par rapport au mouvement tout droit. Cette codification ne renseignant que la géométrie du carrefour, elle ne peut pas être modifiée (Godinot et Bonnel, 2008). Les pénalités retenues pour les mouvements aux nœuds sont présentées dans la table 3.3.

**3.3.1.1.3 Faire le lien entre zones et réseau** Pour compléter la codification du réseau, il est nécessaire d’inclure des éléments permettant de faire le lien entre les

infrastructures (définies comme des ensembles d'arcs et de nœuds) et la demande qui va effectivement utiliser ces infrastructures. Ce lien est établi par l'intermédiaire de nœuds et d'arcs particuliers : les centroïdes de zone et les connecteurs de centroïde. Ces deux éléments sont souvent qualifiés de « fictifs » car ils ne correspondent à aucun élément matériel du réseau mais à des conventions (Bonnell, 2004). Cette fiction est rendue nécessaire par le principe même du zonage (représentation discrète d'un espace continu), ce qui confirme l'importance du choix d'un zonage adapté aux objectifs visés. Le centroïde est un nœud qui concentre tous les attributs d'une zone. Autrement dit, toutes les localisations de la zone (ménages, emplois, origines et destinations des déplacements...) sont concentrées en ce point. Les connecteurs de centroïde sont des liens établis entre les centroïdes des zones et certains nœuds du réseau. Ces derniers deviennent alors effectivement les points d'origine et de destination des différents itinéraires sur le réseau codifié.

À la connaissance de l'auteur, la recherche consacrée à la spécification des connecteurs et des centroïdes est peu développée<sup>27</sup>. En général, on peut trouver dans les manuels de modélisation (comme, par exemple, Ortúzar et Willumsen, 2001 ; Bonnell, 2004) ainsi que dans certains documents méthodologiques des administrations publiques (Ismart, 1990 ; Cambridge Systematics, Inc, 2009) quelques indications (heuristiques) sur la manière de définir les connecteurs. On peut les regrouper dans les « règles » suivantes :

- les connecteurs doivent représenter un accès réaliste aux réseaux ;
- ils ne doivent pas traverser des barrières (naturelles, comme des rivières, ou construites par l'homme, comme des autoroutes) ;
- un nombre suffisant de connecteurs doit être prévu pour éviter de charger trop de déplacements sur une seule infrastructure et la capacité des connecteurs eux-mêmes est supposée être infinie ;
- il ne faut pas connecter les centroïdes à des intersections importantes (risque de biaiser les mouvements tournants) ou à des bretelles donnant accès directement aux autoroutes.

En principe, le positionnement du centroïde n'a pas d'impact sur l'affectation des déplacements. Dans la pratique, certaines caractéristiques des connecteurs sont dérivées directement du positionnement du centroïde et du choix des nœuds de réseau à relier. Ainsi, par exemple, le temps d'accès au réseau, attribut du connecteur de centroïde, est souvent déterminé en appliquant une vitesse constante à la longueur du connecteur. Le positionnement du centroïde et le choix des nœuds à connecter doit donc être fait avec attention.

La méthode retenue dans SIMBAD comporte deux étapes. La première consiste à positionner géographiquement le centroïde au centre de la surface bâtie de la zone, qui concentre la très grande majorité des origines et destinations de déplacements. Il est alors possible d'utiliser la distance entre le centroïde et les nœuds de réseau pour définir la longueur des connecteurs et donc le temps d'accès au réseau (en supposant une vitesse de connexion de 30 km/h, qui reflète les conditions de circulation des voies de desserte locale).

---

27. On peut néanmoins citer deux études récentes sur le sujet : Friedrich et Galster (2009) et Qian et Zang (2012). Il y a également d'autres travaux sur un sujet proche : celui du sous-zonage (*subzoning*, voir Horowitz, 2001).

La deuxième étape consiste à générer automatiquement les connecteurs (DAVISUM dispose d'un outil ayant cette fonctionnalité) et à en choisir certains. Le choix des connecteurs est réalisé en suivant la procédure suivante (Godinot et Bonnel, 2008) :

- élimination des types de nœuds ne pouvant supporter l'arrivée d'un connecteur par l'application de filtres selon les types de nœuds et les types de tronçons (par exemple entrée/sortie d'autoroute car on oblige à utiliser ensuite l'autoroute quelle que soit la destination du déplacement. . .) ;
- choix des nœuds les plus proches avec un nombre maximal de nœuds égal à 3 au sein du Grand Lyon et égal à 5 hors Grand Lyon. Dans cette recherche, seuls les nœuds dont la distance au centroïde est inférieure à 0,5 km au sein du Grand Lyon et 1,5 km en dehors sont considérés. Au sein du Grand Lyon, la connexion sur une voie à sens unique est possible car certains IRIS ne sont composés que de voies à sens unique, alors qu'en dehors du Grand Lyon ces liens sont exclus pour augmenter la pertinence des liens retenus (les liens à sens unique sont peu nombreux hors Grand Lyon).

Un contrôle est ensuite réalisé pour s'assurer d'une part que l'ensemble des connecteurs sont utiles et d'autre part que le nombre de connecteurs est suffisant pour chacune des zones. L'utilité des connecteurs est évaluée grâce à une méthode classique qui consiste à affecter en deux tranches égales une matrice dont toutes les cases sont égales à l'unité. Les connecteurs qui ne chargent pas sont alors enlevés. Cette opération permet d'alléger la codification du réseau et donc de réduire le temps de calcul nécessaire lors de l'affectation de la matrice de demande. La vérification d'un nombre suffisant de connecteurs est effectuée grâce à l'analyse de la capacité cumulée des liens connectés. Il est nécessaire que cette capacité cumulée soit supérieure à la demande générée et attirée par la zone de sorte à ne pas générer de la congestion fictive lors du chargement de la demande (Godinot et Bonnel, 2008).

**3.3.1.1.4 Calage du réseau** Le calage du réseau s'effectue en affectant une matrice de demande et en comparant le trafic chargé sur les liens ou sur des coupures avec les données de comptage disponibles.

L'année de référence du projet SIMBAD est 1999. La constitution de la matrice de référence en heure de pointe du soir pour 1999 (HPS99) fait appel aux données de l'EMD de 1995 et à celles de l'enquête cordon de 1990. Elle nécessite certaines hypothèses, dont les principales sont présentées dans ce qui suit (le détail peut être trouvé dans Godinot et Bonnel, 2008) :

- définition de matrices de passage des découpages EMD et cordon vers le découpage en IRIS qui permet d'éclater les matrices de déplacements issues de ces enquêtes vers les matrices IRIS ;
- estimation des flux d'échange et de transit en 1990 et en 2005 en tenant compte du fait que l'étendue des périmètres 1990 et 2005 de l'enquête cordon est bien plus limitée que celle de l'aire urbaine (Nicolas *et al.*, 2009, section 3.3) ;
- estimation de la part de transport de marchandises à partir des données de comptages permettant de séparer flux VL (véhicules légers) et PL (poids-lourds) ;
- estimation des taux de croissance des trafics d'échange et de transit à partir des enquêtes cordon de 1990 et 2005. Cette estimation est rendue plus complexe

par le fait que les découpages et les périmètres des deux enquêtes ne sont pas identiques ;

- estimation des taux de croissance des flux internes au périmètre, à partir de la croissance des trafics sur les postes de comptages, pour lesquels les données sont disponibles sur plusieurs années, et de l'évolution de la mobilité automobile repérée dans les précédentes EMD (l'enquête de 2006 n'était pas encore disponible lors de la réalisation de ce travail).

L'affectation de cette matrice HPS99 est réalisée grâce à une affectation à l'équilibre selon le premier principe de Wardrop (Bonnell, 2004) afin de prendre en compte les effets de la congestion sur les choix d'itinéraire. Cette affectation prend comme variable de référence un temps généralisé de déplacement qui est relativement simple. Il n'intègre, en effet que le temps de déplacement ainsi qu'une pénalité temporelle correspondant à un équivalent du péage pour l'utilisation de certains tronçons autoroutiers (Godinot et Bonnell, 2008). Les coûts d'utilisation du véhicule ne sont pas intégrés<sup>28</sup>.

Afin d'analyser les résultats de l'affectation, nous utilisons l'indice *GEH*, qui permet de comparer deux séries de volume de trafic :

$$GEH = \sqrt{\frac{2(M - C)^2}{(M + C)}}$$

où *M* représente les flux modélisés et *C* les flux des données de comptages.

Les résultats de l'affectation sont comparés avec tous les comptages disponibles. Les résultats sont présentés dans les tables 3.4 et 3.5. Généralement, dans les études de circulation de projets d'infrastructure de transport, on considère qu'il est souhaitable que 85 % du volume de trafic ait un GEH inférieur à 5 %. Or, dans la mesure où l'objectif visé dans SIMAD n'est pas une analyse de projets spécifiques mais l'analyse macro-économique de politiques de transport, une précision moindre a été retenue.

#### 3.3.1.2 Codification de l'historique du réseau routier

Une fois que le réseau de référence a été établi et calibré, on peut procéder à la codification des réseaux pour les dates des enquêtes ménages déplacements et à la production de matrices de temps généralisés. La définition d'une typologie pour les différents objets du réseau routier sert de support à la codification de l'historique. La production des matrices de temps généralisés requerrait la définition de matrices de demande pour 1985, 1995 et 2006.

Dans un premier temps, nous donnons les lignes générales de la méthode de codification de l'historique. Puis, nous présentons la construction des matrices de demande.

##### 3.3.1.2.1 Principes de codification de l'historique des réseaux routiers

La méthode de codification de l'historique des réseaux devait s'adapter à la fois aux données disponibles et aux capacités de DAVISUM compte tenu de la licence disponible.

---

<sup>28</sup>. Notons que l'estimation des coûts d'usage des véhicules pour les années d'enquête n'aurait pu qu'être relativement fruste, compte tenu des données disponibles.

TABLE 3.4 – Charges finales en entrée et en sortie de l'AUL

Nom de la ligne écran	Sens Entrant			Sens sortant		
	Charge	Comptage	GEH	Charge	Comptage	GEH
Axe VIENNE Nord – Secteur extérieur 1	3 328	3 271	1,0	3 929	3 292	10,6
Axe ILE D'ABEAU – Est – Secteur extérieur 2 et 10	2 986	2 992	0,1	3 044	2 992	0,9
Axe MEXIMIEUX – Ouest – Secteur extérieur 3et4	1 618	1 355	6,8	1 348	1 355	0,2
Axe ANSE – Nord – Secteur extérieur 6	3 515	3 777	4,3	3 538	3 799	4,3
Axe ROANNE – Secteur extérieur 7	755	660	3,6	669	654	0,6
Axe St ETIENNE – Secteur extérieur 9et8	2 915	2 914	0,0	2 764	2 919	4,1
Axe MORESTEL – Secteur extérieur 10	156	193	2,8	175	193	1,3
Axe DOMBES – Secteur extérieur 5	519	422	4,5	531	422	5,0
Axe AUVERGNE – Secteur extérieur 8	128	133	0,4	232	133	7,3

Source : d'après Godinot et Bonnel (2008).

TABLE 3.5 – Répartition de la valeur de l'indicateur GEH en fonction du type de réseau (en %).

Type	GEH ∈ [0; 10]	GEH ∈ ]10; ∞[	Poids des postes de comptage en UVP
Urbain	57	43	31
Périurbain	75	25	9
Echangeurs	55	47	8
Autoroute, Périphérique, VRU	82	18	52
Tout poste de comptage	66	34	100

Source : d'après Godinot et Bonnel (2008).

**Les données disponibles** La tâche la plus difficile a sans doute été la recherche de données. Nous avons, dans un premier temps, cherché des données similaires à celles utilisées pour le réseau de référence. Nous nous sommes donc procurés la base Navteq© 2006 dans l'idée d'importer dans celle-ci les différents attributs des tronçons étant demeurés identiques entre 1999 et 2006, et de donner automatiquement des caractéristiques aux tronçons « nouveaux » (ceux correspondant aux infrastructures construites entre 1999 et 2006). Malheureusement, les différences entre les deux bases étaient trop importantes. Les objets de la base de 2006 n'étaient pas codifiés de manière compatible avec la base de 1999 (les identifiants des objets changeaient entre les deux bases) ce qui compliquait énormément un travail d'importation des attributs des objets. Cette base a donc été utilisée comme une simple carte où l'on repérait manuellement les changements du réseau structurant.

Si le repérage manuel des évolutions du réseau structurant était la seule solution viable pour la codification du réseau de 2006, elle ne pouvait que s'imposer également

pour 1995 et 1985 du fait de l'absence des bases Navteq© à ces horizons temporels. Nos différentes prises de contact avec les divers organismes (Grand Lyon, DDE, Mairies...) susceptibles de disposer des données de réseaux sur les trois périodes ont été plutôt infructueuses : aucun organisme ne dispose d'un mécanisme de mise en archive des différentes modifications du réseau à l'échelle de l'aire urbaine. De plus, les données sous un format SIG quelconque n'existent que pour des états du réseau très récents (années 2000). Les données pouvant être utilisées pour 1985 et 1995 ne sont disponibles que sur des supports papier.

Nos principales sources d'information ont donc été des cartes routières (de type carte Michelin) — recueillies principalement au fonds Rhône-Alpes de la Bibliothèque Municipale de Lyon, à la cartothèque de l'Université Lumière Lyon 2, mais également auprès de particuliers —, des informations ponctuelles qui nous ont été transmises par certaines personnes rattachées aux administrations locales ainsi que deux sites internet : <http://portail.sara-infras.com> et <http://routes.wikia.com/>, contenant des informations détaillées sur les dates de mise en service.

**Les principes de codification** La codification de l'historique des réseaux nécessite de se poser, au préalable, la question suivante, relative aux comportements des automobilistes : est-ce que les comportements de conduite sont les mêmes aujourd'hui qu'il y a 20 ou 30 ans ? Plus précisément, est-ce que les distances entre voitures, les vitesses, et plus généralement tous les comportements associés à l'écoulement de trafic pour un type de voie sont restés inchangés ?

En réalité, nous avons associé un certain nombre de caractéristiques à chaque type d'élément du réseau, et les flux qui passent par chaque élément du réseau auront des propriétés héritées des caractéristiques associées à cet élément. Si on suppose que les comportements de conduite ont changé, alors il faudrait modifier, pour chaque année, les caractéristiques associées à chaque type d'objet (notamment les paramètres des courbes débit-temps) pour reproduire le changement de comportements.

Dans la pratique courante de la modélisation, notamment à l'échelle d'une agglomération, il est d'usage de faire l'hypothèse que les comportements de conduite sont constants dans le temps. On rajoute ainsi l'infrastructure que l'on souhaite évaluer en lui donnant les caractéristiques définies pour des objets similaires à l'année de référence.

Notre travail cherche à tester l'hypothèse de stabilité du modèle à quatre étapes. Le cas des réseaux est cependant problématique en ce que nous ne disposons pas des données nécessaires pour faire les tests de stabilité (on aurait pu penser à des enquêtes auprès des conducteurs pour étudier l'évolution de leurs attitudes de conduite, mais ces données n'existent pas). Nous devons donc nous limiter à supposer que les comportements de conduite sont relativement stables sur la période étudiée, en donnant quelques arguments en faveur de cette hypothèse :

- l'impact limité de l'évolution technologique sur les vitesses de circulation : on peut rouler à des vitesses similaires avec une voiture d'il y a vingt ans et une voiture actuelle (les vitesses autorisées restent très inférieures aux performances maximales des véhicules) ;
- si les distances de freinage ont pu être réduites, le non respect généralisé des distances de sécurité laisse penser que la distance de freinage de la voiture n'a

- pas un impact fort sur la façon de conduire ;
- la mise en place des normes en matière de distances de sécurité aurait pu modifier les comportements de conduite, mais les autorités ne disposent pas encore de moyen de contrôle (et donc de dissuasion) fiable des distances entre véhicules ;
- la conduite écologique est une pratique en développement mais pas encore généralisée.

En revanche, pour la prospective, on peut s'interroger sur l'impact de la mise en place des radars automatiques dans la mesure où ces derniers ont contribué à un plus grand respect des vitesses maximales autorisées. Celui-ci doit cependant être tempéré par le fait que, pour un travail réalisé à l'heure de pointe, les vitesses maximales sont le plus souvent contraintes par la congestion.

DAVISUM a la capacité de modifier la valeur d'un attribut d'un objet en le remplaçant par la valeur d'un autre attribut du même objet. Cette propriété du logiciel a été déterminante dans la codification de l'historique du réseau. Elle nous a permis de créer des attributs utilisateurs (attributs « optionnels » des objets de réseau, définis par l'utilisateur) correspondant à une année de modélisation et à y stocker le numéro type de l'objet à l'année correspondante pour « modéliser tous les réseaux à l'intérieur d'un seul réseau » DAVISUM.

En réalité le système fonctionne par activation des objets correspondant à l'année sur laquelle on souhaite travailler et, nécessairement, désactivation des autres objets. Un exemple simple permet d'illustrer notre propos : imaginons un tronçon A qui a été mis en service entre 1995 et 2006. Il n'existait donc pas en 1995 et il existait en 2006. Parmi les attributs du tronçon A, intéressons nous aux 3 suivants : « numéro type », « caractéristiques 1995 » et « caractéristiques 2006 ». Le premier est un attribut de base de DAVISUM et les deux autres sont des attributs utilisateurs. Supposons que le tronçon A est de type 17 en 2006. Nous enregistrons cette information dans l'attribut « caractéristiques 2006 ». Le « numéro type » du tronçon permet de définir certaines caractéristiques par défaut de celui-ci, dont les principales sont la capacité et la vitesse à vide. Un tronçon dont le numéro type est égal à 17 a une capacité égale à 6 000 véhicules par heure et une vitesse à vide de 110 km/h. Un tronçon dont le numéro type est égal à 0 a une capacité et une vitesse à vide nulles. Supposons que nous souhaitons faire une affectation sur le réseau de 1995. Nous devons donc « activer » les tronçons correspondant à l'année 1995 (et, par conséquent, désactiver les autres). Pour ce faire, nous faisons une sélection multiple de tous les tronçons et nous donnons à leur attribut « numéro type » la valeur de l'attribut « caractéristiques 1995 ». Pour le tronçon A, cette valeur serait 0. Ainsi, le tronçon A aurait une capacité et une vitesse à vide nulles et il serait exclu lors de la procédure d'affectation.

Cette méthode a l'avantage de ne pas avoir à multiplier les réseaux et les fichiers .ver (fichiers de travail de DAVISUM). En revanche, elle peut poser certains problèmes de connexion, qui font qu'elle fonctionne mieux pour ajouter des tronçons au réseau que pour en enlever. Ainsi, par exemple, comme les zones représentant le trafic extérieur à l'aire urbaine sont prioritairement connectées aux voies de grand gabarit, la suppression de celles-ci (par exemple l'A42 qui n'était pas terminée en 1985) conduit à des erreurs de calcul. Certains connecteurs ont du être rajoutés à ceux du réseau de référence pour tenir compte de ces changements (rappelons que pour les zones ne représentant pas le trafic extérieur la connexion ne se fait jamais directement sur

une autoroute). Ces connecteurs sont également activés ou désactivés en fonction de l'année sur laquelle on souhaite faire l'affectation.

**Matrices de demande** Le calage du réseau routier a été réalisé en utilisant une matrice de la demande construite à partir de l'enquête ménages déplacements de 1995 et des enquêtes cordon de 2005 et de 1990 (Godinot et Bonnel, 2008 ; Nicolas *et al.*, 2009).

Pour construire les matrices de demande pour 1985 et 1995 et 2006, nous utilisons les données suivantes, en ajustant la méthode proposée dans Godinot et Bonnel (2008) en fonction des données disponibles :

- pour 2006, nous utilisons l'enquête ménages déplacements de 2006, qui couvre le périmètre de l'aire urbaine (à l'exception des communes de Dargoire et Tararas, du département de la Loire). Cette enquête permet donc de reconstituer l'ensemble des flux internes au périmètre. Pour les flux d'échange et de transit, nous utilisons la matrice de calage de 1999 à laquelle nous appliquons les coefficients de proportionnalité issus des taux de croissance de trafic observés dans les données de comptage : 2 % par an pour les VP et 1,3 % pour les PL.
- pour 1995, nous utilisons une méthode similaire : les données de l'enquête ménages déplacements de 1995 permettent de définir les déplacements internes au périmètre de celle-ci. Les déplacements d'échange et de transit sont définis à partir de la matrice de 2006 à laquelle on applique les coefficients mentionnés dans le point précédent.
- pour 1985, la méthode est la même que pour 1995, en utilisant les données de l'enquête ménages 1985. Les coefficients de proportionnalité sont un peu différents. Les données de comptages ont permis de définir un taux de croissance de trafic de 2 % par an pour les VP et 1,3 % pour les PL entre 1990 et 2006. Entre 1979 et 1990, la croissance était supérieure : de l'ordre de 5 % pour les VP et 3,2 % pour les PL.

Toutes les matrices utilisées ont du être converties de leurs zonages originaux à un zonage en IRIS. En particulier, pour les enquêtes ménages déplacements, nous avons utilisé un SIG pour définir les surfaces bâties des intersections entre les secteurs de tirage de celles-ci et les IRIS. Celles-ci ont permis d'établir des coefficients de passage entre les deux zonages. Il faut noter que les secteurs de tirage des trois enquêtes sont très proches mais pas totalement identiques. Pour 1985, nous avons dû numériser nous-mêmes les secteurs de tirage de l'enquête, cette information n'étant pas disponible par ailleurs.

Les matrices construites étaient des matrices à la journée et on a utilisé le même coefficient d'ajout du trafic de marchandises et de passage à l'heure de pointe que Godinot et Bonnel (2008) : 12,84. Les matrices de demande ont les caractéristiques données dans la table 3.6.

TABLE 3.6 – Matrices de demande VP

Indicateur	1985	1995	2006
Matrices complètes sur l'AUL (trafic d'échange et transit inclus)			
Nombre de déplacements	158 328	215 835	228 606
Nombre de déplacements moyen par O-D	0,26	0,35	0,37
Périmètre AUL (trafic d'échange et transit exclus)			
Nombre de déplacements	139 955	190 266	197 039
Part des intrazonnes (en %)	4,38	4,43	5,24
Part des O-D ayant 0 déplacement (en %)	18,69	17,85	17,19
Périmètre des EMD évolutif. (trafic d'échange et transit exclus)			
Nombre de déplacements	100 413	144 156	197 039
Part des intrazonnes (en %)	3,43	3,70	5,24
Part des O-D ayant 0 déplacement (en %)	8,36	6,74	17,19

*Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006 et de la matrice HPS99 de Godinot et Bonnel (2008).*

### 3.3.2 Codification du réseau de transports collectifs : vers une codification automatique à partir des données publiques ?

#### 3.3.2.1 Réseau de référence : principes de codification retenus pour le projet SIMBAD

Les principes retenus sont similaires à ceux que nous avons exposés pour le réseau routier. Il convient cependant de noter que la description d'un réseau de transports collectifs est de nature différente de celle du réseau viaire puisqu'il faut prendre en compte une organisation par ligne, avec des itinéraires, des points d'arrêt, des horaires de passage, des fréquences. . . Ces éléments n'ont pas d'équivalent pour la voiture. Il est donc nécessaire de superposer au graphe du réseau support une description des lignes de transports collectifs.

Les transports en commun urbains retenus pour la codification ne couvrent qu'une partie de l'aire urbaine lyonnaise, correspondant au Périmètre des Transports Urbains (PTU) de Lyon<sup>29</sup>. Seulement le réseau urbain du Grand Lyon (dépendant du SYTRAL) et le réseau des cars départementaux (dans la limite du PTU) ont été codifiés (contrainte en matière de données). La codification du réseau ferroviaire est en projet.

Le zonage retenu est logiquement le même que celui qui a été retenu pour la codification du réseau routier. Le découpage le plus fin possible, compte tenu des données démographiques disponibles, a été utilisé.

Dans le prolongement de cette logique, le réseau de transports collectifs a été codifié de manière fine. Tous les arrêts de transports collectifs ont été codifiés y compris lorsque plusieurs arrêts d'une même ligne sont localisés au sein d'une même

29. Le PTU correspond aux 55 communes du Grand Lyon (58 aujourd'hui). D'autres communes ont des accords particuliers qui permettent leur desserte. Il s'agit des communes de Brindas, Chaselay, Grézieu-la-Varenne, Genas, Pollionnay, Sainte-Consorce et Vaugneray (toutes ces communes participent par le biais du versement transport au financement du SYTRAL).

zone. Toutefois, des contraintes similaires à celles décrites pour le réseau routier ne permettent pas d'associer cette finesse de codification à une finesse d'analyse, qui irait par exemple jusqu'à l'échelle des arrêts en termes de charge (montée et descente). Par ailleurs, le choix d'une codification la plus automatique possible à partir des bases d'accès aisé ne permet pas un calage fin du réseau. Le réseau est finement codifié sur le plan spatial, mais l'analyse doit rester agrégée. Cela étant, les temps de parcours (et temps généralisé) doivent être utilisables, selon les mêmes critères que ceux des véhicules particuliers.

**3.3.2.1.1 Sources de données** Comme pour le réseau routier, le recours à des bases de données « facilement » accessibles a été privilégié. Le réseau de référence du projet SIMBAD correspond à l'année 2001. Le guide du réseau édité par les TCL ainsi que les plans de cette date contiennent la grande majorité des informations nécessaires à la codification : tracés de lignes, position des arrêts, fréquence des services en heure creuse et en heure de pointe. La seule donnée manquante correspond aux temps de parcours. Ceux-ci ont été récupérés auprès du service méthodes des TCL<sup>30</sup>.

La digitalisation du réseau a été facilitée par la récupération du travail de Caubel (2006), notamment en ce qui concerne la digitalisation des points d'arrêt.

**3.3.2.1.2 Codification des lignes de transports collectifs** La codification des lignes de transports collectifs est assez complexe au sein de DAVISUM. Elle nécessite la définition d'un support du réseau, constitué de nœuds et de tronçons, sur lequel on « superpose » des objets contenant des informations permettant de caractériser les lignes de transport en commun. Ces objets sont :

- le **Système de transport** fait référence à l'offre de transport du point de vue du matériel roulant : métro, funiculaire, tramway et bus pour le réseau lyonnais ;
- les **Arrêts**. DAVISUM utilise une structure hiérarchisée des arrêts :
  - le **Point d'arrêt** rend compte du choix des lieux d'arrêt commercial des lignes TC. Il est rattaché à un nœud. On a retenu une codification où il y a un seul objet point d'arrêt par système de transport pour les lieux d'arrêt des TCL ;
  - la **Zone d'arrêt** correspond plus ou moins à un quai physique. Une zone d'arrêt est assignée à un arrêt précis et peut regrouper plusieurs points d'arrêt. C'est au niveau des zones d'arrêt que DAVISUM comptabilise les montées et les descentes (et non au niveau des points d'arrêt) ;
  - l'**Arrêt** regroupe tous les points et les zones d'arrêt. Il n'est pas directement assigné à un nœud du réseau ou à un tronçon. Sa fonction est de définir les correspondances entre zones d'arrêt. Les zones d'arrêts sont ainsi reliées entre elles par une matrice de transition qui contient des temps de marche à pied de transition pour chaque système de transport (voir plus loin dans la définition du temps généralisé).
- la **Ligne** permet d'introduire les lignes du réseau d'un point de vue que l'on pourrait qualifier d'« administratif ». Les caractéristiques plus techniques sont définies dans les éléments qui suivent :

---

30. Notons que depuis le réaménagement récent des lignes TCL dans le cadre d'Atobus, les temps de parcours sont disponibles dans les fiches usuellement consultables aux agences TCL.

- l'**Itinéraire de ligne** contient le tracé de la ligne défini comme une succession de points d'arrêt. Une ligne peut avoir plusieurs itinéraires de lignes ;
- le **Tracé d'itinéraire de ligne** contient le tracé précis de la ligne défini comme une succession de tronçons. Une ligne peut avoir plusieurs tracés d'itinéraires de lignes ;
- le **Profil de temps de parcours** contient le temps de parcours inter-arrêts, ainsi que le temps d'arrêt commercial à chacun des arrêts pour chacune des lignes. Une ligne peut avoir plusieurs profils de temps de parcours selon les services ;
- le **Service** correspondent à la circulation du matériel roulant à un moment précis de la journée : chaque itinéraire est parcouru autant de fois qu'il y a de services.

La codification de ces objets a été réalisée par étapes. Pour les bus, les navettes et les trams, les nœuds et tronçons du réseau viaire servaient de base de codification. À cette base, il a fallu ajouter des nœuds et tronçons spécifiques lorsque les itinéraires de ligne empruntaient des chemins autorisés seulement aux systèmes de transport caractéristiques de l'offre TC. Dans le cas du métro et des funiculaires, il a fallu codifier des réseaux « parallèles », traduisant le fait que circuler « sous » une avenue est différent de circuler « sur » une avenue (voir les dernières lignes de la table 3.2 pour les types associés aux tronçons, mais aussi aux nœuds, utilisés exclusivement par les transports en commun).

Une fois que la base du réseau TC a été intégralement définie, les autres éléments ont été intégrés à partir des données disponibles. Notons que, pour la définition des temps de parcours, l'horaire de passage aux arrêts intermédiaires (non présents dans les données fournies par le service méthodes des TCL) a été estimé en considérant une vitesse constante entre les arrêts pour lesquels l'horaire de passage est disponible (Skalli, 2006 ; Pittion, 2006). Le réseau de référence codifié est représenté dans l'annexe C.3.

**3.3.2.1.3 Codification automatique des connecteurs de centroïde** Nous avons, dans la préparation des données nécessaires à ce travail, modifié la méthode de génération automatique des connecteurs par rapport à celle qui avait été retenue, dans un premier temps, pour le projet SIMBAD, telle que présentée dans Nicolas *et al.* (2009).

Nous conservons la philosophie présentée pour le réseau routier visant à une production et une codification automatique des connecteurs de centroïde. Le principe d'application est identique, mais sa traduction beaucoup plus complexe.

Le maillage du réseau routier conduit à la présence d'un grand nombre de nœuds de réseau au sein de chacun des IRIS. Dans le cas des transports collectifs, le réseau est restreint : il est bien moins important en termes d'arcs et de nœuds concernés et on peut dire que certains secteurs sont exclus de l'accès aux TC car trop éloignés. Comprendre cet éloignement et le traduire en termes de règles concrètes de création de connecteurs est bien entendu l'enjeu de cette partie.

Si on analyse la desserte des IRIS en TC comme des intersections d'une couche surfacique (IRIS) avec une couche de points (nœuds servant de « support » aux points d'arrêt) on peut constater que :

- il peut y avoir plusieurs points d'arrêt au sein d'un même IRIS. Ces points d'arrêt peuvent ou non appartenir à la même ligne de transports collectifs ;
- il peut ne pas y avoir de point d'arrêt au sein d'un IRIS. Ceci peut être dû tout simplement au fait que l'IRIS n'appartient pas au PTU, soit au périmètre administratif de desserte. Mais l'absence d'arrêts peut être due au caractère arbitraire de la définition des IRIS : on peut par exemple avoir deux IRIS adjacents en centre ville, l'un étant desservi, c'est-à-dire disposant d'une intersection avec un point d'arrêt et l'autre non ;
- il peut y avoir des points d'arrêt qui se trouvent exactement à la limite de deux ou plusieurs IRIS : certaines grandes voies sont privilégiées pour les routes des différentes lignes et servent également de limites aux IRIS.

Ces trois constats rendent peu pertinente l'inclusion des limites des IRIS dans l'analyse. Dans la mesure où leur centroïde a été placé de manière à correspondre au barycentre des surfaces bâties, nous avons opté pour définir une méthode prenant appui seulement sur la distance qui sépare les centroïdes des points d'arrêt. De cette manière, une zone de petite taille, adjacente à une zone similaire mais intersectant un point d'arrêt, pourra être connectée.

L'aire de chalandise des points d'arrêt varie selon le système de transport et l'éloignement du centre. La taille des IRIS varie également selon la distance au centre. Il s'agit dans les deux cas d'une question de densité. La méthode doit inclure ce critère pour déterminer la distance maximale de connexion.

Il faut également prendre en compte le nombre total de connecteurs à générer. Il faut éviter de surcharger le réseau, ce qui pénaliserait les temps de calcul.

Nous avons développé une méthode qui répond en grande partie à ces contraintes. Elle procède par étapes successives :

- (i) Lors de la première étape, nous distinguons les IRIS en deux catégories selon la densité de la population : D (dense) pour les IRIS de plus de 4 600 h/km<sup>2</sup> et ND (non dense) pour les autres. Cette segmentation répond à une volonté de capturer les IRIS de « petite » surface dans la catégorie D. Notons que les IRIS retenus pour représenter la desserte TC sont ceux appartenant au PTU de Lyon, qui est inchangé pour 1985, 1995 et 2006. Le PTU est constitué de 506 IRIS (dont 496 appartiennent au Grand Lyon). 305 IRIS appartiennent à la classe D et 201 à la classe ND (305 et 191, respectivement, pour le Grand Lyon).
- (ii) La seconde étape consiste à utiliser la fonction de génération automatique des connecteurs de DAVISUM pour générer des connecteurs pour chaque système de transport. Cette fonction prend comme argument une distance maximale de connexion que nous avons définie de telle sorte qu'elle permette de traduire au mieux les distances réelles de marche à pied au sein du tissu urbain. Rappelons qu'il est généralement admis dans les études de planification que l'aire de chalandise moyenne des arrêts de transports collectifs est comprise entre 300 et 700 m selon la nature du système de transports collectifs et la densité du réseau. Pour les IRIS considérés comme denses (D) nous avons retenu les distances rectilinéaires maximales de connexion suivantes : 300 m pour les bus et navettes, 450 m pour les trams et 500 m pour les métros et funiculaires. Pour les IRIS non denses (ND), nous avons retenu 500 m pour les bus et navettes, 900 m pour

les trams et 700 m pour les métros et funiculaires. La supériorité de la distance maximale accordée au tram par rapport au métro est due au besoin de raccorder au réseau certains IRIS adjacents à la zone commerciale de la Porte des Alpes à Saint-Priest.

- (iii) La troisième étape consiste à limiter le nombre de connecteurs. Si plusieurs arrêts appartiennent à la même ligne de transports collectifs, il n'est pas indispensable qu'ils soient tous connectés au centroïde car la répartition de la demande entre ces arrêts ne peut être correctement estimée dans une logique de connexion directe des centroïdes et des points d'arrêt<sup>31</sup>. Nous avons donc retenu un seul connecteur par ligne de transports en commun et par centroïde : celui étant le plus court en longueur.
- (iv) La quatrième étape a consisté à attribuer un temps de parcours à chaque connecteur. Ce temps de parcours correspond au temps de marche moyen d'accès au point d'arrêt à partir de la zone à laquelle correspond le connecteur. Nous avons opté pour suivre la méthode conventionnelle de supposer que la vitesse moyenne de marche en ligne droite est de 1 m/s soit encore 3,6 km/h. Comme les trajets en ville suivent le tracé des voies, nous avons appliqué un coefficient multiplicatif de correction à cette vitesse ( $1/\sqrt{2}$ ). Pour obtenir le temps de parcours, nous avons divisé la longueur rectilinéaire du connecteur par la vitesse corrigée.
- (v) La dernière étape consiste à créer une jointure des connecteurs pour toutes les dates afin d'obtenir un seul fichier. On aurait également pu créer plusieurs fichiers, mais cette méthode permet de gérer de manière plus simple la transition d'une année à une autre par simple application de filtres (utilisés également pour les lignes et les points d'arrêt, voir la section 3.3.2.2.1). On obtient au total 2 474 connecteurs TC.

L'application de la méthode ci-dessus a permis de connecter toutes les communes (arrondissements pour Lyon) du PTU à l'exception de Saint-Romain-au-Mont-d'Or (dont le centroïde est trop éloigné de l'unique arrêt TC).

À ce stade de l'analyse, nous n'avons pas pris en compte les trajets multimodaux car leur modélisation est complexe (de l'ordre de l'artisanal dans DAVISUM selon Bhourri, 2002) et qu'ils représentent une part modale encore modeste sur l'agglomération lyonnaise (1 % d'après les données de l'EMD de 2006).

Il convient de noter que le générateur de connecteurs de DAVISUM n'est pas très performant et que, pour une méthode automatisée plus riche, il conviendrait de s'en passer et de faire le travail « manuellement » (dans un autre logiciel comme R, par exemple, R Development Core Team, 2011) et d'importer la table de connecteurs.

**3.3.2.1.4 Choix de la fonction de temps généralisé** Le calcul des temps généralisés nécessite, en plus de la codification des réseaux, de traiter les correspondances et l'estimation des temps d'attente. DAVISUM offre la possibilité de traiter

---

31. Une solution à ce problème consiste à connecter le centroïde à plusieurs nœuds du réseau de voirie — profitant de la corrélation entre nombre d'intersections et densité de population — et de faire un calcul d'affectation prenant en compte des chemins de marche à pied sur la voirie. Nous avons testé cette méthode mais la taille de notre réseau viaire a rendu les calculs impraticables, pour des questions de mémoire vive de l'ordinateur utilisé et du type de licence de DAVISUM.

TABLE 3.7 – Temps spécifiques de correspondances entre systèmes de transports aux arrêts (en minutes).

Systèmes de transport	Métro et Funiculaire	Tramway	Bus et Navette
Métro et Funiculaire	1	2	2
Tramway	2	1	2
Bus et Navette	2	2	1

*Source : traitement Cabrera Delgado.*

très finement les correspondances entre les modes de transports collectifs. Cela passe par la définition de la matrice des temps de marche entre zones d'arrêt, selon les systèmes de transport, en correspondance au niveau des arrêts. Rappelons que quatre systèmes de transport ont été définis : bus et navette, tramway, métro et funiculaire, train. Ce dernier mode n'est toutefois pas activé dans la version actuelle. Tous les arrêts ont pu être traités, en utilisant le logiciel R pour automatiser le plus possible le traitement. Les valeurs retenues sont données dans la table 3.7. Elles ont été modifiées dans le cadre de ce travail par rapport à celles retenues originellement dans le projet SIMBAD, telles que données dans Nicolas *et al.* (2009). Les valeurs retenues résultent du calage du réseau avec les nouveaux connecteurs.

Le mode de calcul du temps dépend du choix de la méthode d'affectation. Deux options sont retenues dans les simulations du projet SIMBAD :

**Affectation à la cadence :** Cette affectation « tient compte du réseau de lignes et utilise les horaires uniquement pour déterminer les temps d'attente moyens ». C'est le mode d'affectation recommandé par PTV pour les réseaux urbains pour lesquels on dispose d'un niveau de fréquence correct pour l'ensemble des lignes. Ce choix est également recommandé pour les études prospectives à moyen/long terme pour lesquels il est difficile de construire des horaires précis en situation de prévision. Dans cette affectation, un temps d'attente moyen, fonction de l'intervalle de passage, est calculé pour la prise de la première ligne et pour chacune des correspondances ;

**Affectation à l'horaire :** À l'inverse de la précédente, cette affectation tient compte des horaires de passage. Cette affectation est recommandée lorsque la fréquence de passage devient faible voire très faible. Dans ce cas, on considère que l'utilisateur connaît les horaires et que les horaires sont respectés. L'utilisateur n'a donc pas d'attente pour la première ligne de transports collectifs. En revanche, en correspondance, l'attente est calculée de manière « exacte » à partir des horaires de passage des lignes en correspondance.

Seulement la deuxième a été retenue dans le présent travail. En effet, l'utilisation de la première dans SIMBAD tient essentiellement à des contraintes techniques en matière de temps de calcul. Comme pour le présent travail il n'est pas nécessaire de faire plusieurs affectations à la suite (ce qui est le cas de SIMBAD) et que les nouvelles versions du logiciel DAVISUM réduisent les temps de calcul (par l'exploitation des processeurs multi-cœurs), la contrainte technique ne s'applique pas vraiment. Par

TABLE 3.8 – Fonction de temps généralisé.

Temps élémentaire	Facteur de pondération
Temps en véhicule	1
Temps rabattement = temps de marche du connecteur à l'origine et à la destination + pénalité modale	1
Temps de correspondance (somme de la pénalité de correspondance et du temps de marche à pied)	1
Temps attente départ (demi intervalle entre deux passages)	2
Temps attente correspondance (demi intervalle entre deux passages)	2
Pénalité supplémentaire pour une rupture de charge	4 min

*Source : traitement Cabrera Delgado.*

ailleurs, la disponibilité des horaires exacts de départ de chaque service ainsi que des horaires de passage à plusieurs points d'arrêt (du fait du caractère rétrospectif de notre étude) rendent l'affectation à l'horaire plus précise dans le calcul des temps de parcours entre les différentes O-D.

La formulation finale de la fonction de temps généralisés permet une pondération des différents temps élémentaires composant un déplacement en transport collectif. En plus des temps de correspondance, déjà présentés, il est possible d'ajouter une pénalité de rupture de charge à chacune des correspondances. Le calage du réseau à conduit à retenir une pénalité de 4 minutes qui traduit une préférence relative pour des trajets directs par rapport à des trajets avec correspondance. Le temps de rabattement, comme on l'a vu, dépend de la longueur des connecteurs. L'estimation de ces pondérations résulte du calage du réseau de transports collectifs (table 3.8).

**3.3.2.1.5 Calage du réseau de transports collectifs** Le calage du réseau de référence dans le projet SIMBAD s'appuie sur les données et les résultats du logiciel TERESE (SEMALY, 2000) développé par EGIS RAIL et appliqué sur l'agglomération lyonnaise. Ces résultats peuvent être considérés comme proches de la « réalité » dans la mesure où EGIS RAIL a réalisé de nombreuses études sur l'agglomération lyonnaise avec le logiciel TERESE dont plusieurs prévisions de trafic qui ont généralement conduit à des résultats très satisfaisants (GESMAD, 2000).

Afin de renforcer la comparabilité des sorties des deux modèles (SIMBAD et TERESE), la matrice des déplacements transports collectifs 2001 d'EGIS RAIL a été utilisée. Plusieurs tests ont été réalisés afin de caler les paramètres de la fonction de temps généralisé, ainsi que les matrices de pénalités de correspondance entre modes et les matrices de temps de marche à pied entre zones d'arrêt. Les résultats, utilisant les coefficients donnés précédemment, sont présentés dans la table 3.9). Ils font apparaître des résultats proches de ceux produits par le logiciel TERESE à la fois sur le plan de la répartition entre les modes de transports collectifs et sur celui de la charge des axes lourds.

### 3.3.2.2 Codification de l'historique du réseau TC

Comme pour le réseau routier, une fois que le réseau de référence a été établi et calibré, on peut procéder à la codification des réseaux pour les dates des enquêtes

TABLE 3.9 – Résultats du calage du réseau de transports collectifs de référence

		EGIS RAIL (TERESE)		SIMBAD (DAVISUM)	
		charge	en %	charge	en %
Nombre de déplacements	Matrice	217 296		217 296	
	Affecté	209 951		217 296	
Répartition modale	Bus	121 651	39,00	137 977	39,80
	Funiculaire	3 150	1,00	2 567	0,80
	Métro	155 369	49,40	167 919	49,00
	Tramway	34 582	11,00	35 738	10,40
	Total	314 752	100,00	342 491	100,00
Charges des principales lignes de transports collectifs	T1	17 750	4,25	15 813	4,62
	T2	16 832	5,35	19 925	5,82
	A	57 737	18,34	59 928	17,50
	B	28 839	9,16	34 704	10,13
	C	8 356	2,65	12 094	3,53
	D	60 437	19,2	61 193	17,87

Source : *TERESE (EGIS RAIL) et traitement LET (SIMBAD)*.

ménages déplacements et à la production de matrices de temps généralisés. La méthode de codification retenue diffère sensiblement de celle retenue pour le réseau routier. Cette différence est liée aux différences entre les caractéristiques des modes de transport, évoquées précédemment (§3.3.1.1.2 et §3.3.2.1.2).

**3.3.2.2.1 Principes de codification de l'historique des réseaux TC** La méthode de codification de l'historique des réseaux devait s'adapter à la fois aux données disponibles et aux capacités de DAVISUM compte tenu de la licence disponible.

**Les données disponibles** La tâche la plus difficile a sans doute été la recherche de données. Dans la mesure où la gestion des transports en commun est relativement centralisée, nous nous attendions à ce que les données relatives à l'historique des réseaux seraient d'un accès relativement direct. Dans les faits, il est apparu que, si l'exploitant ou l'autorité organisatrice du réseau détiennent toutes les informations sur l'état actuel du réseau, ils ne disposent pas d'archives consultables relatives à l'état du réseau dans le passé. Nous avons donc dû nous reporter sur des sources alternatives. Comme pour le réseau routier, nous avons trouvé une bonne partie des informations nécessaires dans le fonds Rhône-Alpes de la Bibliothèque Municipale de Lyon. On a pu y trouver des cartes du réseau, qui contiennent des informations relatives au tracé (points d'arrêts, itinéraire) des lignes, ainsi que des guides du réseau, contenant les jours de service et les fréquences des lignes. Nous avons pu récupérer des données relatives aux temps de parcours des différentes lignes pour 2006. Pour les autres années, nos discussions avec le service méthodes des TCL ont permis de faire le constat d'une relative stabilité des vitesses de circulation entre 1985 et 2001.

**Codification dans DAVISUM** Contrairement au réseau routier où on a codifié les réseaux de 1985, 1995 et 2006 « au sein d'un même réseau », les différentes informations nécessaires à la définition du niveau de service des TC (itinéraires des

lignes, fréquences des services, temps de parcours...) font qu'il faut codifier individuellement chaque ligne pour chaque année, ce qui conduit à une multiplication des lignes codifiées.

Alors que pour le réseau routier nous avons été contraints de nous limiter à la codification des évolutions du réseau structurant, pour le réseau TC nous avons pu définir les évolutions précises de la très grande majorité des lignes et les intégrer dans la codification.

Les sources de données pour la codification des différentes composantes du réseau, dans l'organisation proposée au sein de DAVISUM, ainsi que les besoins de modifications sont donnés ci-dessous :

**Arrêt, Zone d'arrêt, Point d'arrêt :** Comme on l'a vu dans la section 3.3.2.1.4, les lieux de montée et descente et de correspondance sont organisés de manière hiérarchique et les points d'arrêt sont liés avec un nœud. En raison de ce lien, pour la codification des différents points d'arrêt, il a fallu procéder comme pour les tronçons et les nœuds. Nous avons, dans un premier temps, codifié l'ensemble des zones d'arrêt et points d'arrêt pour toutes les années, en évitant les répétitions. Trois attributs utilisateur binaire ont été définis pour signaler l'existence de ces éléments aux trois dates d'analyse. Ensuite, le choix des éléments d'intérêt pour une année particulière se fait par l'intermédiaire de filtres. Les arrêts peuvent ainsi contenir des éléments actifs et inactifs. Nous n'avons pas défini d'attribut particulier pour filtrer les arrêts. En réalité, on peut dire qu'ils ne sont « actifs » qu'à partir du moment où une ligne intégralement définie, et active à la date en question, passe par eux, ce qui se voit reflété sur les zones et les points d'arrêt. Un filtre appliqué sur ces derniers permet donc également d'activer ou désactiver les arrêts. Ainsi, par exemple, l'arrêt de métro « Stade de Gerland » est défini, mais il n'est actif qu'en 2006, où l'arrivée de la ligne B le « connecte » au reste du réseau TC. L'inconvénient de cette méthode de codification est que certains arrêts, zones d'arrêt et points d'arrêt ont changé de nom, ce qui n'est pas reflété dans notre réseau. Les données utilisées pour codifier ces éléments sont essentiellement les cartes du réseau.

**Ligne, Itinéraire de ligne, tracé de l'itinéraire :** Une fois que l'ensemble des arrêts nécessaires était en place, compte tenu des capacités de sélection et de filtrage des différents objets au sein de DAVISUM, il est apparu qu'il fallait codifier séparément chacun de ces éléments pour chacune des années sur lesquelles on travaille. Ainsi, le réseau TC est formé de plus de 450 lignes pour les quatre années de codification (en prenant en compte le réseau pour l'année de référence de SIMBAD). La définition du réseau d'une année se fait par filtrage des lignes par nom, le nom des lignes contenant des informations sur l'année. Ainsi, la ligne de bus 1, en 2006 est nommée « 2006\_Bus1 », ce qui permet d'utiliser les filtres DAVISUM pour « désactiver » les lignes (et donc les arrêts) qui ne sont pas pertinents pour l'année de simulation. Les sources principales de données pour codifier les lignes et leurs itinéraires sont les plans du réseau TCL.

**Profil de temps de parcours, éléments de profil, service :** Une fois que les lignes ont été codifiées, il était nécessaire de définir les temps de parcours et les services des différents itinéraires. DAVISUM donne des temps par défaut en

TABLE 3.10 – Nombre de lignes codifiées par année

Système de Transport	1985	1995	2001	2006	Total
Métro	3	4	4	4	15
Tram	0	0	2	2	4
Funiculaire	2	2	2	2	8
Bus et Navettes	77	105	151	127	460
Total	82	111	159	135	487

*Source : traitement Cabrera Delgado.*

fonction d'une vitesse commerciale (par défaut 50 km/h) et de la longueur des tronçons servant de support aux itinéraires. Or ces temps sont différents des temps d'exploitation constatés. La définition des profils de temps de parcours se fait par étapes en fonction des données disponibles pour chaque année :

- pour 2006, comme c'était le cas pour le réseau de référence de SIMBAD, nous disposons des données de temps de parcours inter-arrêts pour la totalité des lignes modélisées. Ces données ne portent pas sur tous les points d'arrêt d'une ligne mais seulement sur quelques points importants. En faisant l'hypothèse que les vitesses de circulation sont constantes entre deux points d'arrêt pour lesquels on dispose du temps de parcours inter-arrêts, il a été possible de déterminer un temps inter-arrêt pour l'ensemble des arrêts (Noureddine, 2011).
- pour déterminer les temps de parcours en 1985 et 1995 nous avons fait l'hypothèse de stabilité des vitesses de circulation entre 1985 et 2001 ce qui a permis de définir les profils de temps de parcours de la manière suivante :
  - pour les lignes dont l'itinéraire n'a pas du tout changé sur la période considérée, nous avons utilisé les mêmes profils de temps de parcours qu'en 2001 ;
  - certaines lignes ont subi des reconfigurations importantes. Il était nécessaire d'établir des vitesses de circulation moyennes au niveau des tronçons suivis dans leurs itinéraires :
    - si les tronçons empruntés au cours des itinéraires de 1985 ou 1995 étaient également empruntés par des lignes en 2001, alors il est possible de calculer les temps de parcours sur ces tronçons comme étant la moyenne des temps de parcours de l'ensemble des lignes les empruntant en 2001, en prenant soin de distinguer les itinéraires normaux et express ;
    - si les tronçons n'étaient pas empruntés par des lignes en 2001 (par exemple pour les bus qui ont été remplacés par des trams) alors il a été nécessaire de définir des temps de parcours à partir des vitesses de circulation sur des tronçons similaires. Nous avons retenu dans un premier temps la vitesse moyenne de circulation dans les tronçons du même type. Les types 19, 25 et 53 n'étaient pas empruntés par des itinéraires de ligne en 2001. Nous leur avons attribué une vitesse moyenne de circulation à partir de la continuité des axes empruntés. On leur a donc attribué les vitesses moyennes des types 33, 46 et 83 respectivement.

La table 3.10 contient les informations sur le nombre de lignes codifiées par système de transport. On remarque que l'évolution majeure entre 1985 et 1995 est la mise en service de la 4<sup>e</sup> ligne de métro et que l'évolution majeure entre 1995 et 2006 est la mise en service des deux lignes de tram. Il convient de signaler que lors de cette période il y a également eu des prolongements des lignes de métro B et D (voir les figures C.1, C.2, C.3 et C.4 de l'annexe C).

**3.3.2.2 Détermination des temps généralisés** De manière similaire à ce que nous avons du faire pour la modélisation de l'historique du réseau VP, nous devons faire un certain nombre d'hypothèses concernant les comportements des personnes de l'agglomération lyonnaise en matière de choix de l'itinéraire en transport en commun. En effet, nous ne disposons pas de données fines sur la période étudiée qui permettraient de dégager des différences dans l'agencement des systèmes de transport privilégiés lors d'un déplacement. Nous sommes donc obligés de faire l'hypothèse que les coefficients de la fonction de temps généralisé obtenus lors du calage du réseau de référence restent inchangés dans le temps).

La détermination des temps pour les années 1985, 1995 et 2006 se fait en affectant une matrice unité. Ceci est possible par le type d'affectation choisi qui, dans le cas de l'affectation à l'horaire, est une forme de plus court chemin. Notons que ceci n'est possible que parce que notre analyse nécessite des temps de parcours entre les différentes O-D, et non des charges des différentes lignes.

## 3.4 Synthèse et remarques conclusives

Ce chapitre a permis de décrire la procédure de construction de la « donne » nécessaire à notre évaluation de l'impact que peut avoir l'hypothèse de stabilité des paramètres du modèle à quatre étapes en matière de prévision. Le mot d'ordre était — il ne pouvait être autrement dans une étude diachronique — la comparabilité des données dans le temps, sans laquelle les variations des valeurs des paramètres des modèles étudiés ne reflètent pas seulement des changements dans les comportements de mobilité.

En ce qui concerne les données que l'on associe généralement à l'aspect « demande » des modèles de prévision, la source que nous avons retenue est naturellement celle des enquêtes ménages déplacements, dont l'objectif est en partie d'alimenter les types de modèles que nous étudions dans le présent travail. Nous avons analysé les conditions permettant leur utilisation dans le cadre d'une étude diachronique. Le respect de la méthode « standard Certu » dans la réalisation des EMD permet, en principe, d'assurer une comparabilité des différentes enquêtes dans le temps. Dans la pratique, leur utilisation à des fins de comparaison temporelle repose sur le fait de pouvoir considérer que certains biais intrinsèques à la méthode d'enquête sont négligeables. Parmi les biais potentiels, le plus important semble être celui de la non-réponse, qui a connu une augmentation importante au cours du temps. Par ailleurs, même si la méthode de collecte de données n'a pas changé, les données collectées peuvent différer d'une année à une autre. Ainsi, par exemple, une variable socioéconomique importante comme le revenu n'est pas disponible pour toutes les dates d'enquête. L'étude des déplacements des étudiants (études secondaires et supérieures) se voit

limitée par la prise en compte ou non des résidences universitaires dans la base de tirage. Elle trouve également des limites liées à des inconsistances de nomenclature pour certaines variables relatives à l'occupation des personnes enquêtées ou à leur catégorie socioprofessionnelle. En termes d'évolutions temporelles des caractéristiques des habitants de l'agglomération lyonnaise, l'analyse des EMD a permis de mettre en évidence une similarité avec d'autres villes des pays développés. On retrouve le vieillissement de la population, la généralisation du travail féminin, la baisse de la taille des ménages et l'étalement de la ville concomitant avec un accès plus facile à l'automobile.

Les données relatives à l'offre de transports ont présenté un défi bien plus important pour ce travail. Dans le modèle à quatre étapes, les niveaux de service des réseaux sont souvent introduits sous la forme d'un temps généralisé, constitué d'une somme d'équivalents temporels des différentes caractéristiques de la qualité de service (dont les plus importantes sont les temps de parcours). Le temps généralisé est obtenu à partir d'une codification des réseaux de transport, prenant appui sur la théorie des graphes. Cette opération est extrêmement coûteuse et chronophage, ce qui fait que ces données sont produites au cas par cas, et sûrement pas de manière standardisée à des intervalles de temps réguliers. En France, les différents organismes chargés de la gestion des transports particuliers et collectifs disposent de réseaux codifiés pour l'état actuel du réseau sans qu'il y ait vraiment une procédure d'archivage définie. Par ailleurs, ces organismes ne coopérant pas forcément, la compatibilité de réseaux produits par des organismes différents semble douteuse. Nous avons donc entrepris la codification des réseaux viaires et de transports collectifs sur l'agglomération lyonnaise, pour les années des EMD. Pour ce faire, nous avons défini une méthodologie assurant la consistance des temps produits et étant reproductible. Celle-ci a permis d'obtenir des temps de parcours acceptables entre les différentes origines et destinations de l'aire étudiée, même si, pour les réseaux viaires, la charge des petites voiries urbaines n'est pas toujours bien reproduite. Une des limites de la méthode a été de devoir supposer que les comportements de conduite, et donc les courbes débit-temps, sont constants. Si cette hypothèse semble plausible en rétrospective, elle l'est peut-être moins en matière prospective, compte tenu du développement des moyens techniques de contrôle du respect des vitesses et, plus généralement, des règles de conduite.

Finalement, il convient de signaler que le travail de codification de l'historique des réseaux réalisé a un intérêt dépassant celui de l'étude de la capacité prédictive des modèles transport traditionnels. Il constitue une donnée importante pour l'évaluation des politiques de transport ayant eu lieu dans l'agglomération lyonnaise, dans une perspective d'accessibilité. Ainsi, une première série de travaux, s'intégrant dans l'analyse des effets de ségrégation spatiale des politiques de transport, est en cours (Bouzouina *et al.*, 2011, 2012). Au-delà de l'évaluation *ex-post*, l'intérêt de disposer d'une méthodologie reproductible dans le temps permet de poursuivre le travail de codification à des fins prospectives. Ceci est d'autant plus intéressant que, comme on l'a déjà remarqué, les disponibilités de réseaux codifiés en France, surtout dans une optique multimodale, sont limitées par la complexité des relations entre les différents organismes chargés de la gestion des infrastructures de transport.



---

# Génération

Ce chapitre est consacré à l'étude de la stabilité temporelle des modèles de génération usuellement utilisés. Nous avons vu dans le chapitre 1 qu'il existe une multiplicité d'approches pour quantifier la mobilité quotidienne des habitants d'une aire urbaine. Dans ce qui suit, nous allons nous servir de cette multiplicité d'approches pour analyser l'évolution du nombre de déplacements et essayer de comprendre l'évolution des comportements sous-jacente. La première section présente la méthodologie suivie, notamment le choix d'unité d'analyse de la mobilité et les indicateurs utilisés pour tester la stabilité des coefficients ainsi que le pouvoir prédictif des modèles. Les trois sections suivantes sont construites de manière analogue : chacune est consacrée à un type de modèle de génération pour lequel elle étudie la qualité de reproduction des observations aux dates de calage, évalue la stabilité des coefficients dans le temps et examine le pouvoir prédictif du modèle à 10 et 20 ans. Ainsi, la deuxième section est consacrée à l'étude des modèles les plus utilisés aujourd'hui : les modèles de classification croisée au niveau des ménages. Le choix de ce type de modélisation pour commencer le chapitre n'est pas anodin. En effet, une analyse des évolutions temporelles ne peut se faire sans des segmentations de la population de même nature que celles que l'on trouve dans les modèles de génération par classification. En les abordant en premier, on peut observer au passage l'évolution de la mobilité dans l'aire d'étude. La troisième section étudie les modèles agrégés de régression zonale. La quatrième section sera consacrée à l'étude de modèles de régression au niveau des ménages. Une dernière section permettra de regarder une nouvelle approche non paramétrique : les algorithmes de type arbre de régression. En effet, comme on l'a évoqué dans la section 1.2.1.3, ces modèles permettent d'automatiser, à un certain degré, la phase de spécification. Leur inclusion dans ce chapitre se veut une manière d'aller au-delà du simple test de stabilité des paramètres, pour une spécification invariante dans le temps.

## 4.1 Une méthodologie pour étudier les évolutions temporelles

Cette section définit la méthodologie permettant d'évaluer la stabilité des modèles de génération. Un premier point aborde l'unité de mesure de la mobilité retenue. Un

deuxième point décrit les différents types de modèles évalués. Un dernier point décrit les outils d'évaluation des performances descriptives et prédictives des modèles ainsi que de la stabilité des paramètres.

### **4.1.1 Choix de la « sortie » comme unité de mesure de la mobilité**

#### **4.1.1.1 Avantages d'une analyse en termes de nombre de sorties**

Comme nous l'avons vu dans le chapitre 1, l'objectif de l'étape de génération est de quantifier la mobilité de l'aire étudiée. Il est convenable de se doter d'une unité pour mesurer cette mobilité. Contrairement aux usages dans les approches anglo-saxonnes, où l'on privilégie le déplacement pour évaluer la mobilité, nous avons choisi de travailler plutôt en termes de sorties du domicile (Axhausen, 2000 ; Bonnel, 2004), c'est-à-dire la séquence de déplacements en lien avec les différentes activités réalisées par les résidents de l'aire d'étude à partir du moment où ceux-ci sortent de leur domicile jusqu'au moment où ils y reviennent. Le choix de la sortie plutôt que du déplacement comme unité d'analyse repose sur :

- (i) un désir de comparabilité entre les données des différentes enquêtes : la définition du déplacement est restée stable dans la méthode « standard Certu » (CERTU, 2008). En revanche, selon la qualité des enquêteurs et du contrôle, il est possible que l'application de la définition ait pu changer d'une enquête à une autre, notamment pour les arrêts courts comme acheter du pain ou prendre de l'essence qui sont censés être identifiés au cours de l'enquête. Le choix de la sortie comme unité d'analyse permet de minimiser l'impact de ce risque d'inconsistance dans les données ;
- (ii) une symétrisation des flux des matrices O-D par motif à la journée, en cohérence avec les données produites par le modèle de distribution gravitaire (voir point 4.1.1.2.2) ;
- (iii) une meilleure intégration au niveau des étapes de génération et de répartition modale. Au niveau de la génération, la définition en termes de déplacements peut conduire à perdre le motif principal de la sortie du domicile pour les sorties complexes (plus de 2 déplacements), réduisant d'autant la pertinence des lois de génération. Au niveau de la répartition modale, les sorties comportent une forte homogénéité modale, au moins pour les modes motorisés. Dans une approche en termes de sorties, le motif principal justifiant la sortie peut être conservé et la cohérence modale interne à la sortie peut être assurée (Kitamura, 1984 ; Bonnel, 2004).

#### **4.1.1.2 Évolution de la mobilité en termes de « sorties »**

**4.1.1.2.1 La complexité des objets « sorties »** L'analyse en termes de sorties est rapidement complexe en raison de la variabilité du nombre d'activités les constituant et surtout de la très grande diversité des combinaisons d'activités possibles. Il est donc difficile de définir des typologies d'enchaînement d'activités simples. La table 4.1 présente l'évolution des comportements de mobilité des habitants de l'agglomération lyonnaise à partir d'une nomenclature des sorties comprenant 30 classes

## 4. Génération

TABLE 4.1 – Sorties réalisées pour 1 000 personnes suivant une nomenclature semi-détaillée en 30 classes.

N30	Description	EMD 1995			EMD 2006			
		1985	P1985	P1995	P1985	P1995	AUL	P2006
1	DOMI+ACHA/SER+DOMI	245	245	244	235	233	222	219
2	DOMI+TRAV+DOMI	300	238	237	239	239	236	236
3	DOMI+LOISIR+DOMI	195	212	212	235	234	233	232
4	DOMI+ACCT+DOMI	102	121	125	124	125	140	146
5	DOMI+COL/LYC+DOMI	123	113	115	109	109	108	105
6	DOMI+PRIMAIRE+DOMI	86	75	78	72	72	76	76
7	DOMI+AUTRE+DOMI	24	37	37	13	13	14	15
8	DOMI+UNIV+DOMI	22	37	36	27	26	22	19
9	DOMI+ACHA/SER+ACHA/SER+DOMI	23	30	29	38	38	38	41
10	DOMI+TRAV+ACHA/SER+DOMI	22	25	25	21	22	21	21
11	DOMI+TRAV+LOISIR+TRAV+DOMI	19	22	21	20	21	20	19
12	DOMI+LOISIR+LOISIR+DOMI	11	17	16	17	17	15	15
13	DOMI+ACHA/SER+LOISIR+DOMI	10	14	14	14	14	14	15
14	DOMI+ACCT+TRAV+DOMI	11	14	14	12	13	14	14
15	DOMI+ACCT+ACHA/SER+DOMI	10	13	13	12	12	14	14
16	DOMI+TRAV+TRAV+DOMI	8	12	12	10	9	10	10
17	DOMI+TRAV+TRAV+TRAV+DOMI	8	11	11	8	8	9	9
18	DOMI+LOISIR+ACHA/SER+DOMI	9	10	10	11	11	11	11
19	DOMI+ACCT+ACCT+DOMI	4	8	9	9	9	10	12
20	DOMI+TRAV+LOISIR+DOMI	9	10	9	10	10	9	9
21	DOMI+TRAV+ACCT+DOMI	6	8	8	8	8	9	9
22	TRAV+DOMI	6	7	7	6	6	6	6
23	boucle complexe TRAV	85	100	100	86	89	92	96
24	boucle complexe PRIMAIRE	13	20	20	17	18	22	23
25	boucle complexe COL/LYC	20	22	22	18	18	18	19
26	boucle complexe UNIV	14	21	21	13	13	11	10
27	boucle complexe ACHA/SER	35	57	57	51	51	54	59
28	boucle complexe ACCT	14	24	24	22	23	23	24
29	boucle complexe LOISIR	18	26	26	17	17	17	17
30	boucle complexe AUTRE	5	5	6	1	1	2	2
Total		1 457	1 555	1 557	1 475	1 477	1 490	1 504

*Avertissement : nous attirons l'attention du lecteur sur le fait que la table ci-dessus contient des évolutions pour un grand nombre de catégories, ce qui peut avoir un impact au niveau de la significativité statistique de ces évolutions. En réalité, la décomposition en 30 classes permet ici simplement de mettre en évidence la multiplicité des formes de chaînage dans les sorties, dans une optique exploratoire. L'analyse des intervalles de confiance sera réservée, dans ce qui suit, à l'étude des modèles de génération.*

*Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006.*

(Gitton, 2006). Pour assurer la comparabilité, les chiffres donnés pour les différents périmètres — pour les enquêtes de 1995 et 2006 — correspondent aux sorties réalisées par les habitants d'un périmètre, à l'intérieur de ce périmètre (les sorties de transit et d'échange en sont donc exclues). Les sorties y sont caractérisées par un motif au départ, le domicile, puis par une suite de motifs à la destination de chaque déplacement les constituant. On distingue huit motifs de déplacement, autres que le domicile, que l'on peut classer hiérarchiquement, de celui considéré comme étant le plus contraint à celui perçu comme étant le moins contraint (en anglais *non-flexible* ou *flexible*, Kostyniuk et Kitamura, 1984), de la manière suivante : Travail  $\succ$  Enseignement (primaire, secondaire et supérieur)  $\succ$  Achats-Services  $\succ$  Accompagnement

‣ Loisirs ‣ Autres. Le classement hiérarchique des motifs de déplacement permet de définir le motif principal de la sortie, en retenant le plus élevé des motifs présents dans la sortie. Ce principe est notamment appliqué dans la définition des « boucles [sorties] complexes » (catégories 23 à 30). Notons que les sorties simples, consistant en un aller-retour pour l'un des huit motifs retenus (types 1 à 8), représentent 75 % de la mobilité totale pour mille personnes en 1985 et autour de 70 % pour les autres années et périmètres. Si l'on s'intéresse à l'évolution de la mobilité totale pour mille personnes, définie en termes de sorties, on constate une augmentation de celle-ci entre 1985 et 1995 (environ 7 %, à périmètre constant) puis une diminution entre 1995 et 2006 (environ 5 %, à périmètre constant). Lorsqu'on regarde le détail, à partir de la classification en 30 types, on remarque que les évolutions constatés au niveau de la mobilité totale ne sont pas forcément celles que l'on observe pour chaque type de sortie. En effet, seulement les sorties simples pour les motifs autres (7) et enseignement supérieur (8), 5 types de sorties contenant le motif travail (10, 14, 16 17 et 23) et 3 types de sorties complexes ayant pour motifs principaux l'enseignement primaire (24), l'enseignement supérieur (26) et les loisirs (29) connaissent une évolution analogue à celle de la mobilité totale, dont elles ne représentent qu'environ 15 % (19 % en 1995). Pour les autres types de sorties, on observe des évolutions à la hausse ou à la baisse entre 1985 et 2006. Par exemple, les sorties simples pour le motif achats-services (1) connaissent une stabilité entre 1985 et 1995 puis une diminution entre 1995 et 2006. Cette diminution est d'autant plus importante que l'on considère un périmètre d'analyse élargi en 2006 (baisse de 4 % sur le périmètre de l'EMD de 1985 et de 9 % sur celui de l'AUL). À l'opposé, les sorties simples pour les motifs loisirs (3) et accompagnement (4) augmentent entre 1985 et 2006. Pour le motif loisir, l'augmentation est progressive (9 % entre 1985 et 1995 et 11 % entre 1995 et 2006, à périmètre constant) et l'élargissement du périmètre d'analyse en 2006 ne semble pas avoir d'impact. Pour le motif accompagnement, l'augmentation a surtout lieu entre 1985 et 1995 (21 %) mais cette fois l'élargissement du périmètre d'analyse en 2006 se traduit par une forte augmentation (la mobilité dans le périmètre de l'AUL est supérieure à celle du périmètre de l'EMD de 1985 de 13 %). On peut penser que ceci est lié à l'étalement urbain et à la présence plus importante de ménages avec des enfants en périphérie (voir section 3.2.2). Si l'on observe l'évolution des comportements autour du motif travail — qui, on admet souvent, structure les mobilités urbaines — on remarque une très forte diminution des sorties simples domicile-travail-domicile entre 1985 et 1995. Les travailleurs de l'agglomération lyonnaise ne rentrent plus manger chez eux aussi souvent (évolution similaire à celle observée par Kostyniuk et Kitamura, 1984, à partir de données d'enquête dans le Sud-Ouest du Michigan entre 1965 et 1980). On remarque également une augmentation des sorties incluant en même temps le motif travail et le motif accompagnement, ce qui peut traduire l'augmentation du travail féminin et l'accompagnement des enfants à l'école. On avait déjà noté l'augmentation des sorties complexes pour le motif travail entre 1985 et 1995 puis leur baisse entre 1995 et 2006 (sur le périmètre de 1985) pour atteindre le même niveau qu'en 1985. L'élargissement du périmètre d'analyse en 2006 se traduit par une augmentation des sorties complexes pour le motif travail, au point d'approcher les niveaux observés en 1995. Si l'on combine toutes ces évolutions, on observe que le nombre de sorties contenant le motif travail, pour mille personnes, baisse de manière continue entre 1985 et 2006 (environ 6 % tous les 10 ans). L'élargissement du périmètre d'analyse

a un impact limité. En termes de part des sorties contenant le motif travail, celle-ci passe d'environ 33 % des sorties en 1985 à 29 % en 1995 et 28 % en 2006. On a donc, de manière générale, une baisse de la mobilité, en termes de sorties, pour le motif travail. Si on s'intéresse aux motifs autres que travail et enseignement, on observe que le nombre de sorties ne contenant pas ces 4 motifs (l'enseignement compte pour 3 en fonction du niveau), pour mille personnes, augmente entre 1985 et 1995 (environ 16 %, à périmètre constant) puis diminue légèrement entre 1995 et 2006 (baisse d'environ 2 %, à périmètre constant). L'impact du changement du périmètre d'analyse est faible. En termes de part des sorties, celle-ci passe d'environ 48 % des sorties en 1985 à 53 % en 1995 et 54 % en 2006. La part des sorties pour les motifs autres que travail et enseignement s'accroît donc.

**4.1.1.2.2 Simplification de la nomenclature et demi-sorties** La complexité des objets « sorties » pose une difficulté dans la définition d'un modèle de génération : celui-ci doit produire un nombre de sorties pour chaque classe définie dans la typologie. Dans le cas de la typologie utilisée pour le tableau 4.1, cela conduirait à concevoir un modèle permettant de générer 30 classes de sorties. Ceci peut être problématique dans les modèles de génération usuellement utilisés. Ces derniers comportent en général moins d'une dizaine de types à générer, qui sont liés à une décomposition de la mobilité, en termes de déplacements, par motifs. La construction d'un modèle avec un nombre élevé de types de sorties est délicate dans la mesure où le nombre de variables explicatives est limité et, surtout, la taille des échantillons des EMD limite les découpages possibles. Une première solution à ce problème est d'essayer de limiter le nombre de types de sorties. Une analyse préalable des sorties complexes, à partir des données de l'enquête ménages déplacements de 2006, montre que la distance supplémentaire générée par les activités autres que celle correspondant au motif principal de la sortie est très faible par rapport à la distance aller-retour entre le domicile et le lieu du motif principal de la sortie. Il est donc possible, si on ne retient que le motif principal de la sortie comme critère de segmentation, de réduire le nombre de catégories à 8 (une par motif de déplacement retenu). Cette solution a l'avantage de produire des résultats directement utilisables par l'étape de distribution. Tout se passe comme si on travaillait exclusivement avec des déplacements basés au domicile. On distingue deux demi-sorties : l'une ayant pour origine le domicile et pour destination le lieu de réalisation de l'activité principale et l'autre ayant pour origine le lieu de réalisation de l'activité principale et pour destination le domicile. En procédant ainsi, on introduit également une part de stabilité dans les modèles de prévision<sup>1</sup> dans la mesure où, comme le notent Kostyniuk et Kitamura (1984), les probabilités dans le choix de chaînage des différentes activités peuvent ne pas être stables dans le temps. De ce fait, ces changements de probabilités (et donc de quantités de déplacements) ne sont pas pris en compte.

La deuxième solution<sup>2</sup> est de garder un certain degré de complexité dans la dé-

1. Dans les tests de stabilité et de capacité prédictive dans ce chapitre ainsi que dans les chapitres suivants, cela conduit à supposer que l'on se place dans les meilleures conditions possibles pour que les modèles fonctionnent en prévision. Les erreurs observées à ce niveau doivent donc être considérées avec plus d'attention.

2. Celle-ci a été retenue, par exemple, dans le cadre du projet SIMBAD. Une classification des individus a été réalisée pour regrouper ceux ayant des profils de mobilité similaires. On a défini

finition des catégories de sorties et d'essayer d'optimiser l'utilisation des variables explicatives. Celle-ci peut être appliquée avec plus ou moins de succès avec les méthodes algorithmiques « supervisées » — où les classes de la variable à expliquer sont fixées d'avance par l'analyste (comme dans les arbres de segmentation) — ou « non-supervisées » — où les classes de la variables à expliquer sont déterminées de manière endogène par le modèle (comme dans les approches de type réseaux de neurones).

Notons finalement que, dans ce qui suit, nous ferons référence plutôt à des demi-sorties qu'à des sorties. Cette distinction permet de conserver les notions d'origine et de destination des déplacements sous la forme d'origine et de destination principale de la demi-sortie. Ceci permet de faciliter le passage aux matrices O-D nécessaires dans les phases de distribution (chapitre 5) et de répartition modale (chapitre 6).

## 4.1.2 Modèles de génération

### 4.1.2.1 Choix du ménage comme unité de prise de décision

Il est possible d'analyser les demi-sorties émises comme étant le résultat de décisions prises isolément par les individus ou de manière collective au sein des ménages. Dans le contexte français, on est passé d'études prenant plutôt le ménage comme unité de prise de décision à des études de plus en plus fondées sur une analyse à l'échelle individuelle (Bonnell, 2004). Dans le contexte anglo-saxon, plusieurs études ont mis en avant les avantages du choix de l'individu comme unité d'analyse. Cependant, dans la littérature empirique et dans la pratique, le choix du ménage comme unité d'analyse, au niveau de la génération, demeure fréquent (Ortúzar et Willumsen, 2001).

Parmi les avantages généralement attribués à la modélisation de la génération à l'échelle de l'individu, on peut citer le fait que, au fond, ce sont les individus qui se déplacent, et non les ménages ou les zones (Domencich et McFadden, 1975). Une analyse à l'échelle individuelle pourrait se faire à partir d'enquêtes personnes (et non plus ménages) ce qui permettrait de réduire la taille des échantillons (en termes d'individus à enquêter) pour obtenir des niveaux de précision statistique similaires (Supernak *et al.*, 1983). Cela permettrait également d'envisager des enquêtes téléphoniques ou web qui sont souvent réalisées sur une base individuelle. Il est possible d'inclure les caractéristiques du ménage parmi les variables caractérisant l'individu (Pas, 1978). Dans un contexte de prévision, certains changements démographiques sont plus faciles à modéliser à l'échelle de l'individu que du ménage (Bonnell, 2004).

---

33 profils à partir du nombre et des types de sorties effectuées dans une journée. L'algorithme de classification a permis de révéler l'importance du statut de la personne, ainsi que d'autres variables relatives au type de ménage auquel elle appartient, dans la définition des motifs des sorties et donc des profils de mobilité. Cette information a ensuite été utilisée pour construire une segmentation des ménages. Cette dernière a permis de mettre en place une procédure de micro-simulation aléatoire pour traiter la génération. Dans celle-ci, on associe à chaque ménage d'une population, pour laquelle on cherche à faire des prévisions, le comportement de mobilité d'un donneur ayant les mêmes caractéristiques (en fonction des variables retenues pour la segmentation de ces derniers), et pour lequel on connaît la mobilité (données EMD). Cette manière de procéder possède l'avantage de permettre de conserver la variété des comportements. Pour plus de détail, voir Gitton (2006) et Nicolas *et al.* (2009).

Dans le cas de la présente étude les arguments relatifs à la taille de l'échantillon et à la prévision des variables démographiques ne s'appliquent pas vraiment, compte tenu des sources de données et du caractère rétrospectif de la démarche (les données ont déjà été collectées). En revanche, les trois principaux arguments en faveur d'une modélisation à l'échelle du ménage gardent de leur pertinence. Le premier est qu'une partie des déplacements concerne l'entretien et le fonctionnement du ménage et que ces déplacements sont donc gérés de manière collective au niveau du ménage. Le deuxième concerne la répartition des moyens en termes de revenus et de motorisation, qui a également lieu au niveau du ménage (Bonnell, 2004). Le troisième tient au fait qu'il y a dans les modèles à l'échelle des personnes, des problèmes spécifiques en lien avec une proportion plus importante d'unités d'analyse qui ne se déplacent pas (Monzon *et al.*, 1989).

Pas (1978) étudie comparativement le pouvoir explicatif des modèles à l'échelle du ménage et à l'échelle de la personne. Il conclut que le modèle à l'échelle de la personne explique le mieux la variabilité inter-personnelle en termes de nombre de déplacements et que les modèles à l'échelle personnelle et du ménage estiment les émissions à l'échelle zonale avec une performance similaire, légèrement favorable au modèle à l'échelle personnelle. Badoe et Chen (2004b) reviennent sur la question, avec une approche plus générale, incluant un test de capacité prédictive (et non plus seulement explicative). Ils notent que pour la plupart des études réalisées à l'échelle de la personne les données proviennent d'enquêtes ménages, et que ceci peut poser des problèmes par rapport au respect de l'hypothèse d'indépendance des observations usuellement nécessaire aux modèles (notamment ceux de régression). Ils réalisent une évaluation de la capacité prédictive des deux types d'approche et trouvent que l'approche retenant le ménage donne des résultats légèrement meilleurs. L'inclusion de variables relatives au ménage dans l'approche à l'échelle des personnes donne des résultats significativement meilleurs lors du calage, mais cette amélioration ne trouve qu'un impact positif faible en termes d'une performance prédictive accrue. Ils concluent que, dans les agglomérations où l'unité d'échantillonnage est le ménage, il est convenable de garder cette unité pour l'analyse de la génération.

Compte tenu de ce qui a été exposé jusqu'ici, notre analyse privilégiera les modèles à l'échelle du ménage.

### 4.1.2.2 Choix des modèles à étudier

Notre analyse de la transférabilité des modèles de génération portera sur quatre types de modèles :

- (i) Classification croisée au niveau des ménages ;
- (ii) Régression Zonale ;
- (iii) Régression au niveau des ménages ;
- (iv) Arbre de segmentation au niveau des ménages.

Ces modèles représentent une très grande majorité des approches de modélisation usuellement appliquées (notamment les deux premiers).

Nous avons vu dans le chapitre 3 que la comparabilité des données relatives à l'enseignement est douteuse en raison de différences de codage et d'échantillonnage entre

les EMD. En conséquence, notre analyse se focalise sur deux motifs de déplacement : le motif travail et les motifs autres que travail et enseignement.

**4.1.2.2.1 Les modèles par classification croisée** Les modèles les plus utilisés dans les organismes de planification sont certainement les modèles de segmentation ou classification croisée (notamment dans les pays anglo-saxons). Ces modèles consistent, comme on l’a évoqué dans le chapitre 1, à établir des segments de population à partir de quelques variables explicatives (une ou deux en général) et à déterminer le nombre moyen de sorties produites par chaque segment de population.

En nous inspirant de la méthodologie de Walker et Peng (1991), nous regardons l’évolution dans le temps, en prenant en compte les évolutions des périmètres d’enquête, du nombre moyen de demi-sorties pour le motif travail et pour un motif agrégé constitué des motifs autres que le travail et l’enseignement, en fonction de quelques variables usuelles de classification. Deux indicateurs statistiques sont calculés pour chaque cellule de la matrice de classification croisée :

- le nombre moyen de chaînes émises par le segment de population correspondant à la cellule, en prenant en compte les pondérations liées à la stratification du plan de sondage ;
- les intervalles de confiance à 5 % des coefficients d’émission (à partir d’une hypothèse de normalité)<sup>3</sup>.

Le premier indicateur permet d’observer l’évolution des taux d’émission de chaque segment de population issu de la classification. Le deuxième permet de tester l’hypothèse de stabilité des coefficients observés compte tenu des échantillons disponibles.

Dans ce type de modèles, la variable à expliquer est quantitative et les variables explicatives sont qualitatives. Pour les variables explicatives, nous avons donc eu recours à une transformation des données des EMD, de données quantitatives en données qualitatives<sup>4</sup>. Les variables retenues pour la classification sont celles usuellement retenues dans les études académiques (Kollo et Purvis, 1984 ; Kitamura et Kostyniuk, 1986 ; Walker et Peng, 1991 ; Mwakalonge, 2010, par exemple) et de planification :

- pour le motif travail, nous avons retenu le nombre de personnes du ménage déclarant avoir un emploi (nbperstrav\_F), décomposé en 3 catégories : 0, 1 et 2 ou plus ; le nombre de véhicules disponibles dans le ménage (nbvpsdispo\_F), décomposé en 4 catégories : 0, 1, 2 et 3 ou plus ; et, finalement, le cycle de vie (cycledevie), décomposé en 7 catégories en fonction de l’âge de la personne de référence, de l’existence d’un lien conjugal dans le ménage et de la présence d’enfants (voir l’annexe B pour le détail) ;
- pour les motifs autres que travail et enseignement, nous avons retenu le nombre de personnes du ménage (nbpers\_F), décomposé en 5 catégories : 1, 2, 3, 4 et 5 personnes ou plus ; ainsi que nbvpsdispo\_F et cycledevie.

Les variables relatives au nombre de personnes (ou nombre de personnes travaillant) à la motorisation sont usuellement retenues en tant que variables explica-

3. Calculés à l’aide de la librairie SURVEY (Lumley, 2010) du logiciel R, permettant la prise en compte du plan de sondage.

4. Pour distinguer l’utilisation des versions quantitative et qualitative d’une même variable explicative, on rajoute le suffixe « \_F » (de l’anglais *factor*) à la version qualitative de la variable. Ainsi, par exemple, nbpers désigne le nombre de personnes dans sa version quantitative et nbpers\_F dans sa version qualitative.

tives. Si le pouvoir discriminant du nombre de personnes reste important de manière assez générale (en raison d'un effet taille), l'apport de la motorisation en termes de réduction de la variance totale est une question ouverte. Kitamura et Kostyniuk (1986) observent que la motorisation des ménages a beaucoup augmenté depuis les années 1950 aux États-Unis au point que l'on peut parler d'un état de « maturité » (87% des ménages possède au moins un véhicule en 1980). Cette « maturité » de la motorisation s'accompagne d'une baisse du pouvoir discriminant de la motorisation par rapport aux comportements de mobilité des ménages. En effet, l'accès généralisé à la voiture se traduit par une plus grande hétérogénéité au sein des propriétaires de véhicules et par une diminution de la corrélation entre motorisation et revenu. En revanche, la structure des ménages, caractérisée par le cycle de vie<sup>5</sup>, semble avoir un pouvoir explicatif aussi important que les variables de type nombre de personnes. En effet, on peut penser que les caractéristiques des ménages en termes d'âge de la personne de référence, de la présence de lien conjugal ou d'enfants conditionnent les besoins de celui-ci en termes d'activités à accomplir et les décisions sur la manière d'organiser l'accomplissement de ces besoins. Il est envisageable, par exemple, que les liens conjugaux et de parenté aient une traduction en termes d'une centralisation plus grande de certaines tâches comme l'achat de provisions alimentaires ou qu'ils donnent lieu à la nécessité de certains déplacements spécifiques de type accompagnement. Par ailleurs, dans le cadre d'une analyse par cohortes visant à distinguer les évolutions de la mobilité en véhicules particuliers liées à des changements de comportements et celles liées à des évolutions démographiques, Sun (2009) remarque que l'on n'observe pas de différence significative sur le nombre de déplacements des ménages appartenant au même stade dans le cycle de vie, suivant différentes localisations du logement, ce qui suggère une similarité des styles de vie (alors que la motorisation dépend fortement de la localisation). En revanche, certaines évolutions comportementales ont été observées pour les différents stades du cycle de vie.

Dans le cas de l'agglomération lyonnaise, on pourrait penser qu'on arrive au stade de maturité de la motorisation, même si la motorisation des ménages est moins importante qu'aux États-Unis (voir table 3.1). Les résultats d'analyses de la variance préliminaire pour valider le choix des variables explicatives retenues ont montré une capacité discriminante supérieure pour le cycle de vie que pour la motorisation. Cette dernière semble, par ailleurs, voir son pouvoir discriminant se réduire avec le temps.

**4.1.2.2.2 Les modèles de régression à l'échelle zonale** Le modèle de régression à l'échelle zonale est l'ancêtre des autres approches de modélisation de la génération et de ce fait, ainsi que par sa facilité d'application, il est un bon repère par rapport auquel on peut comparer les résultats des autres approches. Par ailleurs, bien qu'ancienne, on peut encore retrouver cette méthode dans la pratique (TRB, 2007; DREIF, 2008).

Comme pour les modèles par classification croisée, nous calibrons des modèles de régression à l'échelle zonale pour chaque date d'enquête et périmètre disponibles. Pour étudier l'évolution dans le temps, la même spécification est estimée à chaque fois. Le zonage retenu est, bien entendu, identique pour tous les calages et il correspond

---

5. L'intérêt d'introduire la variable cycle de vie dans l'analyse de la mobilité a été mis en évidence par Heggie (1978), Zimmerman (1982) et Clarke *et al.* (1982), entre autres.

au zonage défini à partir des secteurs de tirage des EMD (voir section 3.2.1.2.2). Les indicateurs retenus pour l'analyse sont ceux usuellement retenus (donnés par tous les logiciels) pour le diagnostic : coefficient de détermination ( $R^2$ ) et le t de Student empirique (t-value). Cette méthodologie est proche de celles d'autres études présentées dans la section 2.2.1 (Dodd, 1966 ; Ashford et Holloway, 1972, par exemple).

Les variables disponibles pour le choix de la spécification des modèles sont les suivantes :

- nbpers\_i : nombre de personnes de la zone ;
- nbperstrav\_i : nombre de personnes de la zone déclarant avoir un emploi ;
- nbhom\_i : nombre d'hommes de la zone ;
- nbfem\_i : nombre de femmes de la zone ;
- nbhomtrav\_i : nombre d'hommes de la zone déclarant avoir un emploi ;
- nbfemtrav\_i : nombre de femmes de la zone déclarant avoir un emploi ;
- nbvdispo\_i : nombre de véhicules disponibles pour les résidents de la zone ;
- nbmenavecenf\_i : nombre de ménages comportant au moins un enfant ;
- nbmenpresfoyer\_i : nombre de ménages comportant au moins une personne au foyer.

Le complément de notation « \_i » permet d'indiquer qu'il s'agit d'une variable à l'échelle de la zone. Les variables nbmenavecenf\_i et nbmenpresfoyer\_i ont été introduites pour essayer de rendre compte de la diminution du nombre de demi-sorties pour le motif travail en représentant, en quelque sorte, l'incitation (ou besoin) à déjeuner au domicile.

Avec un zonage qui ne change pas dans le temps, il aurait théoriquement été possible d'approfondir l'étude de la dynamique de la génération en profitant d'une structure de données en forme de panel. Chaque zone correspond, en effet, à une « entité individuelle » observée sur plusieurs dates — ou plus précisément 2 dates, en admettant que l'on utilise les données de 1985 et 1995 pour calibrer un modèle et qu'on conserve celles de 2006 pour réaliser des tests de capacité prédictive du modèle. Nous avons exploré cette possibilité<sup>6</sup>, mais nos constats préliminaires nous ont conduit à l'écarter de cette présentation. En effet, les données de panel permettent, en principe, d'estimer des modèles dynamiques intégrant des variables retardées (endogènes et exogènes). Dans le cas présent, nous ne disposons que de 3 dates, ce qui limite les estimateurs utilisables pour un modèle dynamique (Bond, 2002) à celui d'Anderson et Hsiao (1981, 1982) qui, comme le note Arellano (1989), a une variance très élevée. En utilisant juste 2 dates d'enquête, le modèle dynamique perd de son sens (impossibilité de l'estimer). On aurait pu étudier des modèles non-retardés (*within, pooling* ou *between*, par exemple) pour mettre en évidence des effets individuels liés aux zones et la possibilité de combiner plusieurs données d'enquête dans le cadre d'un modèle statique. Cependant, la non disponibilité d'une méthode informatisée pour faire les prévisions hors échantillon (*out-of-sample prediction*) rendait l'exercice trop coûteux par rapport aux gains en termes analytiques ou pratiques<sup>7</sup>.

---

6. À l'aide de la librairie `plm` (Croissant et Millo, 2008) du logiciel statistique R (R Development Core Team, 2011).

7. Notons que des développements dans ce sens du package `plm` sont prévus, mais pas à un horizon proche. Plus de détails sur la prévision hors échantillon, à partir de modèles calés sur données de panel, peuvent être trouvés dans Baltagi (2007) ou encore, plus spécifique au domaine des transports, dans Kitamura (1990).

**4.1.2.2.3 Les modèles de régression à l'échelle des ménages** Les modèles de régression à l'échelle des ménages permettent *a priori* d'inclure plus de variables que ceux à l'échelle zonale. En effet, à l'échelle zonale les variables explicatives sont souvent construites par agrégation de données disponibles à l'échelle des ménages. Par la procédure d'agrégation, on aboutit à des corrélations importantes entre variables explicatives. Ces corrélations sont moins importantes à l'échelle des ménages. Par ailleurs, à l'échelle des ménages, on peut introduire des variables susceptibles de refléter la coordination des comportements de mobilité au sein de l'unité de prise de décision. Ils sont les plus étudiées dans la littérature académique empirique (voir le chapitre 2).

Les variables disponibles pour le choix de la spécification des modèles sont les suivantes :

- nbpers : nombre de personnes dans le ménage ;
- nbperstrav : nombre de personnes du ménage déclarant avoir un emploi ;
- nbhom : nombre d'hommes dans le ménage ;
- nbfem : nombre de femmes dans le ménage ;
- nbhomtrav : nombre d'hommes du ménage déclarant avoir un emploi ;
- nbfemtrav : nombre de femmes du ménage déclarant avoir un emploi ;
- nbpers18p : nombre de personnes du ménage ayant 18 ans ou plus ;
- nbvdispo : nombre de véhicules disponibles pour l'usage des membres du ménage ;
- nbpermis : nombre de personnes du ménage ayant le permis de conduire ;
- presvp : *dummy*-variable prenant la valeur 1 lorsque le ménage a au moins un véhicule à disposition, 0 sinon ;
- presfoyer : *dummy*-variable prenant la valeur 1 lorsque le ménage a au moins un de ses membres restant au foyer, 0 sinon ;
- presconj : *dummy*-variable prenant la valeur 1 lorsque le lien conjugal est présent dans le ménage ;
- presenf : *dummy*-variable prenant la valeur 1 lorsque le lien de parenté est présent dans le ménage ;
- presret : *dummy*-variable prenant la valeur 1 lorsqu'au moins un des membres du ménage est à la retraite, 0 sinon ;
- preschom : *dummy*-variable prenant la valeur 1 lorsqu'au moins un des membres du ménage est au chômage, 0 sinon.

Nous avons envisagé d'inclure une variable indiquant l'âge de la personne de référence du ménage, mais la relation entre cette variable et la mobilité n'étant certainement pas linéaire, nous ne l'avons finalement pas incluse (nos essais de transformation de la variable pour trouver une relation linéaire significative ont été infructueux). Les spécifications des modèles finalement retenues sont évoquées un peu plus loin dans l'exposé (points 4.4.1).

Lumley (2010) remarque que la non prise en compte des inégalités dans les probabilités de tirage peut distordre les relations établies entre les différentes variables de l'échantillon. Dans les modèles de régression, ignorer les coefficients de redressement revient à avoir des estimations correspondant à une population avec une distribution différente pour certaines des variables. Or, l'objectif primaire des modèles de régression est d'établir des associations qui ont un comportement stable sur un éventail de populations. Il devrait, en conséquence, être possible d'estimer ces associations sans

utiliser les coefficients de redressement. Un examen des biais potentiels pour certaines hypothèses sur les vraies relations est donné par DuMouchel et Duncan (1983).

Nous examinons des modèles calibrés avec et sans prise en compte des coefficients de redressement <sup>8</sup>.

#### 4.1.2.2.4 Les modèles par arbres de segmentation à l'échelle des ménages

Comme nous avons vu dans la section 1.2.1.3, les arbres de segmentation présentent une alternative intéressante dans la modélisation de la génération car ils combinent la flexibilité des méthodes de classification avec la disponibilité d'une méthode plus ou moins standard de choix des variables. Dans les points 4.1.2.2.1 à 4.1.2.2.3) ci-dessus, nous avons défini une méthodologie permettant de tester la stabilité temporelle des paramètres des modèles de classification croisée et de régression. L'objectif de l'étude des arbres de segmentation est un peu différent. On souhaite obtenir quelques informations sur l'évolution du choix des variables explicatives, c'est-à-dire sur l'évolution dans le temps de la spécification elle-même.

Nous avons choisi d'appliquer l'algorithme CART avec une variable à expliquer qualitative <sup>9</sup> : le profil de mobilité, défini en termes du nombre de sorties d'un certain type (voir la section 4.1.1.2.1) réalisées par les membres du ménage. Pour le motif travail, la variable à expliquer contient 5 modalités correspondant à des quantités de sorties réalisées par le ménage (0, 1\_2, 3\_4, 5\_6, 7p). Pour les motifs autres que travail et enseignement, la variable à expliquer contient sept modalités (0, 1\_2, 3\_4, 5\_6, 7\_8, 9\_10, 11p). Le principe de notation est le suivant : 0 indique que le ménage n'a pas réalisé de sortie. x\_y indique que le ménage a réalisé x ou y demi-sorties. En général les ménages réalisent un nombre pair de sorties y. Le nombre impair x est inclus en raison de la prise en compte des déplacements des sorties incomplètes (en dehors de la tranche horaire de recueil des données) xp correspond à « x demi-sorties ou plus ».

Les variables explicatives potentielles auxquelles on applique l'algorithme correspondent aux équivalents numériques de celles qui ont été retenues pour les régressions au niveau des ménages (il n'y a plus besoin de faire attention au problème de corrélation). À ces variables, on ajoute les variables cycle de vie et âge de la personne de référence du ménage, qu'on avait dû exclure des régressions à l'échelle des ménages.

L'application de l'algorithme a été réalisée à l'aide de la librairie `rpart` (Therneau et Atkinson, 2011) du logiciel R (R Development Core Team, 2011). L'algorithme a été appliqué aux données non pondérées ainsi qu'aux données pondérées de manière à refléter le plan de sondage. Les pondérations retenues pour cette section ne cherchent pas à prendre totalement en compte les effets du plan de sondage complexe (via le calcul ajusté des écarts-types, par exemple). Elles cherchent simplement à donner une idée plus juste de la probabilité de tirage de chaque observation dans la construction des arbres. Les poids choisis sont les coefficients de redressement des différentes enquêtes auxquels on applique un coefficient de correction égal au rapport entre le nombre de ménages observés et l'effectif redressé <sup>10</sup>. De cette manière, nous pouvons définir les mêmes critères d'arrêt pour l'ensemble des arbres :

---

8. La prise en compte des coefficients de redressement se fait à l'aide de la librairie `survey` (Lumley, 2010) du logiciel R (R Development Core Team, 2011).

9. Ce choix tient au lien entre ce travail et le projet SIMBAD. Voir l'annexe D.1.

10. Notons que cette manière de procéder est similaire à démultiplier les observations en fonction

- un seuil limite de réduction de l’impureté de 0,000 5. Si le gain apporté par la division est inférieur à ce seuil, celle-ci n’est pas effectuée ;
- une taille minimale de segment divisible égale à 60 ménages ;
- une taille minimale dans les segments terminaux égale à 30 ménages.

Notons que ces critères d’élagage sont peu restrictifs, ce qui permet d’avoir une perspective plus large du pouvoir de discrimination des variables explicatives auxquelles on applique l’algorithme.

#### 4.1.2.3 Évaluer la stabilité temporelle des modèles

La manière la plus rigoureuse, sur le plan statistique, d’évaluer la validité d’un même modèle à plusieurs moments dans le temps est de tester (dans le sens statistique du terme) l’égalité des coefficients de celui-ci dans le temps. Parmi les modèles retenus dans ce chapitre, seulement les trois premières catégories se prêtent bien à un tel exercice (la dernière approche de modélisation a été incluse explicitement pour évaluer, de manière plus large, la stabilité du classement des variables explicatives).

Les coefficients des modèles de classification sont, avant tout, des moyennes évaluées sur une certaine population. Il est donc possible, en théorie, de tester l’égalité de ces deux moyennes par un test de Student d’égalité de la moyenne de deux échantillons (Snedecor et Cochran, 1989). Dans la pratique, il serait nécessaire de modifier ce test si l’on souhaite prendre en compte le plan de sondage (notamment dans le calcul des degrés de liberté pour la définition de la valeur issue de la distribution théorique). Cet ajustement ne faisant, à notre connaissance, pas encore partie des outils de calcul utilisés, nous avons plutôt opté pour tester le recoupement des intervalles de confiance. Si, pour deux moyennes que l’on souhaite comparer, les intervalles de confiance sont disjoints, alors on rejette l’hypothèse de stabilité des coefficients d’émission. Inversement, si l’intersection des intervalles de confiance est différente de l’ensemble vide, alors on ne peut pas rejeter l’hypothèse de stabilité. Le seuil retenu pour le test est de 95 %. Dans ce qui suit, il sera fait référence à ce test aussi sous l’appellation « IC-test ».

Pour les modèles par régressions, nous avons opté pour l’approche classique donnée par Galbraith et Hensher (1982) et Ortúzar et Willumsen (2001). Le test est donné dans l’équation (4.1) ci-dessous :

$$t^* = \frac{\theta_u - \theta_v}{\sqrt{\left(\frac{\theta_u}{t_u}\right)^2 + \left(\frac{\theta_v}{t_v}\right)^2}} < 1,96 \quad (4.1)$$

des coefficients de redressement. En effet, les logiciels statistiques ont une manière particulière de traiter les échantillons avec des observations pondérées. En général, la spécification de pondérations conduit à ce que ces dernières soient traitées comme des « pondérations de fréquence » (de l’anglais *frequency weights*, Lumley, 2010). Celles-ci indiquent le fait que, pour chaque observation, l’échantillon (et non la population) est effectivement constitué d’un nombre d’observations identiques égal à la valeur de la pondération. Dit autrement, les coefficients de redressement ne sont pas traités convenablement en tant que « pondérations de sondage » (de l’anglais *sampling weights*, Lumley, 2010), indiquant qu’une observation de l’échantillon représente un certain nombre d’individus de la population (fonction de la probabilité de tirage). Le traitement approprié des pondérations est limité à certaines librairies spécialisées comme *survey*, mais il n’existe pas à notre connaissance une librairie permettant de le faire dans le cas de l’algorithme CART. Le coefficient de correction appliqué ici permet de conserver la même « échelle » en termes d’effectif total de l’échantillon pour la définition des critères d’arrêt dans la construction des arbres.

où  $\theta$  est un paramètre du modèle estimé à partir des données de la date  $u$  ( $\theta_u$ ) et de la date  $v$  ( $\theta_v$ ).  $t_u$  et  $t_v$  sont les valeurs empiriques du  $t$  de Student correspondantes. Si la relation donnée par l'équation (4.1) est respectée, alors on ne peut pas rejeter l'hypothèse d'égalité (de stabilité) des paramètres. Le niveau de confiance de ce test est 5 %.

#### 4.1.2.4 Indicateurs de qualité des calages et prévisions

Nous retenons trois indicateurs principaux, correspondant à différentes échelles d'analyse. Le premier indicateur concerne la capacité d'un modèle de génération à reproduire la mobilité totale, c'est-à-dire le nombre total de demi-sorties ayant lieu dans l'aire d'étude. Il s'agit tout simplement de l'écart relatif (en pourcentage) donné par l'équation (4.2), ci-dessous :

$$EREL = 100 \cdot \frac{T - N}{N}. \quad (4.2)$$

où  $T$  est le nombre total de demi-sorties estimées par le modèle et  $N$  est le nombre total de demi-sorties observées. Notons que la plupart des modèles analysés dans ce qui suit reproduisent, par construction, assez bien le nombre moyen ou le nombre total des demi-sorties. Pour cette raison, cet indicateur sera surtout utile pour juger de la capacité prédictive des modèles (où des surestimations ou sous-estimations du nombre total de demi-sorties peuvent vraiment avoir lieu).

Le deuxième indicateur porte sur l'échelle zonale. Dans la mesure où, dans la séquence de modélisation, un modèle de génération doit fournir des quantités de demi-sorties émises par les différentes zones de l'aire d'étude, analyser la performance des différents modèles à cette échelle est indispensable. L'indicateur retenu est une forme de racine carrée normalisée de l'erreur quadratique moyenne (ou *SRMSE* de par le sigle de l'indicateur en anglais<sup>11</sup>), donné dans l'équation (4.3) ci-dessous :

$$SRMSEG^{12} = \frac{\sqrt{\frac{\sum_i (N_i - T_i)^2}{I}}}{\frac{\sum_i N_i}{I}} \quad (4.3)$$

où  $I$  est le nombre de zones de l'aire d'étude et  $i$  est un indice correspondant à la zone de sorte que  $i \in \{1 \dots I\}$ .  $N_i$  et  $T_i$  sont respectivement le nombre de demi-sorties observées et modélisées pour la zone  $i$  ; on a donc  $\sum_i N_i = N$  et  $\sum_i T_i = T$ . D'après Fotheringham et Knudsen (1987), l'utilisation du *SRMSE* est préférable à celle des autres types de *RMSE* car ces dernières sont très sensibles aux ordres de grandeur des données. La statistique prend également en compte le nombre de zones ce qui la rend particulièrement attractive pour notre étude dans la mesure où des périmètres différents sont pris en compte dans l'analyse. Sa limite inférieure est zéro — réplique parfaite des données — et elle est généralement inférieure à un — une valeur supérieure à un indique que l'erreur moyenne est plus grande que la

---

11. *Standardized root mean square error*.

12.  $G$  est ajouté à *SRMSE* pour signifier que la statistique correspond aux modèles de génération et la distinguer ainsi de la statistique correspondante appliquée aux modèles de distribution dans le chapitre suivant (5) de la présente étude.

valeur moyenne. La sensibilité du *SRMSE* aux erreurs est linéaire, ce qui lui confère, d'après Fotheringham et Knudsen, une supériorité par rapport aux autres mesures générales de distance « zone à zone » pour comparer un même modèle dans deux contextes différents.

Le troisième indicateur porte sur l'échelle de prise de décision, soit le ménage. Il s'agit de l'erreur moyenne absolue, ou *MAE* de par ses sigles en anglais<sup>13</sup>, qui, concrètement, donne une estimation de l'erreur de modélisation ou prévision en nombre de demi-sorties par ménage. Son calcul s'effectue selon la formule donnée par l'équation (4.4), ci-dessous :

$$MAE = \frac{\sum_{h=1}^{n(H)} |T_h - N_h| w_h}{\sum_{h=1}^{n(H)} w_h} \quad (4.4)$$

où  $H$  est l'ensemble des ménages de l'échantillon et  $h$  est un indice correspondant sur les ménages enquêtés de sorte que  $h \in \{1 \dots n(H)\}$ , où  $n(H)$  est le nombre de ménages de  $H$ . De manière analogue à l'indicateur précédent,  $N_h$  et  $T_h$  sont respectivement le nombre de demi-sorties observées et modélisées pour le ménage  $h$ , ce qui fait que  $\sum_h N_h = N$  et  $\sum_h T_h = T$ .  $w_h$  est le coefficient de redressement associé au ménage  $h$ . Cet indicateur est assez courant dans la littérature ; il a également été utilisé par Koppelman et Wilmot (1982), Badoe et Chen (2004a) et Mwakalonge et Badoe (2012), entre autres. Notons que, par définition, il ne s'applique pas aux modèles de génération par régression zonale.

## 4.2 Étude de la mobilité moyenne par segment de population : est-ce que les émissions moyennes obtenues dans les modèles de classification croisée sont transférables dans le temps ?

Cette section présente les résultats de l'application du modèle de classification croisée aux données de l'agglomération lyonnaise. Elle présente également l'analyse de la stabilité des taux d'émission et des capacités prédictives des différentes classifications retenues, pour le motif travail, ainsi que pour les motifs autres que travail et enseignement.

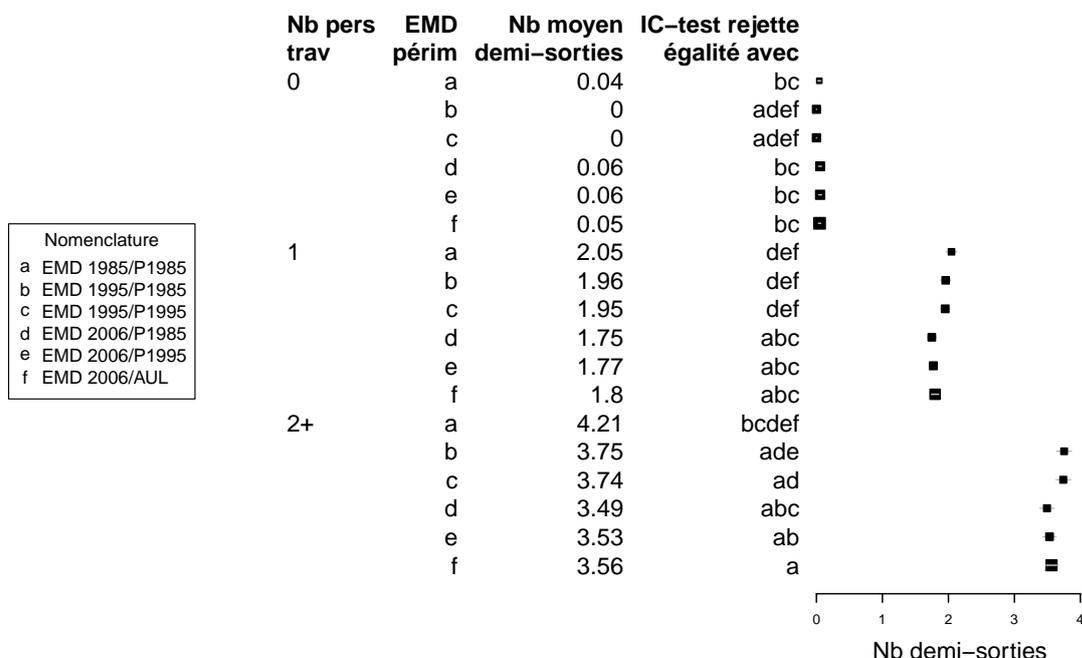
### 4.2.1 Stabilité ou instabilité des taux d'émission dans les classifications croisées ?

#### 4.2.1.1 Une seule variable explicative

Les figures 4.1 à 4.6 présentent le nombre de demi-sorties moyen suivant les modalités des variables `nbperstrav_F`, `nbpvdispo_F` et `cycledevie`, pour le motif

13. *Mean absolute error.*

FIGURE 4.1 – *Forest plot* du nombre de demi-sorties par ménage, pour le motif Travail, en fonction du nombre d’actifs employés (pour les différentes dates d’enquête et périmètres).



*Note de lecture : sur la septième ligne, on peut lire que les ménages comportant une personne active employée, enquêtés en 1985 et résidant dans le périmètre de 1985 (a), réalisent en moyenne 2,05 demi-sorties au sein de ce périmètre. Le nombre de demi-sorties moyen est donné par un carré dont la taille est proportionnelle au nombre de ménages enquêtés dans le segment correspondant. Les lignes latérales (peu visibles ici en raison du nombre limité de modalités retenues) indiquent l’intervalle de confiance à 5%. L’IC-test correspondant permet de rejeter l’hypothèse d’égalité du coefficient avec ceux des ménages similaires enquêtés en 2006, pour tous les périmètres (d, e et f).*

*Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006.*

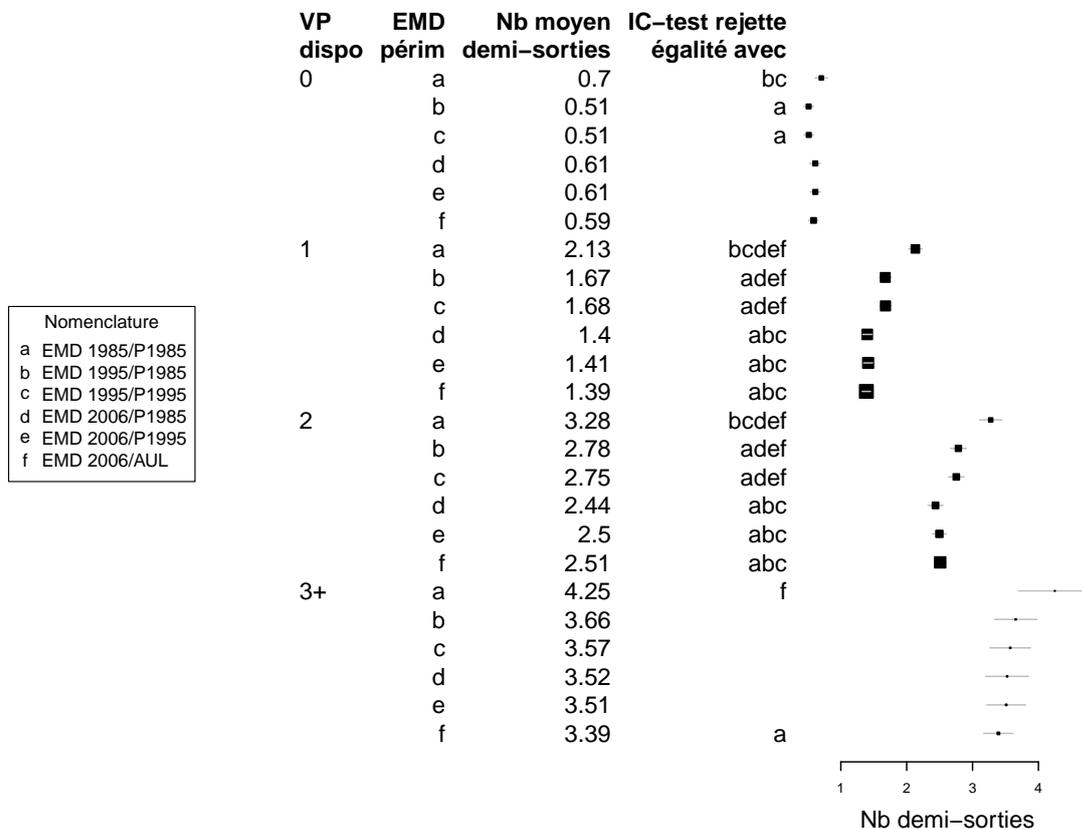
travail et nbpers\_F, nbpvdispo\_F et cycledevie pour un motif regroupant les motifs autres que le travail et l’enseignement.

**4.2.1.1.1 Le motif Travail** La figure 4.1 présente la classification pour le motif travail à partir du nombre de personnes actives employées. On remarque naturellement un lien positif entre celles-ci et le nombre de demi-sorties pour le motif travail, pour l’ensemble des enquêtes et périmètres. Pour les ménages n’ayant pas de personnes déclarant avoir un emploi, les taux d’émission sont nuls (ou quasiment nuls). Il y

a, en effet, des taux légèrement positifs en 1985 et 2006, qui sont liés aux étudiants qui travaillent (la variable statut les classe comme étudiants et non comme actifs employés). En 1995, cette catégorie n'apparaît pas dans l'échantillon (il n'y a pas de personnes réalisant des sorties pour le motif travail et pour le motif études supérieures dans la journée). L'IC-test d'égalité des paramètres indique une différence significative entre 1995 et les deux autres années (provenant en partie du fait que l'écart-type pour les coefficients de 1995 est nul). Pour les ménages ayant une personne active employée, on passe à des taux d'émission compris entre 2,05 en 1985 et 1,75 en 2006, sur le périmètre de 1985. Il y a donc une baisse de la mobilité moyenne d'un peu moins de 15 % pour cette catégorie au cours du temps, notamment entre 1995 et 2006 (la différence entre les coefficients de 2006 et les autres est significative). Pour les ménages ayant deux actifs employés ou plus, les taux d'émission se situent entre 4,21 en 1985, et 3,49 en 2006, sur le périmètre de 1985, soit une baisse d'environ 17 %. Comme pour les ménages avec un actif employé, la mobilité baisse avec le temps et, cette fois, les différences sont significatives entre toutes les dates d'enquête, sur le périmètre de 1985. L'élargissement du périmètre d'analyse conduit à une baisse infime du coefficient en 1995 et à une légère augmentation de celui-ci en 2006, pour les ménages ayant des actifs employés, sans que cette augmentation donne lieu au rejet de l'hypothèse de stabilité, pour chacune de ces années. En revanche, ces évolutions ont pour effet de rendre les différences entre les paramètres de 1995, sur le périmètre de 1995, et de 2006, sur le périmètre de 1995 et sur celui de l'AUL, non significatives, pour les ménages ayant deux actifs employés ou plus.

La figure 4.2 présente la classification pour le motif travail à partir du nombre de véhicules disponibles pour l'usage du ménage. Comme pour le nombre de personnes ayant un emploi, la relation entre nombre de véhicules disponibles et mobilité est croissante pour les trois dates d'enquête. Les ménages enquêtés sont majoritairement motorisés et disposent d'un ou deux véhicules. Pour ces catégories, on observe de très fortes variations du nombre de demi-sorties moyen d'une enquête à l'autre (les taux d'émission restent très proches en comparant différents périmètres pour une même date d'enquête). Pour les ménages ayant un seul véhicule à disposition, on constate une baisse du nombre de demi-sorties d'environ 35 % entre 1985 et 2006. Pour les ménages disposant de deux véhicules, la baisse est d'environ 25 %. Pour ces deux catégories, les coefficients sont significativement différents pour toute paire de dates d'enquête. Les ménages non motorisés ont des taux d'émission assez faibles (inférieurs à un) ce qui rend les écarts relatifs assez importants. Ces écarts restent cependant moins importants que ceux constatés pour les ménages disposant d'un ou deux véhicules (l'écart maximum est d'environ 20 %). L'évolution des taux d'émission, pour cette catégorie, n'est pas seulement décroissante dans le temps ; elle est décroissante entre 1985 et 1995 et croissante entre 1995 et 2006. Ces éléments font qu'on pourrait parler d'une relative stabilité pour les ménages non motorisés. Ceci est d'ailleurs corroboré par le fait que les coefficients sont significativement différents entre 1985 et 1995, mais ceux de 2006 ne sont significativement différents d'aucun autre. Les ménages enquêtés ayant à disposition trois véhicules ou plus sont relativement moins nombreux que les autres et de ce fait les écarts-types des coefficients d'émission sont relativement importants. On remarque une baisse assez marquée de la moyenne entre 1985 et 1995 pour ensuite avoir une baisse continue mais très faible. Cette baisse est valable au niveau des années d'enquête mais également au niveau des périmètres, de

FIGURE 4.2 – *Forest plot* du nombre de demi-sorties par ménage, pour le motif Travail, en fonction de véhicules à disposition du ménage (pour les différentes dates d'enquête et périmètres).



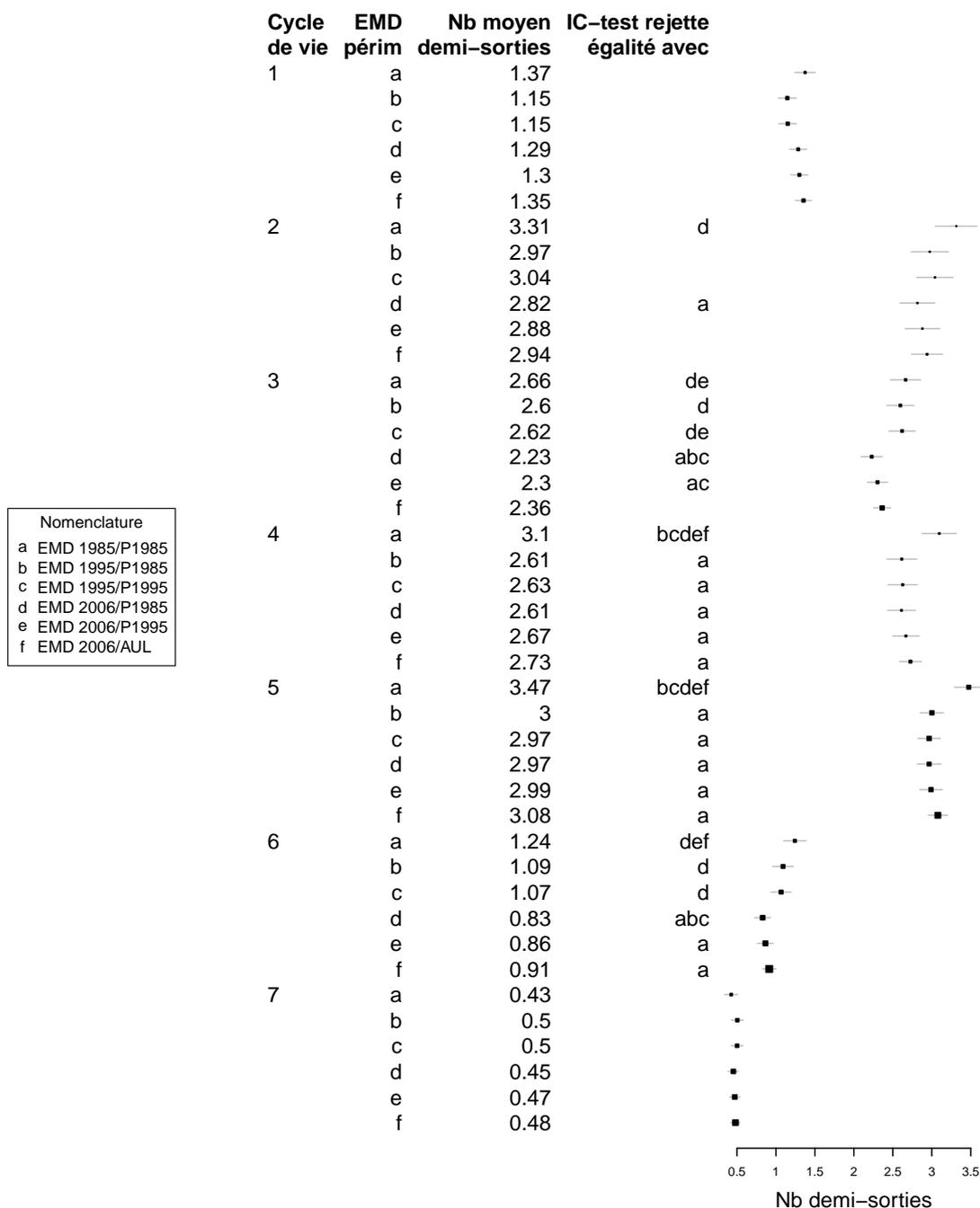
Note de lecture : voir figure 4.1.

Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006.

telle sorte que la seule différence significative peut être constatée entre le coefficient correspondant aux données de l'EMD de 1985 et celui correspondant aux données de l'EMD de 2006 sur le périmètre de l'AUL. L'écart maximum du coefficient est d'environ 20 %.

La figure 4.3 présente la classification pour le motif travail à partir du cycle de vie du ménage. La mobilité selon les différents stades du cycle de vie est similaire à celle que l'on peut trouver dans d'autres études utilisant cette variable (Kitamura et Kostyniuk, 1986 ; Kumar et Levinson, 1993). Le nombre de demi-sorties par ménage augmente puis diminue avec l'âge de la personne de référence, les stades ayant la mobilité la plus importante correspondent à ceux composés simultanément de parents et leurs enfants. Lorsqu'il n'y a pas d'enfant dans le ménage, la mobilité est plus élevée pour les ménages où la personne de référence est « plutôt jeune » que pour les ménages ayant une personne de référence « plutôt âgée ». Si l'on perçoit les mêmes tendances à toutes les dates d'enquête, on est loin de voir une stabilité des

FIGURE 4.3 – *Forest plot* du nombre de demi-sorties par ménage, pour le motif Travail, en fonction du stade dans le cycle de vie du ménage (pour les différentes dates d'enquête et périmètres).



Note de lecture : voir figure 4.1. La nomenclature de la variable Cycle de vie peut être trouvée dans l'annexe B (page 214).

Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006.

taux d'émission. Ceux de 1985 sont systématiquement plus élevés que les autres, à l'exception de la dernière catégorie (personne de référence du ménage « plutôt âgée », sans conjoint). En particulier, pour les catégories 4 et 5 (ménages dont le plus jeune des enfants a au moins 5 ans), il y a une différence d'environ 15 % entre le coefficient de 1985 et les autres années. Le test par intervalles de confiance rejette l'hypothèse de stabilité entre 1985 et les autres années, pour lesquelles les valeurs des coefficients sont très proches et l'hypothèse de stabilité ne peut pas être rejetée. Les catégories 3 et 6 présentent une certaine stabilité entre 1985 et 1995, avec une rupture (à la baisse) en 2006. La mobilité de la catégorie 7 est stable mais très faible.

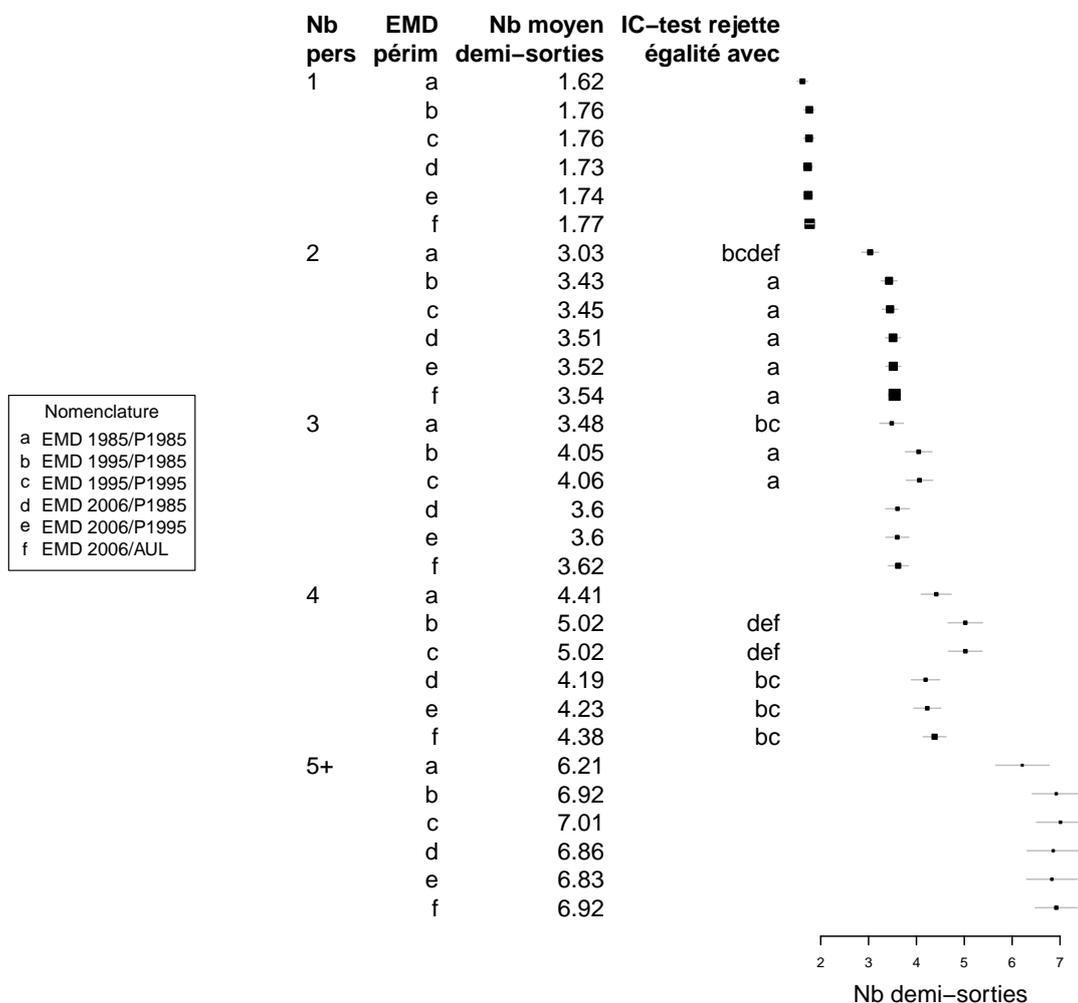
On a constaté, dans la section 4.1.1.2.1, que les sorties simples ont connu une baisse importante, entre 1985 et 1995, puis une relative stabilisation, entre 1995 et 2006. On a également noté que les sorties complexes ont augmenté entre 1985 puis diminué entre 1995 et 2006 sur le périmètre de 1985, l'élargissement du périmètre d'analyse se traduisant en 2006 par une augmentation de la mobilité. La composition de ces effets fait que l'on observe une décroissance continue des demi-sorties pour le motif travail entre 1985 et 2006. Le panorama dressé dans les figures 4.1 à 4.3 suggère des évolutions significatives dans les taux d'émission, en fonction des différentes variables de classification. Ces évolutions ne sont pas réparties de manière homogène dans toutes les modalités : certaines catégories présentent des hausses de la mobilité, d'autres des baisses ou les deux à la fois (et dans des ordres différents). Pour les classifications reposant sur le nombre d'actifs employés et sur le nombre de véhicules disponibles, on retrouve, à peu près, les tendances générales de baisse de la mobilité sur chacun des segments. Le cycle de vie présente plus d'hétérogénéité dans les évolutions de la mobilité au sein de chaque segment. Ceci n'est pas étonnant dans la mesure où cette variable est censée représenter de manière plus fine que les autres les interactions potentielles entre les différents membres du ménage, dans la mesure où elle prend en compte l'existence de liens particuliers entre les membres du ménage (le lien conjoint-conjoint ou parent-enfant, par exemple). Ces interactions potentielles sont à l'origine des possibilités de chaînage des déplacements. Si l'on s'intéresse au choix des variables pour les modèles de classification, dans une optique de stabilité des comportements observés à l'échelle des ménages<sup>14</sup>, le nombre de personnes ayant un emploi semble donner les classes pour lesquelles les taux d'émission sont relativement les plus stables alors que les évolutions associées au nombre de véhicules disponibles sont les plus importantes.

**4.2.1.1.2 Les motifs autres que Travail et Enseignement** La figure 4.4 présente la classification pour les motifs autres que le Travail et l'Enseignement à partir du nombre de personnes du ménage. L'accroissement de la taille du ménage a naturellement un effet positif sur la quantité de demi-sorties réalisées. On remarque une stabilité du nombre de demi-sorties en 1995 et 2006, pour les ménages à une ou deux personnes. En 1985, pour ces mêmes ménages, le nombre de demi-sorties est plus faible. L'écart est d'environ 8 % pour les ménages à une personne et de 13 % pour les ménages à deux personnes. Le test de la stabilité des coefficients d'émission à partir des intervalles de confiance ne permet de rejeter l'hypothèse de stabilité pour aucune des enquêtes, quel que soit le périmètre d'analyse retenu, pour la première catégorie.

---

14. Ce critère venant compléter la maximisation de la variance inter-classes (ANOVA).

FIGURE 4.4 – *Forest plot* du nombre de demi-sorties par ménage, pour les motifs autres que le motif Travail ou Enseignement, en fonction du nombre de personnes du ménage (pour les différentes dates d'enquête et périmètres).

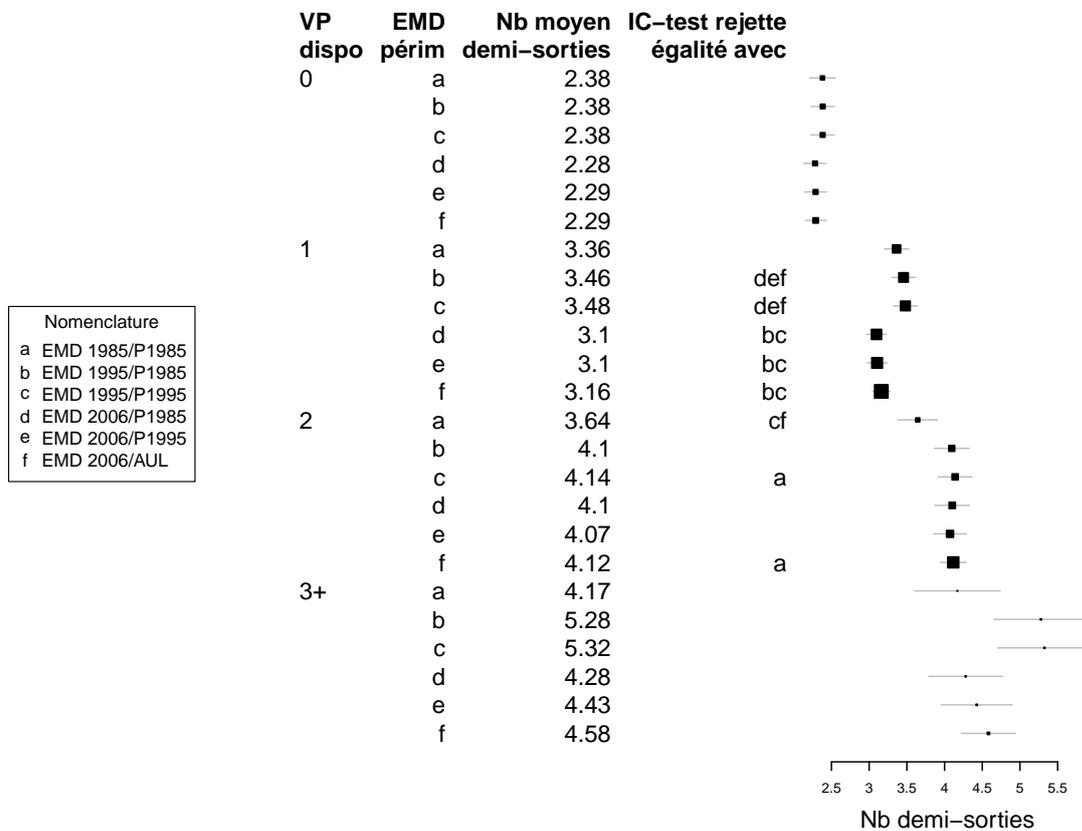


Note de lecture : voir figure 4.1.

Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006.

Pour la deuxième catégorie, la stabilité des paramètres n'est rejetée qu'entre 1985 et les autres années. Pour les ménages composés de 3 ou 4 personnes, on observe une situation un peu différente. Il y a une augmentation de la mobilité d'environ 16 % entre 1985 et 1995 puis une diminution d'environ 25 % entre 1995 et 2006, pour les premiers. Pour les seconds, l'augmentation est d'environ 14 % entre 1985 et 1995 et la diminution se situe entre 13 et 17 %, selon le périmètre retenu, entre 1995 et 2006. Ces écarts se retrouvent partiellement dans les tests de stabilité. Celle-ci doit être rejetée pour certaines des comparaisons mettant en jeu les coefficients obtenus pour 1995. Pour les ménages de 3 personnes, la différence entre les coefficients de 1985 et ceux

FIGURE 4.5 – *Forest plot* du nombre de demi-sorties par ménage, pour les motifs autres que le motif Travail ou Enseignement, en fonction de véhicules à disposition du ménage (pour les différentes dates d’enquête et périmètres).



Note de lecture : voir figure 4.1.

Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006.

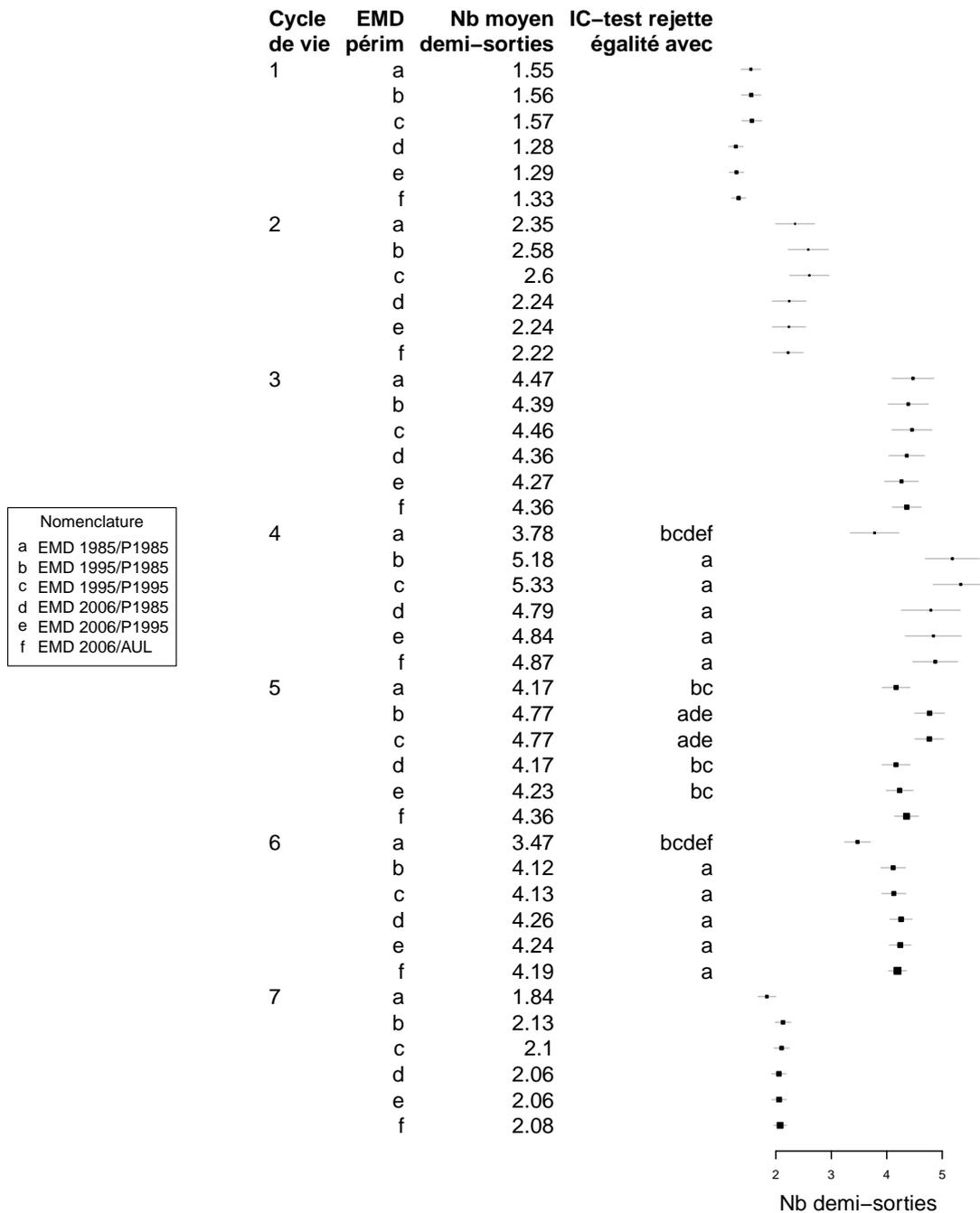
de 1995 est significative alors que, pour les ménages de 4 personnes, c’est la différence entre les coefficients de 1995 et ceux de 2006 qui l’est. Pour les ménages composés de 5 personnes ou plus, on trouve un cas de figure se situant entre les deux décrits précédemment : on a une augmentation des coefficients entre 1985 et 1995 puis une très légère diminution entre 1995 et 2006, l’écart le plus important étant d’environ 12%. Le test de stabilité des coefficients ne permet de rejeter cette hypothèse dans aucun des cas, mais ceci est dû en partie à un effectif enquêté relativement faible rendant les intervalles de confiance plus larges que ceux des autres catégories. L’écart entre le coefficient de 1985 est de 0,7 ou 0,8 demi-sorties, selon le périmètre, ce qui est un changement assez important en termes absolus.

La figure 4.5 présente la classification pour les motifs autres que le Travail et l’Enseignement à partir du nombre de véhicules à disposition du ménage. Comme pour le nombre de personnes, la relation avec la mobilité est croissante, bien que le passage de 2 à 3 véhicules ou plus semble avoir un impact marginal (sauf en 1995). Les

ménages non motorisés ont une mobilité très stable : le taux de variation maximum des taux d'émission sur la période étudiée est d'environ 4 % (baisse entre 1995 et 2006). Le test de stabilité à partir des intervalles de confiance ne permet pas de rejeter l'hypothèse de stabilité pour aucune des comparaisons possibles entre enquêtes. On retrouve également une stabilité assez marquée pour les ménages disposant de 2 véhicules entre 1995 et 2006. L'écart maximum sur cette période est inférieur à 2 %. La valeur des coefficients a connu une augmentation assez importante par rapport à 1985 (un peu moins de 14 %). En revanche, cette augmentation n'est statistiquement significative que lorsque l'on compare les enquêtes à périmètre évolutif (1985 avec 1995/P1995 et 2006/AUL). Les ménages disposant d'un seul véhicule connaissent une décroissance de la mobilité entre 1995 et 2006 d'un peu moins de 11 % (soit entre 0,3 et 0,4 demi-sorties). L'IC-test révèle une différence significative entre les coefficients de 1985 et 1995, d'une part, et ceux de 2006 d'autre part. Les ménages ayant à disposition 3 véhicules ou plus connaissent une augmentation du nombre de demi-sorties réalisées d'environ 27 % entre 1985 et 1995, puis une diminution d'entre 14 et 20 % en 2006 (les coefficients de 2006 augmentent avec la taille du périmètre considéré). Les écarts-types importants pour cette catégorie font que les taux d'émission pour 1995 ne sont pas statistiquement différents de ceux des autres années.

La figure 4.6 présente la classification pour les motifs autres que le Travail et l'Enseignement à partir du stade dans le cycle de vie du ménage. Le profil de la relation entre cycle de vie et mobilité des ménages diverge quelque peu de celle constatée pour le motif travail (figure 4.3). On peut toujours observer une relation d'abord croissante et puis décroissante mais les passages d'une modalité à une autre sont plus progressifs pour les trois premiers cycles. La mobilité des catégories 6 et 7, logiquement plus importante pour ce motif que pour le motif travail (en raison de l'inclusion de la mobilité des retraités), connaît une rupture marquée entre les gens vivant en couple et les autres : les couples sont bien plus mobiles. En termes d'évolutions observées pour chaque modalité de la variable de segmentation, on observe des cas de figure similaires à ceux qu'on a pu constater pour d'autres variables. Les ménages dans les stades 2 et 3 (couples ou ménages avec enfants en bas âge) présentent une certaine stabilité des taux d'émission sur l'ensemble de la période. Pour la première de ces catégories, l'écart maximal constaté est de 15 % entre 1995 (ensemble du périmètre de 1995) et 2006 (aire urbaine). Pour la deuxième catégorie, les écarts sont plus faibles (moins de 5 % entre 1985 et 2006, sur le périmètre de 1985). Les IC-tests ne permettent de rejeter l'hypothèse de stabilité dans aucun cas. Les ménages se situant dans les stades 6 et 7 connaissent une stabilité de la mobilité entre 1995 et 2006, l'écart maximum sur cette période étant aux alentours de 3 % pour les deux segments. Entre 1985 et 1995, la mobilité s'est accrue pour les deux catégories : environ 19 et 15 %, respectivement. Les écarts-types étant assez faibles, ces taux de variation se traduisent par le rejet de l'hypothèse de stabilité entre 1985 et le reste de la période étudiée, pour les ménages dans le stade 6 mais pas pour ceux dans le stade 7, ce qui suppose une grande stabilité de la mobilité pour les ménages de cette catégorie. Les ménages se situant au 4<sup>e</sup> stade du cycle de vie présentent une grande variabilité des comportements pour chaque année d'enquête, ce qui se traduit par des écarts types importants. De ce fait, alors que l'on peut observer une très forte augmentation des taux d'émission entre 1985 et 1995 (allant jusqu'à 41 %) suivie d'une diminution

FIGURE 4.6 – *Forest plot* du nombre de demi-sorties par ménage, pour les motifs autres que le motif Travail ou Enseignement, en fonction du stade dans le cycle de vie du ménage (pour les différentes dates d'enquête et périmètres).



Note de lecture : voir figure 4.1. La nomenclature de la variable Cycle de vie peut être trouvée dans l'annexe B (page 214).

Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006.

située entre 6 et 10 %, selon le périmètre considéré, entre 1995 et 2006, le test de stabilité ne rejette que celle entre 1985 et les autres dates. Dit autrement, on obtient le même résultat que celui obtenu pour le stade 6 qui, étant de composition plus homogène, a des taux d'émission moins variables. Pour les ménages du stade 1, on observe une stabilité du taux d'émission entre 1985 et 1995 (écart le plus important de 0,12 %) puis une baisse, comprise entre 15 et 18 %, entre 1995 et 2006. Cette variation est relativement faible et les tests de stabilité conduisent à admettre une stabilité sur l'ensemble de la période (aucune différence n'est significative). Pour les ménages au stade 5, on observe une augmentation des taux d'émission (un peu plus de 14 %) entre 1985 et 1995 puis une diminution (comprise entre 9 et 13 %, selon les périmètres) entre 1995 et 2006. Les taux d'émission de 1985 et 2006 sont, en conséquence, très proches (leur écart maximum est inférieur à 5 %). Le test de stabilité conduit à rejeter l'hypothèse de stabilité pour les comparaisons entre les taux de 1995 et les autres, sauf pour les coefficients calculés pour 2006 sur l'ensemble de l'AUL. Ces derniers ne sont significativement différents d'aucun autre des coefficients.

Dans des schémas de classification prenant en compte une seule variable explicative, les contraintes en matière d'effectif et de variabilité des comportements au sein de chaque catégorie retenue conduisent à dire que, de manière générale, les différences des coefficients, d'une année à une autre, inférieures à 8–9 % peuvent être considérées comme négligeables — elles conduisent à un non-rejet de l'hypothèse de stabilité des taux d'émission. Les différences supérieures à 25 % conduisent systématiquement à rejeter l'hypothèse de stabilité. La zone comprise entre ces deux valeurs est plus délicate dans la mesure où l'hétérogénéité à l'intérieur des classes (variables non retenues pour la segmentation) et la variabilité des comportements au sein d'une classe conduisent à des résultats contrastés. Quel est l'impact de ces écarts en termes de coefficients au moment de faire des prévisions ? Cette question sera étudiée dans la section 4.2.2. Le point suivant (4.2.1.2) s'intéresse à l'introduction d'une deuxième variable dans le but de réduire l'hétérogénéité au sein des classes.

#### 4.2.1.2 Deux variables explicatives

Les tables 4.2 à 4.5 présentent le nombre de demi-sorties moyen suivant le croisement de variables `nbperstrav_F` et `nbpvdispo_F` ainsi que `nbperstrav_F` et `cycledevie`, pour le motif travail, et `nbpers_F` et `nbpvdispo_F` ainsi que `nbpers_F` et `cycledevie`, pour un motif regroupant les motifs autres que le travail et l'enseignement.

**4.2.1.2.1 Le motif Travail** La table 4.2 donne l'évolution des coefficients d'émission de demi-sorties pour le motif travail, suivant une classification selon le nombre de personnes du ménage déclarant avoir un emploi et le nombre de véhicules disponibles. Le croisement des deux variables n'est donné que pour les ménages ayant au moins une personne active employée<sup>15</sup>. On observe, dans la plupart des cas, une baisse continue de la mobilité entre 1985 et 2006, à périmètre constant (allant de 13 à 22 % de la valeur de 1985). L'élargissement du périmètre d'analyse en 2006 se traduit par une

15. Comme on l'a vu dans le point précédent, la mobilité pour le motif travail des ménages n'ayant pas d'actif employé est négligeable.

TABLE 4.2 – Analyse de la stabilité temporelle du coefficient d'émission de demi-sorties par ménage, pour le motif Travail. Classification réalisée à partir de deux variables : nombre de personnes déclarant avoir un emploi et nombre de véhicules disponibles.

Modalité	Coefficient d'émission moyen ( $\mu$ ) et écart-type ( $se$ )												IC-test : différence significative entre $\mu$					
	1985/P1985 (a)		1995/P1985 (b)		1995/P1995 (c)		2006/P1985 (d)		2006/P1995 (e)		2006/AUL (f)		a	b	c	d	e	f
Nb pers trav	$\mu_a$	$se_a$	$\mu_b$	$se_b$	$\mu_c$	$se_c$	$\mu_d$	$se_d$	$\mu_e$	$se_e$	$\mu_f$	$se_f$						
1 Personne																		
0	1,89	0,10	1,65	0,08	1,65	0,08	1,66	0,08	1,66	0,08	1,66	0,07						
1	2,05	0,05	2,04	0,05	2,03	0,05	1,76	0,04	1,78	0,04	1,81	0,04	def					
2	2,17	0,10	1,99	0,07	1,98	0,07	1,78	0,08	1,83	0,07	1,86	0,06	def					
3+	2,18	0,26	2,01	0,20	1,87	0,19	1,92	0,25	1,91	0,24	1,79	0,15						
2 Personnes et +																		
0	3,60	0,32	2,99	0,35	3,03	0,34	3,29	0,35	3,39	0,34	3,38	0,33						
1	4,12	0,09	3,59	0,10	3,63	0,10	3,34	0,08	3,37	0,08	3,41	0,08	bedef	a	a	a	a	a
2	4,19	0,11	3,74	0,08	3,70	0,07	3,43	0,07	3,48	0,07	3,49	0,05	bedef	ad	a	ab	a	a
3+	5,28	0,34	4,48	0,19	4,40	0,18	4,27	0,18	4,22	0,16	4,13	0,13	ef					a

Note de lecture :  $a = EMD\ 1985$ ;  $b = EMD\ 1995\ Périn.;  $c = EMD\ 1995\ Périn.;  $d = EMD\ 2006\ Périn.;  $e = EMD\ 2006\ Périn. et  $f = EMD\ 2006\ Périn.. Sur la deuxième ligne et dernière colonne on observe que, pour les ménages ayant une personne déclarant travailler (comme activité principale) et un véhicule disponible, l'IC-test permet de rejeter l'hypothèse d'égalité du taux d'émission calculé avec les données de l'EMD de 2006 sur le périmètre de l'aire urbaine (f) avec ceux calculés avec les données de 1985 (a) et 1995, sur les deux périmètres (b et c). L'effectif observé est supérieur ou égal à 30 ménages pour toutes les classes.$$$$$

Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006.

légère augmentation du nombre de demi-sorties émises en moyenne. Une exception à cette tendance générale peut être observée au niveau des ménages non motorisés, à deux actifs. Pour cette catégorie, le taux d'émission décroît entre 1985 et 1995 puis s'accroît entre 1995 et 2006 (tout en restant inférieur au niveau de 1985). On observe que l'utilisation de deux variables dans la segmentation a eu pour effet de rendre les écarts types des coefficients d'émission assez importants (ceux-ci pouvant aller au-delà de 0,3, ce qui donne une amplitude de plus d'une demi-sortie aux intervalles de confiance à 95 %). Pour les catégories ayant des écarts-types le plus faibles (autour de 0,1), soit les ménages ayant à disposition 1 ou 2 véhicules, il existe des différences significatives entre les coefficients d'émission. Pour les ménages à un seul actif, les différences se situent notamment entre les paramètres de 1985 et ceux de 2006 (les paramètres de 1995 sont soit proches de ceux de 1985, pour les ménages ayant un seul véhicule, soit compris entre les 2 et non significativement différents, pour les ménages ayant 2 véhicules.). Pour les ménages à deux actifs, les différences se situent principalement entre les coefficients d'émission de 1985 et les autres. Au niveau de l'impact de la motorisation dans la mobilité des ménages pour le motif travail, on observe, pour les ménages ayant un seul actif, que le fait d'être motorisé a un impact différent selon les années d'enquête. L'impact semble être le plus fort en 1995 puisque l'écart entre la mobilité des ménages non motorisés et ceux ayant un véhicule est d'environ 0,4 demi-sorties, soit environ 25 %, alors qu'en 1985 et 2006 il n'est que d'entre 0,1 et 0,15 demi-sorties (autour de 8 %). Pour les ménages à deux actifs ou plus, l'accès à la voiture conduit à une hausse de la mobilité relativement importante en 1985 et surtout en 1995 (augmentation d'entre 0,5 et 0,6 demi-sorties, ce qui représente une augmentation d'environ 14 % en 1985 et de 20 % en 1995). En 2006, l'impact de l'accès à la motorisation est très limité (moins de 1 %), l'accès au deuxième véhicule ayant un impact un peu plus important. Ces résultats sont consistants avec la baisse du pouvoir de discrimination de la motorisation (Kitamura et Kostyniuk, 1986). Ils doivent cependant être tempérés par l'importance des écarts-types pour les ménages non motorisés à deux actifs.

La table 4.3 donne l'évolution des coefficients d'émission de demi-sorties pour le motif travail, suivant une classification selon le nombre de personnes du ménage déclarant avoir un emploi et le stade dans le cycle de vie du ménage. Le premier élément que l'on peut remarquer sur ce tableau concerne l'existence de deux classes pour lesquelles le nombre d'observations est inférieur à 30 (sur certaines années et périmètres). Il s'agit des ménages ayant deux actifs ou plus, et se trouvant au stade 1 ou 7 du cycle de vie — ménages pour lesquels il n'y a pas de lien conjugal entre les membres ni de présence d'enfants. Il s'agit donc de ménages de type « colocation » de plusieurs actifs. En termes d'évolutions temporelles, pour les classes avec un effectif « suffisant » (au moins 30 observations) l'analyse qui peut être faite ressemble à celle réalisée pour la classification faisant intervenir `nbperstrav_F` et `nbvppdispo_F`. Dans la plupart des cas, on observe une baisse continue de la mobilité entre 1985 et 2006, à périmètre constant (celui de 1985). L'élargissement du périmètre d'analyse en 2006 se traduit par une augmentation légère du nombre de demi-sorties. Pour les classes avec des écarts-types des coefficients d'émission relativement faibles, la baisse généralisée de la mobilité se traduit par un rejet de l'hypothèse de stabilité. Il est toutefois nécessaire de compléter cette perspective d'ensemble par l'analyse de certains cas particuliers. Le premier concerne les ménages à un actif se trouvant au stade 1 du

TABLE 4.3 – Analyse de la stabilité temporelle du coefficient d'émission de demi-sorties par ménage, pour le motif Travail. Classification réalisée à partir de deux variables : nombre de personnes déclarant avoir un emploi et cycle de vie.

Modalité	Coefficient d'émission moyen ( $\mu$ ) et écart-type ( $se$ )												IC-test : différence significative entre $\mu$							
	1985/P1985 (a)		1995/P1985 (b)		1995/P1995 (c)		2006/P1985 (d)		2006/P1995 (e)		2006/AUL (f)		a	b	c	d	e	f		
Nb pers trav	$\mu_a$	$se_a$	$\mu_b$	$se_b$	$\mu_c$	$se_c$	$\mu_d$	$se_d$	$\mu_e$	$se_e$	$\mu_f$	$se_f$								
1 Personne																				
1	1,80	0,07	1,77	0,07	1,77	0,07	1,75	0,06	1,77	0,06	1,82	0,05								
2	2,55	0,19	1,91	0,17	1,91	0,16	1,77	0,15	1,81	0,15	1,86	0,14	def				a	a	a	
3	1,99	0,10	2,03	0,10	2,09	0,10	1,68	0,09	1,75	0,08	1,74	0,07			df		c	c	c	
4	2,22	0,12	1,99	0,10	2,00	0,10	1,87	0,11	1,88	0,11	1,90	0,09								
5	2,20	0,09	2,16	0,08	2,11	0,08	1,90	0,08	1,91	0,08	1,94	0,07								
6	2,10	0,11	2,04	0,11	1,98	0,11	1,63	0,10	1,65	0,10	1,66	0,10		def	df		ab	ab	a	
7	1,85	0,12	1,92	0,09	1,90	0,09	1,63	0,08	1,65	0,07	1,69	0,07							ab	
2 Personnes et +																				
1	3,57	0,47	3,39	0,43	3,39	0,43	3,71	0,49	3,73	0,49	3,67	0,42								a
2	3,66	0,16	3,72	0,14	3,78	0,13	3,46	0,13	3,52	0,13	3,53	0,12								
3	3,73	0,15	3,45	0,12	3,45	0,12	2,92	0,09	2,98	0,09	3,04	0,07	def	def	def	def	abc	abc	abc	abc
4	4,05	0,15	3,47	0,14	3,42	0,13	3,34	0,11	3,40	0,11	3,38	0,09	def	def	a	a	a	a	a	a
5	4,79	0,12	4,13	0,10	4,09	0,10	3,98	0,10	4,00	0,10	4,04	0,08	bedef	a	a	a	a	a	a	a
6	3,93	0,19	3,71	0,20	3,72	0,19	3,43	0,16	3,47	0,15	3,54	0,12								
7	2,84	0,27	3,49	0,52	3,49	0,52	5,03	0,95	4,71	0,68	4,31	0,57	e							

Note de lecture : voir table 4.2. Les cases surhignées en gris correspondent à un effectif observé inférieur à 30 ménages. La nomenclature de la variable Cycle de vie peut être trouvée dans l'annexe B (page 214).

Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006.

cycle de vie. Pour cette classe, on observe une grande stabilité des comportements. En effet, alors que les écarts types des coefficients d'émission sont les plus faibles parmi toutes les classes obtenues dans ce découpage (ils correspondent à moins de 4 % des coefficients d'émission) le test par intervalles de confiance ne permet de rejeter l'hypothèse de stabilité dans aucune des comparaisons possibles (selon les années et périmètres d'enquête). On observe une petite baisse de la mobilité entre 1985 et 2005 à périmètre constant (moins de 3 %) alors qu'en prenant en compte toute l'information pour 2006 on a une augmentation d'environ 1 %. Notons que cette classe est essentiellement constituée des actifs vivant seuls, pour lesquels il est donc possible de dire que les comportements de mobilité n'ont presque pas changé sur la période étudiée. Le deuxième cas particulier correspond aux ménages à un actif se trouvant au stade 2 du cycle de vie. Cette classe est intéressante car malgré le fait que les écarts-types des coefficients d'émission sont plutôt élevés, l'hypothèse de stabilité est rejetée pour la comparaison entre 1985 et 2006. Autrement dit, la baisse de la mobilité des couples sans enfant où une seule personne travaille est très importante, ce qui peut suggérer une diminution de l'influence de la présence d'une personne au foyer dans la décision de retourner au domicile déjeuner. Dans ce sens, en 1985 on observe une augmentation du coefficient d'émission d'environ 42 % lorsque l'on compare les ménages à un actif au stade 1 du cycle de vie aux ménages au stade 2. En 2006, sur le périmètre de 1985, l'écart n'est que de 1 %. Compte tenu des écarts-types observés, on peut dire que la différence est significative en 1985 et qu'elle ne l'est pas en 2006.

**4.2.1.2.2 Les motifs autres que Travail et Enseignement** La table 4.4 contient l'évolution des coefficients d'émission de demi-sorties pour les motifs autres que travail et enseignement, suivant une classification selon le nombre de personnes du ménage et le nombre de véhicules disponibles. Compte tenu du nombre de modalités plus important de la variable nbpers\_F par rapport à la variable nbperstrav\_F utilisée pour le motif travail, certaines catégories ont un effectif inférieur à 30. Il s'agit, en particulier, des ménages à une personne ayant plusieurs véhicules à disposition (3 ou plus). Par ailleurs, le nombre élevé de classes associé à une variabilité des comportements plus importante que pour le motif travail (liée en partie au fait que tous les ménages sont concernés par ce motif de déplacement et non seulement les ménages avec des actifs) conduit à avoir des écarts-types relativement importants. En matière d'évolutions temporelles, on a, pour la plupart des catégories, une augmentation de la mobilité entre 1985 et 1995 puis une baisse de celle-ci entre 1995 et 2006, à périmètre constant. Cette baisse peut être supérieure ou inférieure à l'augmentation de la mobilité entre 1985 et 1995. Le premier cas est néanmoins prépondérant, tant en termes de nombre de classes concernées qu'en termes d'importance de chaque classe (rappelons que l'on observe une diminution de la taille des ménages et une augmentation de la motorisation de ces derniers, comme on l'a évoqué dans la section 3.2.2). Ainsi, par exemple, on observe chez les personnes seules non motorisées, une augmentation de la mobilité entre 1985 et 1995 d'environ 15 % puis une baisse d'environ 9 % entre 1995 et 2006. Pour les ménages à deux personnes et un véhicule disponible, on observe une augmentation de la mobilité d'environ 14 % entre 1985 et 1995 puis une diminution d'environ 5 % entre 1995 et 2006, à périmètre constant. L'élargissement du périmètre d'analyse n'a en général que très peu d'impact sur les coefficients d'émission, ceux-ci

TABLE 4.4 – Analyse de la stabilité temporelle du coefficient d'émission de demi-sorties par ménage, pour les motifs autres que Travail et Enseignement. Classification réalisée à partir de deux variables : nombre de personnes et nombre de véhicules disponibles.

Modalité	Coefficient d'émission moyen ( $\mu$ ) et écart-type ( $se$ )												IC-test : différence significative entre $\mu$					
	1985/P1985 (a)		1995/P1985 (b)		1995/P1995 (c)		2006/P1985 (d)		2006/P1995 (e)		2006/AUL (f)		a	b	c	d	e	f
Nb pers VP dispo	$\mu_a$	$se_a$	$\mu_b$	$se_b$	$\mu_c$	$se_c$	$\mu_d$	$se_d$	$\mu_e$	$se_e$	$\mu_f$	$se_f$						
1 Personne																		
0	1,66	0,07	1,91	0,07	1,90	0,07	1,80	0,07	1,80	0,07	1,81	0,07						
1	1,59	0,09	1,65	0,09	1,65	0,09	1,68	0,07	1,70	0,07	1,75	0,06						
2	1,10	0,49	0,91	0,24	1,06	0,27	1,31	0,32	1,39	0,31	1,38	0,27						
3+	1,74	1,03	2,11	0,42	2,11	0,42	3,32	1,29	3,32	1,29	1,75	0,87						
2 Personnes																		
0	2,87	0,18	2,87	0,20	2,88	0,20	3,27	0,19	3,31	0,19	3,27	0,18						
1	3,28	0,12	3,73	0,12	3,75	0,12	3,55	0,11	3,56	0,11	3,62	0,10	c					
2	2,41	0,19	3,15	0,15	3,18	0,15	3,59	0,15	3,58	0,14	3,54	0,11	bcdef					
3+	3,68	1,25	3,06	0,55	3,04	0,54	2,75	0,61	2,70	0,55	3,26	0,40	a					
3 Personnes																		
0	4,23	0,38	4,39	0,42	4,39	0,42	3,22	0,45	3,21	0,44	3,22	0,43						
1	3,68	0,18	4,29	0,25	4,33	0,25	3,76	0,20	3,81	0,20	3,96	0,19						
2	3,03	0,21	3,76	0,21	3,76	0,21	3,46	0,19	3,40	0,18	3,32	0,15						
3+	3,66	0,45	3,96	0,40	4,02	0,38	3,84	0,40	3,87	0,36	3,87	0,29						
4 Personnes																		
0	4,77	0,64	3,60	0,57	3,60	0,57	4,30	0,69	4,30	0,69	4,31	0,64						
1	4,43	0,24	5,37	0,30	5,36	0,30	4,34	0,26	4,32	0,25	4,41	0,23						
2	4,38	0,26	4,79	0,27	4,77	0,25	3,99	0,22	4,04	0,20	4,31	0,17						
3+	4,20	0,45	5,50	0,56	5,66	0,60	4,57	0,43	4,74	0,41	4,63	0,31						
5 Personnes et +																		
0	6,36	0,80	5,68	0,81	5,78	0,78	5,95	0,80	5,95	0,80	5,86	0,75						
1	6,47	0,45	6,54	0,39	6,76	0,40	6,52	0,36	6,51	0,35	6,51	0,34						
2	6,12	0,48	7,12	0,42	7,11	0,41	7,63	0,51	7,42	0,48	7,32	0,36						
3+	5,21	0,75	8,39	0,71	8,23	0,69	5,61	0,56	6,07	0,57	6,75	0,45	bc ad ad bc					

Note de lecture : voir table 4.2. Les cases surlignées en gris correspondent à un effectif observé inférieur à 30 ménages.

Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006.

demeurant à peu près constants. Il convient de noter deux cas particuliers pour lesquels la mobilité augmente avec le temps, concernant les ménages à taille inférieure ou égale à deux personnes. Le premier cas est celui des personnes seules disposant d'un véhicule. Pour cette catégorie, on observe une hausse très légère au cours du temps et avec l'élargissement du périmètre d'analyse (les coefficients d'émission sont entre 6 et 10 % plus élevés en 1995 et 2006 que ceux de 1985). Le deuxième cas est celui des ménages à deux personnes ayant deux véhicules à disposition. Pour ces ménages, on constate une augmentation bien plus importante (hausse d'environ 30 % entre 1985 et 1995 et d'environ 14 % entre 1995 et 2006), à périmètre constant. L'élargissement du périmètre d'analyse ne semble pas avoir d'impact pour cette catégorie. L'augmentation de la mobilité au cours du temps pour cette catégorie est d'autant plus notoire qu'elle est l'une des rares (3/20) pour lesquelles les IC-tests permettent de rejeter l'hypothèse de stabilité des coefficients. Celle-ci est également rejetée entre 1995 et les autres années (à périmètre constant) pour les ménages constitués de 5 personnes ou plus et ayant 3 ou plus véhicules à disposition. L'écart entre les coefficients de 1995 et les autres est de plus de 2,5 demi-sorties, ou encore près de 60 % de la mobilité de cette catégorie de ménages en 1985 et de 50 % en 2006. Notons que, pour cette catégorie de ménages, l'élargissement du périmètre d'analyse en 2006 se traduit par une augmentation de la mobilité d'environ une demi-sortie, ce qui rend l'écart avec les coefficients de 1995 non significatif. Le dernier cas de rejet de l'hypothèse de stabilité correspond à celui des ménages à deux personnes et un véhicule, mais seulement lorsqu'on compare le coefficient de 1985 à celui de 1995 sur le périmètre de 1995. L'écart entre la valeur de ces paramètres est de 14 %, ce qui est peu en comparaison des autres écarts donnant lieu au rejet de l'hypothèse de stabilité. Si l'on s'intéresse à l'impact qu'a le niveau de motorisation pour les différentes tailles de ménages, en termes de mobilité, on observe des cas contrastés. Pour les personnes vivant seules, l'accès à la voiture a tendance à réduire un peu la mobilité, sur l'ensemble des années et périmètres étudiés. Une explication possible de cette baisse peut être trouvée dans la facilité accrue, avec la voiture, d'enchaîner les déplacements (notamment les déplacements pour le motif achats-services pour lequel Doubleday, 1977, remarquait que la disponibilité de la voiture peut induire des parcours d'achats successifs dans la mesure où elle donne une capacité de charge importante, que les piétons n'ont pas). Une autre explication possible concerne le regroupement hebdomadaire au détriment des achats quotidiens. Un cas de figure similaire peut être observé pour les ménages à 3 personnes, en 1985 et 1995. Pour tous les autres cas, l'accès à la voiture augmente le niveau de mobilité. Lorsque les ménages sont motorisés, ceux qui disposent d'un seul véhicule ont des coefficients d'émission plus importants que ceux qui disposent de deux véhicules. Pour les ménages ayant 3 véhicules ou plus, la mobilité peut être plus élevée que celle des ménages à 2 véhicules, mais elle peut aussi être moins élevée.

La table 4.5 contient l'évolution des coefficients d'émission de demi-sorties pour les motifs autres que travail et enseignement, suivant une classification selon le nombre de personnes du ménage et le cycle de vie. Comme dans la classification précédente, le nombre important de modalités des deux variables retenues conduit à avoir certaines classes avec un effectif très faible, voir nul. Par ailleurs, il y a d'autres cas où l'effectif est nul par construction (incompatibilité des modalités) ; ces classes n'ont pas été incluses dans la table (on peut donner l'exemple des ménages à une

personne se trouvant au stade 2 du cycle de vie, qui implique un lien conjugal et donc la présence d'une personne supplémentaire). Les ménages ayant 3 personnes ou plus correspondent très majoritairement à des ménages où il y a un lien de parenté (ménages avec des parents et leurs enfants). Ainsi, les classes correspondant aux autres stades de vie ont un effectif observé insuffisant. Un autre cas de classe avec un effectif faible est celui des ménages à deux personnes où le lien parental existe et l'enfant a moins de 5 ans. En termes d'évolutions temporelles, la majorité des classes connaît une augmentation de la mobilité entre 1985 et 1995 (entre 3 et 60 %) puis une diminution de celle-ci entre 1995 et 2006 (entre 3 et 46 %), à périmètre constant. Certaines classes connaissant des tendances un peu différentes doivent être évoquées de manière particulière, compte tenu des évolutions démographiques observées dans la section 3.2.2. D'abord, les ménages uni-personnels « jeunes » connaissent une diminution de la mobilité à un taux relativement constant entre 1985 et 2006 (diminution d'environ 7 % tous les 10 ans). Puis, les ménages uni-personnels plutôt âgés connaissent, au contraire, une augmentation de leur mobilité quotidienne. Celle-ci augmente d'environ 22 % entre 1985 et 1995 pour ensuite rester à peu près constante entre 1995 et 2006. Finalement, les couples un peu âgés sans enfant dans le foyer connaissent une augmentation de 19 % entre 1985 et 1995 puis une augmentation de 4 % entre 1995 et 2006. Au niveau du test de stabilité des coefficients, si on écarte l'ensemble des classes pour lesquelles l'effectif observé est insuffisant et les classes avec des écarts-types très élevés (supérieurs à 0,2), on observe seulement deux cas où l'hypothèse de stabilité est rejetée : celui des ménages constitués de personnes plutôt âgées vivant seules (l'hypothèse de stabilité est rejetée entre 1985 et les autres années) et celui des couples plutôt âgés sans enfants (le rejet de l'hypothèse de stabilité se fait également entre 1985 et les autres années). Au niveau des différences de mobilité selon le stade dans le cycles de vie et la taille du ménage, on peut observer, pour les ménages sans enfant, que l'âge de la personne de référence joue un rôle important. En effet, pour les ménages uni-personnels, les coefficients d'émission des ménages au stade 7 du cycle de vie sont entre 17 et 62 % plus élevés que ceux des ménages au stade 1 (l'écart étant croissant avec le temps). Pour les ménages à deux personnes, lorsqu'il n'y a pas de lien conjugal, l'écart entre la mobilité des ménages ayant une personne de référence « plutôt jeune » et ceux ayant une personne de référence « plutôt âgée » est comprise entre 22 et 97 % (l'écart le plus faible étant constaté en 1995 et le plus élevé en 2006). Lorsque le lien conjugal est présent, les écarts de mobilité entre les ménages ayant une personne de référence « plutôt jeune » et ceux ayant une personne de référence « plutôt âgée » vont d'environ 52 % en 1985 à environ 93 % en 2006. À taille de ménage égale, la mobilité plus importante des ménages ayant une personne de référence « plutôt âgée » peut trouver une explication dans le fait que certains de ces ménages sont constitués, de manière totale ou partielle, de personnes à la retraite. La présence d'actifs ayant un emploi dans le ménage fait qu'une partie des déplacements autres que le travail et l'enseignement sont réalisés au sein de sorties ayant le travail comme motif principal. Au niveau des ménages ayant des enfants, le classement en fonction de l'âge du benjamin ne donne pas des résultats stables au cours du temps en termes de mobilité. Certaines classes sont les plus mobiles à une date et elles ne le sont plus à une autre. Une dernière remarque peut être faite vis-à-vis de l'impact des liens conjugaux et parentaux au sein des ménages, en termes de mobilité : on se serait attendu à ce que la coordination des activités favorisée par

TABLE 4.5 – Analyse de la stabilité temporelle du coefficient d'émission de demi-sorties par ménage, pour les motifs autres que Travail et Enseignement. Classification réalisée à partir de deux variables : nombre de personnes et cycle de vie.

Modalité	Coefficient d'émission moyen ( $\mu$ ) et écart-type ( $se$ )										IC-test : différence significative entre $\mu$						
	1985/P1985 (a)	1995/P1985 (b)	1995/P1995 (c)	2006/P1985 (d)	2006/P1995 (e)	2006/AUL (f)	a	b	c	d	e	f					
Nb pers Cycle de vie	$\mu_a$	$se_a$	$\mu_b$	$se_b$	$\mu_c$	$se_c$	$\mu_d$	$se_d$	$\mu_e$	$se_e$	$\mu_f$	$se_f$					
1 Personne	1,47	0,09	1,35	0,08	1,36	0,09	1,26	0,07	1,26	0,07	1,29	0,07					
7	1,72	0,07	2,09	0,07	2,07	0,07	2,03	0,07	2,04	0,07	2,05	0,06	bcdef	a	a	a	a
2 Personnes	1	2,16	0,40	2,36	0,29	2,36	0,29	1,34	0,25	1,45	0,27	0,27					
2	2,25	0,17	2,51	0,17	2,53	0,17	2,20	0,16	2,20	0,15	2,18	0,14					
3	1,51	0,35	2,39	0,75	2,60	0,75	2,50	0,54	2,57	0,52	2,64	0,47					
4	1,59	0,38	2,55	0,36	2,55	0,36	2,19	0,36	2,14	0,35	2,35	0,32					
5	2,53	0,23	2,68	0,23	2,70	0,23	2,64	0,22	2,74	0,22	2,66	0,20					
6	3,44	0,12	4,10	0,11	4,12	0,11	4,25	0,10	4,23	0,10	4,19	0,08	bcdef	a	a	a	a
7	3,51	0,55	2,88	0,44	2,92	0,43	2,65	0,61	2,45	0,56	2,55	0,45					
3 Personnes	1	3,17	0,76	6,01	1,28	1,28	1,84	0,98	1,84	0,98	1,84	0,98					
2	5,57	1,65	11,35	2,62	11,35	2,62	3,20	0,40	3,20	0,40	3,75	0,52		def	def	bc	bc
3	3,31	0,24	3,10	0,24	3,09	0,23	3,13	0,21	3,05	0,20	2,99	0,18					
4	3,02	0,27	4,56	0,34	4,58	0,33	4,14	0,34	4,05	0,32	4,08	0,27		a	a		
5	3,66	0,18	4,22	0,20	4,27	0,19	3,70	0,18	3,80	0,17	3,82	0,15	bc				
6	4,51	0,61	5,11	0,85	5,11	0,85	4,96	0,70	4,96	0,70	4,58	0,84					
7	3,98	1,15	14,00	0,00	14,00	0,00	3,33	2,10	1,74	1,54	4,33	2,40	bc	adef	adef	bc	bc
4 Personnes	1	4,00	0,00	4,00	0,00	4,00	2,01	0,72	2,01	0,72	2,42	0,79	def				
2	4,38	0,31	4,38	0,30	4,45	0,29	5,98	1,94	5,98	1,94	4,30	2,01		b	a	a	a
3	4,78	0,39	5,89	0,45	6,01	0,44	4,05	0,25	3,93	0,23	4,04	0,19					
4	4,24	0,21	5,04	0,25	4,89	0,24	4,78	0,37	4,99	0,36	5,10	0,29					
5							3,92	0,21	3,99	0,20	4,22	0,18		de	de	bc	bc
6	9,08	2,62	3,09	1,75	3,09	1,75	11,57	0,48	11,57	0,48	11,57	0,48		def	def	bc	bc
7							9,90	1,41	9,90	1,41	7,23	2,34					
5 Personnes et +	1	2,00	0,00	16,00	0,00	0,00	6,00	0,00	6,00	0,00	6,00	0,00	bcdef	adef	adef	abc	abc
2																	
3	6,61	0,44	6,84	0,42	6,93	0,42	6,52	0,36	6,46	0,35	6,68	0,30					
4	5,20	0,95	8,26	0,79	8,52	0,84	9,21	1,18	8,51	1,07	8,16	0,82					
5	6,09	0,40	6,64	0,35	6,67	0,33	6,50	0,37	6,55	0,36	6,74	0,32					
6							2,00	0,00	2,00	0,00	2,00	0,00					
7																	

Note de lecture : voir table 4.2. Les cases surlignées en gris correspondent à un effectif observé inférieur à 30 ménages. La nomenclature de la variable Cycle de vie peut être trouvée dans l'annexe B (page 214).

Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006.

ces liens conduise à observer une mobilité moins importante pour les ménages où il sont présents. Or, on constate que les ménages à deux personnes sans enfants sont plus mobiles lorsque le lien conjugal existe que lorsqu'il n'est pas présent. Pour les ménages avec des enfants, on se serait attendu à voir une mobilité plus importante avec l'âge du benjamin (indépendance accrue permettant la réalisation d'activités supplémentaires) mais ce n'est pas le cas : le plus souvent, ce sont les ménages au stade 4 qui présentent une mobilité plus élevée.

#### **4.2.2 Les modèles de classification croisée calibrés en 1985 et 1995 permettent-ils de prédire convenablement les comportements de mobilité de 2006 ?**

Les tests d'égalité des paramètres de modèles estimés à différentes dates suggèrent que ceux-ci ne sont globalement pas transférables dans le temps. Il convient néanmoins de compléter cette première approche par une étude de l'impact que la différence constatée entre les coefficients d'émission des ménages a en termes de capacité prédictive, dans la mesure où celui-ci est l'objectif opérationnel des modèles étudiés ici. Les tables 4.6 et 4.7 donnent les résultats des prévisions des émissions de 2006, réalisées en utilisant les paramètres estimés avec les données de 1985 et 1995. La table 4.6 donne les résultats de ces prévisions à périmètre constant — les estimations ont été réalisées avec les données correspondant au périmètre de l'EMD de 1985 — alors que la table 4.7 est réalisée à partir de l'ensemble des données disponibles lors de chaque enquête (dans la limite de l'aire urbaine lyonnaise)<sup>16</sup>.

Si l'on s'intéresse au motif travail, on observe que les paramètres estimés avec les données de 1995 prédisent mieux la mobilité de 2006 que ceux estimés en 1985, à tous les niveaux d'analyse retenus. Au niveau de la mobilité totale, on constate que l'hypothèse de stabilité des coefficients d'émission conduit à une surestimation, pour les paramètres estimés avec les données de 1985 comme pour ceux estimés avec les données de 1995. On retrouve donc les conséquences des évolutions constatées dans la table 4.1. La forte baisse du nombre de sorties « simples » entre 1985 et 1995 fait que l'utilisation des coefficients calculés en 1985 pour prévoir la mobilité de 2006 se traduise par une forte surestimation (au point de les rendre inutilisables d'un point de vue opérationnel). En 1995, une classification à partir d'une seule variable, le cycle de vie, est celle qui permet de mieux reproduire la mobilité totale de 2006. En 1985, il s'agit de la classification utilisant conjointement le nombre de personnes et le cycle de vie des ménages. Cependant, il convient de noter que, lors des prévisions, on a attribué une mobilité nulle aux non observées en 1985 et 1995, ce qui a pu légèrement influencer les prévisions à la baisse. L'analyse de la mobilité totale pour le motif travail fait apparaître la perte de pouvoir explicatif du niveau de motorisation des ménages, résultat qui fait écho aux travaux de Kitamura et Kostyniuk (1986), puisque cette variable conduit aux pires prévisions en 1985 et 1995. À l'échelle des ménages (*MAE*), on observe que, pour les classifications réalisées à partir seulement du nombre de véhicules disponibles ou du cycle de vie, les comportements de mobilité sont moins bien reproduits par les coefficients calculés en 1985 et en 1995 que par ceux calculés

---

16. L'indicateur *EREL* (voir section 4.1.2.4) n'a naturellement pas été fourni pour 2006 car, par construction, il est nul pour toutes les classifications.

en 2006. Les coefficients de 1995 sont à nouveau meilleurs dans la prévision que ceux de 1985. Pour les classifications réalisées avec seulement le nombre de personnes déclarant avoir un emploi ou en combinant cette variable avec le nombre de véhicules disponibles, on observe que les coefficients de 1985 et 1995 permettent de mieux reproduire les données de 2006 que les paramètres de 2006. On est confrontés ici à une des limites du modèle de classification : l'hypothèse de normalité du nombre de demi-sorties par ménage semble remise en cause. La distribution du nombre de demi-sorties n'est pas symétrique par rapport à la moyenne et de ce fait, l'indicateur *MAE* peut être meilleur pour des coefficients étant différents de la moyenne (donnée par les coefficients de 2006). Notons que les paramètres de 1995 donnent à nouveau de meilleurs résultats que ceux de 1985. Pour la dernière classification, faisant intervenir le nombre de personnes ayant un travail et le cycle de vie, les coefficients de 1985 donnent de moins bons résultats que ceux de 2006 alors que ceux de 1995 donnent des résultats meilleurs. À l'échelle zonale, on retrouve une meilleure performance prédictive des coefficients de 1995. Comme précédemment, le nombre de véhicules disponibles semble conduire aux pires résultats en termes explicatifs et prédictifs. Les meilleurs résultats sont obtenus en utilisant le nombre de personnes ayant un travail (variable seule ou combinée avec le cycle de vie). Les prévisions de 1985 sont marginalement meilleures lorsqu'on inclut le cycle de vie dans la classification. Si on s'intéresse à l'erreur supplémentaire induite par l'utilisation des paramètres de 1985 et de 1995 par rapport à l'utilisation des paramètres de 2006 (tirant parti de la linéarité du *SRMSEG*), on constate que les prévisions avec les coefficients de 1995 ont des erreurs entre 14 et 38 % plus importantes (en excluant la classification avec seulement le nombre de véhicules disponibles) que celles obtenues avec les paramètres de 2006. On note que les *SRMSEG* restent acceptables pour les meilleures classifications, l'écart de 38 % correspondant, en réalité, aux meilleures prévisions (*SRMSEG* le plus faible). Les erreurs induites par l'utilisation des paramètres de 1985 sont environ égales au double de celles induites par les paramètres de 1995.

Pour les motifs autres que travail et enseignement, les résultats sont plus contrastés que pour le motif travail. Au niveau du nombre total de demi-sorties, l'ensemble des classifications pour 1985 et 1995 conduit à une erreur en prévision inférieure à 10 %. La mobilité pour ce motif est donc mieux reproduite que celle pour le motif travail. Les coefficients de 1995 surestiment la mobilité totale, alors que ceux de 1985 la sous-estiment, à l'exception de la classification à partir du nombre de véhicules disponibles qui la surestime très légèrement (moins de 1 %). Les erreurs obtenues avec les coefficients de 1995 sont, en majorité, inférieures en valeur absolue à celles obtenues avec les coefficients de 1985. Le rôle joué par le nombre de véhicules disponibles est complexe : en 1985, les meilleurs résultats en prévision sont obtenus à partir de la classification utilisant uniquement cette variable alors que les résultats les moins bons proviennent de la classification l'associant au nombre de personnes. En 1995, on a la configuration inverse : la meilleure prévision de la mobilité totale est faite par la classification utilisant le nombre de personnes et le nombre de véhicules disponibles alors que les résultats les moins bons sont obtenus avec la classification utilisant uniquement le nombre de véhicules. À l'échelle des ménages, les coefficients de 1985 donnent systématiquement de meilleurs résultats en prévision que ceux de 1995. Les coefficients de 2006 reproduisent mieux les données de 2006 que ceux de 1985 et 1995. Les meilleurs résultats en prévision sont donnés par la classification faisant intervenir

TABLE 4.6 – Prévision de la mobilité de 2006 dans le périmètre de l'EMD 1985 avec les coefficients estimés sur ce même périmètre en 1985 et 1995.

Motif Variable	<i>EREL</i>		<i>MAE</i>			<i>SRMSEG</i>		
	1985	1995	1985	1995	2006	1985	1995	2006
<i>p = Travail</i>								
nbperstrav_F	18,80	7,68	0,79	0,74	0,80	0,22	0,11	0,08
nbvpdispo_F	38,81	12,87	1,53	1,43	1,42	0,46	0,24	0,19
cycledevie	18,64	6,57	1,35	1,33	1,27	0,25	0,16	0,14
nbperstrav_F et nbvpdispo_F	19,99	7,67	0,80	0,75	0,81	0,23	0,12	0,09
nbperstrav_F et cycledevie	16,78	7,42	0,81	0,75	0,78	0,20	0,11	0,08
<i>p = Autres que Travail et Enseignement</i>								
nbpers_F	-7,17	4,50	2,35	2,36	2,33	0,16	0,14	0,13
nbvpdispo_F	0,68	7,27	2,55	2,57	2,52	0,15	0,16	0,15
cycledevie	-7,03	5,61	2,34	2,36	2,32	0,15	0,14	0,12
nbpers_F et nbvpdispo_F	-8,57	4,37	2,35	2,36	2,32	0,17	0,14	0,13
nbpers_F et cycledevie	-7,66	5,85	2,29	2,30	2,24	0,15	0,14	0,12

*Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006.*

TABLE 4.7 – Prévision de la mobilité de 2006 sur l'ensemble de l'aire urbaine avec les coefficients estimés avec l'ensemble de données de 1985 et 1995.

Motif Variable	<i>EREL</i>		<i>MAE</i>			<i>SRMSEG</i>		
	1985	1995	1985	1995	2006	1985	1995	2006
<i>p = Travail</i>								
nbperstrav_F	16,58	5,13	0,80	0,76	0,81	0,20	0,10	0,09
nbvpdispo_F	37,40	11,75	1,57	1,47	1,46	0,45	0,22	0,19
cycledevie	13,17	1,37	1,37	1,35	1,32	0,20	0,14	0,14
nbperstrav_F et nbvpdispo_F	18,93	5,70	0,82	0,77	0,81	0,23	0,11	0,10
nbperstrav_F et cycledevie	14,67	4,94	0,82	0,77	0,79	0,18	0,10	0,09
<i>p = Autres que Travail et Enseignement</i>								
nbpers_F	-8,51	3,55	2,44	2,45	2,42	0,17	0,14	0,14
nbvpdispo_F	-2,05	6,23	2,62	2,65	2,61	0,16	0,16	0,15
cycledevie	-8,44	4,91	2,43	2,46	2,43	0,17	0,14	0,13
nbpers_F et nbvpdispo_F	-10,52	3,25	2,44	2,45	2,42	0,19	0,14	0,13
nbpers_F et cycledevie	-8,54	5,54	2,38	2,39	2,34	0,17	0,14	0,13

*Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006.*

le nombre de personnes et le cycle de vie, en 1985 comme en 1995, même s'il convient de noter que les écarts dans la qualité de prévision des différentes classifications sont relativement faibles. Si on compare les ordres de grandeur de l'indicateur *MAE* avec ceux pour le motif travail, on remarque une grande différence (au moins une demi-sortie d'écart). Celle-ci peut trouver une explication dans le fait qu'il y a une plus grande variabilité au niveau de la mobilité des ménages pour ce motif que pour le motif travail. On pouvait s'y attendre dans la mesure où l'ensemble des ménages est concerné par ce motif alors que seulement certains ménages sont concernés par le motif

travail, ce qui se traduit par une variabilité plus importante car, pour les ménages non concernés par le motif travail la mobilité pour ce motif est constante et nulle (ou quasi-constante et quasi-nulle, si l'on prend en compte les étudiants ne se déclarant pas comme activement employés mais ayant une mobilité non nulle pour le motif travail). Au niveau de la mobilité par zone, on trouve des résultats qui contrastent avec ceux de la mobilité totale. En effet, alors que les coefficients de 1985 et ceux de 1995 permettent de mieux reproduire la mobilité totale pour ce motif que pour le motif travail, au niveau de la mobilité par zone, on a des erreurs plus importantes que celles constatées pour le motif travail. Il y a donc des effets de compensation entre certaines zones pour lesquelles la mobilité est surestimée et d'autres où elle est sous-estimée. Le nombre de véhicules disponibles n'est plus la variable permettant de faire la meilleure classification. Il s'agit désormais de la combinaison entre le nombre de personnes et le cycle de vie. Celle-ci permet de faire la meilleure reproduction des données observées en 2006, tout comme des prévisions d'égale ou de meilleure qualité que les autres classifications, à la fois en 1985 et en 2006. Un autre contraste avec les résultats pour le motif travail est que, même si la qualité des prévisions est meilleure avec les coefficients de 1995 qu'avec ceux de 1985 (à l'exception de la classification à partir du nombre de véhicules disponibles), l'écart est nettement moins marqué puisqu'il n'est que de 21 % au maximum.

La prise en compte de l'ensemble des données — demi-sorties réalisées sur l'ensemble des périmètres des différentes enquêtes (table 4.7) — donne des résultats relativement similaires à ceux observés à périmètre constant, à condition de prendre en compte l'augmentation de la mobilité de 2006 occasionnée par l'élargissement du périmètre à l'aire urbaine. Ainsi, au niveau de la mobilité totale, les prévisions pour le motif travail suivent les mêmes tendances qui ont été décrites jusqu'ici, en termes de classement des classifications retenues en fonction de leur performance prédictive, mais on trouve une amélioration générale des résultats, provenant du fait que l'on prévoit une mobilité accrue (que l'on surestime donc en moindre mesure). Parallèlement, l'augmentation de la mobilité, pour les motifs autres que travail et enseignement se traduit par une amélioration des performances prédictives en 1995 (puisque la mobilité était surestimée) et une détérioration des performances des coefficients de 1985 (toutes les classifications, désormais, sous-estiment, de manière plus importante, la mobilité de 2006). Au niveau des ménages, on remarque une augmentation de la variabilité des comportements en 2006. L'erreur absolue moyenne augmente pour l'ensemble des motifs, des classifications et des coefficients. Il convient de noter que l'augmentation de la variabilité est moins forte pour le motif travail que pour les motifs autres que travail et enseignement. Au niveau zonal, l'augmentation de la variabilité des comportements se traduit par une très légère baisse de la performance descriptive des coefficients de 2006, visible dans la légère augmentation du *SRMSEG* pour 2006, sur l'ensemble des motifs. L'augmentation de la mobilité en 2006 a, en revanche un impact différent sur les prévisions pour le motif travail et les motifs autres. Comme pour la mobilité totale, la performance prospective des coefficients de 1985 et 1995 s'améliore pour le motif travail. Pour les motifs autres que travail et enseignement, elle se détériore pour les coefficients de 1985 sans qu'il y ait de changement visible en 1995.

## 4.3 Le modèle de génération par régression zonale

### 4.3.1 Stabilité des coefficients

La première étape de l'analyse des modèles de régression zonale consiste à déterminer la meilleure spécification du modèle, pour ensuite voir si les coefficients sont stables dans le temps.

#### 4.3.1.1 Spécification du modèle

**4.3.1.1.1 Motif travail** Un des désavantages reconnus de l'approche zonale de la régression est que le nombre de variables explicatives est en général relativement faible. En effet, il est difficile d'agréger certaines informations dont on dispose à l'échelle individuelle ou des ménages pour les retranscrire au niveau zonal. Cette difficulté d'agrégation est accentuée par le fait que les différentes variables obtenues à l'échelle zonale comportent dans la plupart des cas une certaine correspondance avec la taille de la zone (en population, par exemple) ce qui conduit à avoir des variables fortement corrélées et donc inutilisables de manière simultanée dans un modèle de régression linéaire. Ici, dans le cas du motif travail, en faisant un parallèle avec les variables explicatives retenues dans la modélisation par classification croisée, on pourrait être tenté d'inclure dans la régression le nombre de personnes de la zone déclarant avoir un emploi, *nbperstrav*; le nombre de véhicules disponibles dans la zone, *nbvpdispo*; et le nombre de ménages ayant des enfants (pour introduire un effet de type « cycle de vie »), *presenf*. Or, il se trouve que ces variables sont très corrélées entre elles (les coefficients de corrélation entre ces différentes variables pour toutes les dates d'enquête sont supérieurs à 0,7). Les régressions les incluant toutes les trois conduisent à avoir des coefficients non significatifs et parfois avec des signes contre-intuitifs pour les variables nombre de véhicules disponibles et nombre de ménages avec des enfants (cas d'école lorsque les régresseurs ne sont pas « assez » orthogonaux entre eux). Par ailleurs, la variable nombre de personnes déclarant avoir un emploi, *nbperstrav*, semble bien expliquer le nombre de demi-sorties pour le motif travail (figure D.1 de l'annexe D). La comparaison d'un modèle non contraint, par une analyse classique de la covariance, intégrant les trois variables citées précédemment, et d'un modèle contraint, intégrant seulement le nombre de personnes déclarant avoir un emploi (on teste l'hypothèse  $H_0$  que les coefficients pour les autres variables sont égaux à zéro), ne permet pas de déceler de différence significative.

Assez régulièrement, dans les modèles de génération par régression zonale, on observe une augmentation de la dispersion des données avec le régresseur (autre manifestation de l'effet « taille de la zone »). Dans le cas présent, d'un point de vue graphique, le constat est assez évident pour 1995 mais un peu plus difficile à faire pour 1985 et 2006 (figure D.1 en annexe, page 219). Le test de White (1980) met en évidence la présence d'hétéroscédasticité pour toutes les années et périmètres considérés. La présence d'une légère hétéroscédasticité n'est, en général, pas gênante dans le contexte des modèles zonaux car elle n'introduit pas de biais au niveau des coefficients. Elle peut biaiser les calculs des écarts-types des estimateurs. Elle a pour effet de rendre l'estimation par moindres carrés non-efficiente (dans le sens de variance

minimale). Dans le cas de notre étude, elle conduirait à des tests de stabilité à partir d'écart-types non-minimaux. Ceci a pour effet de rendre les tests un peu moins strictes. Dans le cas présent, on cherche à atténuer le problème d'hétéroscédasticité<sup>17</sup>. On applique donc une transformation de type racine carrée au nombre de demi-sorties pour le motif travail et au nombre de personnes déclarant avoir un emploi. Finalement, on retient une spécification où la constante est nulle, ceci en accord avec le fait qu'on imagine mal l'existence de déplacements pour le motif travail en absence de personnes travaillant et que les régressions préliminaires donnaient une constante non significative. La transformation des données fait que les tests ne mettent plus en évidence la présence d'hétéroscédasticité<sup>18</sup>.

**4.3.1.1.2 Autres motifs que Travail et Enseignement** Comme pour le motif travail, nous avons testé les spécifications correspondant aux combinaisons des variables disponibles à l'échelle zonale. Trois variables explicatives ont été retenues : le nombre de personnes résidant dans la zone, le nombre de véhicules disponibles et le nombre de ménages avec des enfants (variables faisant écho à celles retenues pour la classification croisée). Ces trois variables sont fortement corrélées, notamment le nombre de personnes et le nombre de véhicules disponibles. Nous avons donc opté, comme pour le motif travail, pour une spécification avec une seule variable, à savoir le nombre de personnes résidant dans la zone. Cette variable est souvent privilégiée dans les études de prévision à l'échelle zonale (voir Ashford et Holloway, 1972 ; DREIF, 2008) et, d'un point de vue graphique, elle semble bien expliquer les sorties pour les motifs autres que travail et enseignement (figure D.2). Les remarques faites par rapport à l'hétéroscédasticité sont valables aussi pour ce cas. Nous avons donc opté pour retenir une spécification employant la transformation racine carrée, sans constante.

#### 4.3.1.2 Résultats de calage

Comme on aurait pu s'attendre à partir des figures D.1 et D.2, les modèles de régression zonale ont une capacité descriptive très bonne.

**4.3.1.2.1 Motif travail** La table 4.8 donne les résultats de calage du modèle de régression zonale pour le motif travail, pour l'ensemble des dates et périmètres étudiés. Le coefficient de la variable explicative  $\sqrt{\text{nbperstrav}_i}$  est significativement différent de zéro et le coefficient de détermination  $R^2$  dépasse 0,99 dans tous les cas. L'indicateur *SRMSEG*, calculé à partir du nombre de demi-sorties observées et des valeurs modélisées au carré (pour tenir compte de la transformation racine carrée) permet de dire que l'erreur du modèle en 2006 est légèrement inférieure à celle des autres années. En termes d'évolution du coefficient de la variable explicative dans le temps, on observe une diminution d'environ 4 % entre 1985 et 1995 puis une diminution similaire entre 1995 et 2006, à périmètre constant. L'élargissement du

17. L'application de transformations aux variables est aisée dans le cas d'une spécification simple comme celle retenue ici. Lorsque l'on introduit plus de variables, comme dans les régressions à l'échelle des ménages, l'impact de l'hétéroscédasticité est généralement négligé

18. Une solution plus rigoureuse aurait été de faire une correction de l'hétéroscédasticité en pondérant les observations (dans une régression de type moindres carrés généralisés).

TABLE 4.8 – Résultats des calages des modèles de régression linéaire à l'échelle zonale pour le motif Travail

Variable	Coef.	écart-type	t-value	Pr(> t )	R <sup>2</sup>	SRMSEG	Transf. nég
Données : EMD 1985 (a)							
$\sqrt{\text{nbperstrav\_i}}$	1,42	0,009	158,9	8,33E-74	0,998	0,10	bcdef
Données : EMD 1995 Périm. 1985 (b)							
$\sqrt{\text{nbperstrav\_i}}$	1,36	0,009	156,6	1,82E-73	0,998	0,10	adef
Données : EMD 1995 Périm. 1995 (c)							
$\sqrt{\text{nbperstrav\_i}}$	1,36	0,010	137,1	1,83E-72	0,997	0,12	adef
Données : EMD 2006 Périm. 1985 (d)							
$\sqrt{\text{nbperstrav\_i}}$	1,31	0,007	181,9	5,77E-77	0,998	0,09	abc
Données : EMD 2006 Périm. 1995 (e)							
$\sqrt{\text{nbperstrav\_i}}$	1,32	0,007	194,6	5,82E-81	0,999	0,08	abc
Données : EMD 2006 Périm. AUL (f)							
$\sqrt{\text{nbperstrav\_i}}$	1,32	0,007	187,7	6,96E-111	0,998	0,10	abc

Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006.

périmètre d'analyse a un impact limité en termes de changement des coefficients pour une année déterminée. Ces observations se retrouvent au niveau du test de stabilité des coefficients, qui permet de rejeter l'hypothèse de stabilité temporelle entre toutes les paires de coefficients correspondant à des années différentes. En revanche, le changement de périmètre d'analyse pour 1995 et 2006 ne se traduit pas par une différence significative des coefficients.

**4.3.1.2.2 Autres motifs que Travail et Enseignement** La table 4.9 donne les résultats de calage du modèle de régression zonale pour les motifs autres que travail et enseignement, pour l'ensemble des dates et périmètres étudiés. Comme pour le motif travail, tous les coefficients de la variable explicative  $\sqrt{\text{nbpers\_i}}$  sont significativement différents de zéro et les  $R^2$  des différentes années et périmètres sont supérieurs à 0,99. Les *SRMSEG* sont plus importants que ceux du motif travail (d'environ 50 % en termes relatifs), le meilleur ajustement étant atteint avec les données de 1995 sur l'ensemble du périmètre d'enquête. Pour ce qui est des évolutions temporelles, on observe une augmentation du coefficient d'environ 8 % entre 1985 et 1995 puis une très légère baisse entre 1995 et 2006 (moins de 2 %). Au niveau du test de stabilité des paramètres, ces évolutions se traduisent par le rejet de l'hypothèse d'égalité pour le coefficient de 1985, comparé à tous les autres.

### 4.3.2 Capacité prédictive des modèles par régression linéaire à l'échelle zonale

De manière similaire à ce que nous avons fait pour les modèles de classification croisée, nous étudions la capacité des modèles de régression linéaire à l'échelle zonale à faire des prévisions à des horizons de 10 et de 20 ans. La table 4.10 donne les résultats des prévisions de la mobilité en 2006 issues de l'application des modèles de régression linéaire à l'échelle zonale estimés en 1985 et en 1995. De manière générale, les modèles calibrés avec les données de 1995 donnent de meilleures prévisions que ceux calibrés avec les données de 1985. L'intuition selon laquelle les changements dans

TABLE 4.9 – Résultats des calages des modèles de régression linéaire à l'échelle zonale pour les motifs autres que Travail et Enseignement.

Variable	Coef.	écart-type	t-value	Pr(> t )	R <sup>2</sup>	SRMSEG	Transf. nég
Données : EMD 1985 (a)							
$\sqrt{\text{nbpers}_i}$	1,12	0,012	97,1	2,73E-62	0,994	0,16	bcdef
Données : EMD 1995 Périm. 1985 (b)							
$\sqrt{\text{nbpers}_i}$	1,21	0,011	112,2	1,17E-65	0,996	0,14	a
Données : EMD 1995 Périm. 1995 (c)							
$\sqrt{\text{nbpers}_i}$	1,21	0,011	114,7	3,80E-68	0,996	0,13	a
Données : EMD 2006 Périm. 1985 (d)							
$\sqrt{\text{nbpers}_i}$	1,19	0,010	114,5	3,92E-66	0,996	0,14	a
Données : EMD 2006 Périm. 1995 (e)							
$\sqrt{\text{nbpers}_i}$	1,19	0,010	115,8	2,24E-68	0,996	0,14	a
Données : EMD 2006 Périm. AUL (f)							
$\sqrt{\text{nbpers}_i}$	1,19	0,009	131,2	5,24E-98	0,995	0,14	a

Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006.

le système urbain ayant lieu en 10 ans seraient moins importants que ceux ayant lieu en 20 ans est corroborée. Remarquons qu'en 2006, l'indicateur *SRMSEG* est moins important pour le motif travail que pour les motifs autres que travail et enseignement (l'écart relatif entre les deux étant compris entre 40 et 50 %). Pour le motif travail, l'écart entre le nombre de demi-sorties observées et modélisées est environ égal, en moyenne, à 10 % du nombre de demi-sorties pour chaque zone (lorsqu'on ne prend en compte que les données correspondant au périmètre de 1985, l'écart est légèrement plus faible que lorsqu'on utilise l'ensemble des données sur l'aire urbaine). Pour les motifs autres que travail et enseignement, quel que soit le périmètre choisi, l'écart est d'environ 14 %. En termes d'évolution des coefficients, nous avons vu que, pour le motif travail, l'écart entre ceux de 1995 et ceux de 2006 est d'environ 4 %. L'écart entre les coefficients de 1985 et de 2006 est d'environ 9 %. Ces écarts se traduisent, en prévision, par des surestimations de la mobilité. L'utilisation des coefficients de 1995 conduit à une surestimation de 7,46 % de la mobilité totale en 2006, sur le périmètre de l'EMD de 1985. L'utilisation du coefficient de 1995 estimé sur l'ensemble du périmètre de 1995 conduit à une surestimation de 5,28 % de la mobilité totale en 2006, pour le motif travail, sur le périmètre de l'aire urbaine. L'utilisation du coefficient de 1985 conduit en 2006 à des surestimations de 16,82 et 15,02 %, sur le périmètre de 1985 et sur celui de l'aire urbaine, respectivement. L'hypothèse de stabilité des coefficients utilisée en prévision combinée à la tendance à la baisse de la mobilité pour le motif travail, au cours du temps, conduit à des surestimations de la mobilité totale qui s'accroissent avec l'éloignement de la date pour laquelle on souhaite avoir des prévisions. On retrouve une configuration analogue lorsqu'on s'intéresse aux prévisions pour chaque zone. L'écart entre les coefficients de 1995 et ceux de 2006 fait que l'erreur moyenne soit supérieure de 1/3 par rapport à celle issue de l'application du modèle calibré en 2006. L'écart entre les coefficients de 1985 et ceux de 2006 conduit à une erreur deux fois supérieure. Pour les motifs autres que travail et enseignement, l'écart entre le coefficient associé au nombre de personnes de chaque zone en 1995 et celui de 2006 est très faible (environ 2 %, le coefficient de 2006 étant inférieur à celui de 1995). L'écart entre le coefficient de 1985 et celui de 2006 est plus important (un peu plus de 6 %, le coefficient de 2006 étant supérieur à

TABLE 4.10 – Capacité des modèles de régression à l'échelle zonale, calibrés en 1985 et 1995, à prévoir la mobilité observée en 2006.

Variable	<i>EREL</i>		<i>SRMSEG</i>		
	1985	1995	1985	1995	2006
Prévision sur le Périmètre de l'EMD 1985					
$p = \text{Travail}$					
$\sqrt{\text{nbperstrav}_i}$	16,82	7,46	0,20	0,12	0,09
$p = \text{Autres que Travail et Enseignement}$					
$\sqrt{\text{nbpers}_i}$	-11,16	2,30	0,18	0,14	0,14
Prévision sur le Périmètre de l'AUL en utilisant toutes les données					
$p = \text{Travail}$					
$\sqrt{\text{nbperstrav}_i}$	15,02	5,28	0,19	0,11	0,10
$p = \text{Autres que Travail et Enseignement}$					
$\sqrt{\text{nbpers}_i}$	-11,62	2,01	0,20	0,14	0,14

Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006.

celui de 1985). Ces écarts se traduisent par une légère surestimation de la mobilité totale de 2006 par le modèle calibré en 1995 (environ 2%). Cependant, au niveau de la mobilité par zone, les résultats des prévisions de 1995 sont comparables à ceux de l'estimation réalisée en 2006. Le modèle estimé en 1985 a tendance à sous-estimer la mobilité totale en 2006 d'entre 11 et 12%. Au niveau zonal, les erreurs sont en moyenne 43% plus importantes que celles obtenues avec les modèles estimés en 1995 et en 2006.

Les résultats présentés ci-dessus ne sont pas surprenants, compte tenu de ce qui a déjà été exposé dans la section 4.1.1.2 concernant l'évolution de la mobilité par personne<sup>19</sup>. Les mobilités pour les deux motifs étudiés connaissent des évolutions à la hausse ou à la baisse et non une fluctuation autour d'une valeur « moyenne ». Il s'en suit que les prévisions sont d'autant plus mauvaises que l'horizon auquel elles sont réalisées est éloigné. Notons que l'évolution de la mobilité totale par personne, sans faire de distinction par motif, est croissante puis décroissante sur la période étudiée et, de ce fait, un modèle fruste tel que celui qui a été utilisé ici pour les motifs autres que travail et enseignement (avec le nombre de personnes étant la seule variable explicative) pourrait donner des prévisions de qualité comparable à des horizons de 10 ou 20 ans. Certains des résultats de la littérature, présentés dans la section 2.2.1, concernant la supériorité en prévision des modèles statiques calibrés en combinant les données de plusieurs enquêtes sont en partie dus au fait que ces études sont réalisées sur la mobilité totale. Ce propos doit néanmoins être tempéré par le fait que la performance prédictive des modèles de classification croisée est supérieure à celle des modèles par régression zonale. On peut en déduire qu'il y a un certain degré de stabilité des comportements pour certaines catégories et que les évolutions de la mobilité par personne, établie d'une manière agrégée, cachent en partie des effets de composition. La section suivante procède donc à étudier un cas intermédiaire, celui des modèles de régression à l'échelle des ménages, pour essayer de préciser la nature de ces évolutions

19. Pour le motif travail, il suffisait de mettre en rapport l'évolution à la baisse des sorties pour le motif travail avec la stabilité de la part de la population active

## 4.4 Le modèle de génération par régression à l'échelle des ménages

### 4.4.1 Spécification

Un des arguments en faveur de l'analyse à l'échelle des ménages plutôt qu'à l'échelle zonale est la possibilité d'employer plus de régresseurs. Ces derniers peuvent être quantitatifs ou qualitatifs. Notons, que dans le cas extrême où l'on n'utiliserait que des variables qualitatives, sous forme de *dummy*-variables<sup>20</sup> comme régresseurs, on reviendrait à une configuration similaire à celle d'un modèle de classification (en effet, il s'agirait de l'analyse de variance qui, dans le cas de la classification croisée, permet de déterminer la variance expliquée par chacun des régresseurs retenus). Or, l'intérêt de garder une forme paramétrique est que celle-ci permet de prendre en compte de manière naturelle les variables continues (ou du moins non-catégorielles)<sup>21</sup>. Pour cette raison, le choix de la spécification retenue ici doit privilégier l'utilisation des variables quantitatives contenues dans les EMD. Ainsi, par exemple, on privilégiera l'utilisation du nombre de personnes du ménage (variable *nbpers*) par rapport à celle de son équivalent sous forme catégorielle (*nbpers\_F*). Le recours aux *dummy*-variables doit se limiter à essayer d'intégrer certains effets qui pourraient se voir exclus autrement, en raison d'une corrélation trop importante des régresseurs, liée à l'effet taille du ménage (cette pratique étant assez répandue ; voir, par exemple, Mwakalonge, 2010).

Nous avons vu dans la littérature que le fait de complexifier les modèles pour améliorer leur performance descriptive n'a pas toujours un impact positif en termes de capacité prédictive du modèle. Ce constat nous a conduit à finalement retenir deux spécifications par mode. Le premier type de spécification est simple. Il contient uniquement la variable *nbperstrav*, pour le motif travail, et *nbpers*, pour les motifs autres que travail et enseignement. Le deuxième type de spécification est étendu et il contient une décomposition par sexe de ces variables auxquelles on ajoute quelques variables supplémentaires (voir tables 4.11 et 4.12). Les spécifications simples et étendues n'ont pas de terme constant

### 4.4.2 Résultats des calages

#### 4.4.2.1 Motif Travail

La table 4.11 donne les résultats des calages réalisés pour le motif travail, en prenant en compte l'impact du plan de sondage. Pour toutes les années et périmètres, la spécification étendue reproduit un peu mieux les données observées que la spécification simple, à l'échelle zonale (*SRMSEG*). Le coefficient associé à la variable *nbperstrav* est toujours significativement différent de zéro. Les coefficients associés à la version décomposée selon le sexe de cette variable (*nbhomtrav* et *nbfemtrav*) le sont également. En termes d'évolutions temporelles, tant pour la spécification simple

20. Variables prenant la valeur zéro ou un, uniquement. Elles sont généralement utilisées pour introduire des variables catégorielles dans la régression.

21. Rappelons qu'une des limites de la classification croisée correspond à la détermination des points de rupture pour déterminer les classes de certaines variables originellement quantitatives.

TABLE 4.11 – Résultats de calage des modèles de génération pour le motif Travail à l'échelle des ménages (avec prise en compte explicite du plan de sondage) et test de stabilité temporelle des paramètres.

Variable	Coef.	écart-type	t-value	Pr(> t )	SRMSEG	Transf. nég
Données : EMD 1985 (a)						
Spécification étendue						
nbhomtrav	1,93	0,057	33,6	1,26E-221	0,103	def
nbfemtrav	1,90	0,065	29,3	3,46E-173		def
nbpermis	0,09	0,031	2,9	4,20E-03		
presfoyer	0,07	0,060	1,2	2,16E-01		
Spécification simple						
nbperstrav	2,01	0,027	74,7	0,00E+00	0,104	bcdef
Données : EMD 1995 Périm. 1985 (b)						
Spécification étendue						
nbhomtrav	1,80	0,048	37,8	3,92E-280	0,100	d
nbfemtrav	1,78	0,048	37,2	4,52E-271		def
nbpermis	0,04	0,019	2,2	2,49E-02		
presfoyer	0,15	0,047	3,3	9,58E-04		
Spécification simple						
nbperstrav	1,84	0,023	81,8	0,00E+00	0,103	adef
Données : EMD 1995 Périm. 1995 (c)						
Spécification étendue						
nbhomtrav	1,78	0,048	37,0	2,79E-270	0,118	
nbfemtrav	1,79	0,047	38,0	1,76E-283		def
nbpermis	0,03	0,018	1,9	6,31E-02		
presfoyer	0,18	0,046	3,8	1,49E-04		
Spécification simple						
nbperstrav	1,83	0,022	81,8	0,00E+00	0,120	adef
Données : EMD 2006 Périm. 1985 (d)						
Spécification étendue						
nbhomtrav	1,67	0,043	38,6	8,85E-292	0,086	ab
nbfemtrav	1,58	0,045	35,2	2,04E-247		abc
nbpermis	0,07	0,015	4,7	3,19E-06		
presfoyer	0,05	0,060	0,8	3,96E-01		
Spécification simple						
nbperstrav	1,70	0,021	80,3	0,00E+00	0,088	abc
Données : EMD 2006 Périm. 1995 (e)						
Spécification étendue						
nbhomtrav	1,68	0,042	40,5	0,00E+00	0,081	a
nbfemtrav	1,59	0,043	36,9	4,16E-271		abc
nbpermis	0,07	0,014	5,1	2,83E-07		
presfoyer	0,06	0,058	1,0	3,09E-01		
Spécification simple						
nbperstrav	1,72	0,020	85,6	0,00E+00	0,084	abc
Données : EMD 2006 Périm. AUL (f)						
Spécification étendue						
nbhomtrav	1,71	0,036	47,8	0,00E+00	0,097	a
nbfemtrav	1,60	0,038	42,4	0,00E+00		abc
nbpermis	0,06	0,012	5,1	3,48E-07		
presfoyer	0,07	0,050	1,4	1,57E-01		
Spécification simple						
nbperstrav	1,73	0,017	102,8	0,00E+00	0,102	abc

Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006.

que pour la spécification étendue, les coefficients associés au nombre de personnes employées (hommes ou femmes) baissent avec le temps. Dans la plupart des cas, la baisse se situe entre 7 et 8 %. Cependant, entre 1995 et 2006, on observe une baisse plus prononcée pour le coefficient correspondant à la mobilité des femmes employées (environ 11 %). De ce fait, alors qu'en 1985 et 1995 les coefficients associés au nombre d'hommes et au nombre de femmes employés sont quasiment identiques, en 2006,

on observe un coefficient plus faible associé au nombre de femmes employées. Les coefficients associés aux variables *nbpermis* et *presfoyer* ne sont pas toujours significativement différents de zéro. En 1985 et 2006, le coefficient associé à *nbpermis* est significativement différent de zéro et celui associé à *presfoyer* ne l'est pas. En revanche, en 1995, on observe le cas inverse. Ce résultat est quelque peu contre-intuitif. On se serait attendu à observer un impact décroissant du nombre de permis (traduisant la « maturité » de la motorisation). Pour la présence d'un adulte au foyer, on aurait espéré que cette variable soit positive et significativement différente de zéro en 1985, traduisant ainsi la facilité accrue pour les actifs à déjeuner chez eux lorsque quelqu'un est déjà présent au domicile. En termes du test de stabilité, pour la spécification simple, la décroissance continue de l'impact du nombre d'actifs sur la mobilité pour le motif travail est retrouvée dans le fait que le test rejette l'égalité des paramètres appartenant à des années différentes. Pour la spécification étendue, l'importance relative des écarts-types par rapport aux coefficients des variables *nbpermis* et *presfoyer* rend impossible le rejet de l'hypothèse de stabilité (et rend le test peu pertinent). Pour les variables *nbhomtrav* et *nbfemtrav* les résultats sont moins nets que ceux de la spécification simple. On observe un rejet de l'hypothèse d'égalité entre 1985 et 2006, mais pour les comparaisons à 10 ans d'écart (1995–2006, par exemple) ce n'est pas toujours le cas.

En comparant la table 4.11 à la table D.1 (voir annexe, page 221), donnant les résultats d'un calage ne prenant pas en compte le plan de sondage et considérant les observations comme si elle provenaient d'un tirage aléatoire simple, on remarque que l'impact du plan de sondage sur les coefficients est limité. Les coefficients présentés sur les deux tables sont positifs. Pour les coefficient associés au nombre d'actifs, on a des écarts d'au plus 2 %, dans la spécification simple. Dans la spécification étendue, l'écart le plus important entre les deux tables pour les coefficients de *nbhomtrav* ou *nbfemtrav* sont de 3 %. En revanche, pour les deux autres variables de la spécification étendue (*nbpermis* et *presfoyer*), on retrouve des écarts plus importants. Par exemple, le coefficient associé à *presfoyer* dans la table 4.11 n'est pas significativement différent de zéro. Sur la table D.1 (annexe D page 221), il est presque deux fois supérieur et il est significativement différent de zéro. Il reste cependant assez faible en valeur absolue, d'autant plus qu'il est associé à une *dummy*-variable. Au niveau du test de stabilité, l'impact de la prise en compte du plan de sondage conduit au rejet de l'hypothèse de stabilité pour une partie plus importante des comparaisons à 10 ans d'écart.

#### 4.4.2.2 Motifs autres que Travail et Enseignement

La table 4.12 donne les résultats des calages réalisés pour les motifs autres que travail et enseignement, en prenant en compte l'impact du plan de sondage. Comme pour le motif travail, la spécification étendue reproduit mieux les données observées que la spécification simple, à l'échelle zonale (*SRMSEG*), pour toutes les années et périmètres. Le gain apporté varie selon les dates d'enquête et le périmètre. Il est compris entre 9 %, en 1995 sur le périmètre de l'EMD de 1985, et 27 %, en 2006 sur le périmètre de l'EMD de 1995. Pour la spécification simple, le coefficient associé au nombre de personnes est toujours significativement différent de zéro. Pour la spécification étendue, les coefficients associés à la version décomposée selon le sexe de

cette variable (nbhom et nbfem) le sont également. Contrairement à ce qui se passe pour le motif travail, les coefficients des variables supplémentaires — c'est-à-dire les variables ne portant pas directement sur la taille de la population concernée par le motif de déplacement — sont en général significativement différents de zéro. Les coefficients pour ces variables sont positifs. On pouvait s'attendre à ce résultat pour les variables presfoyer, presret et preschom dans la mesure où elles caractérisent la présence de personnes n'ayant pas d'activité laborale. Il s'en suit que la possibilité d'« inclure » certains déplacements pour des achats ou des loisirs, par exemple, au sein d'une sortie pour le motif travail n'est pas envisageable. La variable presvp a également un signe positif en 1985 et en 1995 (le coefficient n'est pas significatif en 2006). On aurait pu penser que la motorisation permet d'enchaîner certains déplacements, notamment pour le motif achats-services (Doubleday, 1977), mais elle traduit aussi un revenu plus important et donc une possibilité accrue de réaliser des activités, l'effet combiné étant positif. Le signe positif du coefficient associé à la variable presconj (en 1995 et 2006, en 1985 le coefficient est non significatif) est un peu plus surprenant dans la mesure où on se serait attendu à ce que la présence de ce lien au sein d'un ménage conduise à une coordination plus importante des déplacements qui se traduirait par un nombre de demi-sorties moins important, et donc un signe négatif du coefficient. On peut penser que la coordination au sein des couples ne joue pas un rôle important dans la réduction de la mobilité<sup>22</sup>. En termes d'évolutions temporelles, pour la spécification simple, on observe, de manière similaire à ce qui se passe au niveau zonal, que le coefficient associé au nombre de personnes s'accroît entre 1985 et 1995 (augmentation d'un peu plus de 15 %, à périmètre constant) pour ensuite décroître légèrement entre 1995 et 2006 (baisse d'un peu moins de 4 %, à périmètre constant). L'élargissement du périmètre d'analyse a un impact très faible sur ce coefficient. Au niveau de la spécification étendue, les coefficients connaissent des évolutions différentes selon les variables. Les coefficients associés à nbhom et preschom connaissent une augmentation entre 1985 et 1995 (14 et 11 %, respectivement) puis une diminution entre 1995 et 2006 (environ 13 % pour les deux), ce qui place le coefficient de nbhom en 2006 à son niveau de 1985 et celui de preschom à un niveau un peu plus bas que celui de 1985. Les coefficients associés à nbfem, presconj et presret connaissent une augmentation continue entre 1985 et 2006 (47 % pour nbfem et 26 % pour presret alors que celui associé à presconj augmente de 77 % entre 1995 et 2006 et n'est pas significatif en 1985). Ainsi, le nombre de femmes a un impact plus important en termes de mobilité en 2006 que le nombre d'hommes alors que c'était l'inverse en 1985. Par ailleurs, l'impact de la présence de retraités en termes de mobilité s'accroît, ce qui, par des biais différents, reflète le vieillissement de la population (entre autres, l'espérance de vie étant plus importante, la mobilité des retraités, en meilleure santé, l'est aussi). Les coefficients associés à presvp et presfoyer connaissent une diminution continue entre 1985 et 2006 (22 % entre 1985 et 1995 pour presvp et il n'est pas significatif en 2006 ; 28 % entre 1985 et 2006 pour presfoyer). L'impact en baisse de la présence de véhicules disponibles, au point d'arriver à la non-significativité du coefficient en 2006 peut traduire, d'une certaine manière, l'arrivée au stade de « maturité » de la motorisation. Le test de

22. Rappelons que nous avons vu quelque chose de similaire au niveau des classifications croisées, avec la variable cycle de vie.

## 4. Génération

TABLE 4.12 – Résultats de calage des modèles de génération pour les motifs autres que Travail et Enseignement à l'échelle des ménages (avec prise en compte explicite du plan de sondage) et test de stabilité temporelle des paramètres.

Variable	Coef.	écart-type	t-value	Pr(> t )	SRMSEG	Transf. nég
Données : EMD 1985 (a)						
Spécification étendue						
nbhom	0,86	0,082	10,4	3,18E-25	0,136	
nbfem	0,68	0,071	9,5	2,94E-21		bcdef
presvp	0,47	0,104	4,5	8,17E-06		def
presconj	0,09	0,138	0,7	5,03E-01		d
presfoyer	1,30	0,142	9,2	7,67E-20		
presret	1,40	0,099	14,1	3,43E-44		def
preschom	1,55	0,213	7,3	4,60E-13		
Spécification simple						
nbpers	1,18	0,024	49,8	0,00E+00	0,177	bcdef
Données : EMD 1995 Périm. 1985 (b)						
Spécification étendue						
nbhom	0,98	0,076	12,9	2,32E-37	0,134	
nbfem	0,90	0,063	14,3	1,87E-45		a
presvp	0,37	0,090	4,1	4,84E-05		
presconj	0,26	0,129	2,0	4,70E-02		
presfoyer	1,26	0,154	8,2	2,65E-16		
presret	1,64	0,089	18,5	3,02E-74		
Données : EMD 2006 Périm. 1985 (d)						
preschom	1,72	0,186	9,3	2,48E-20		
Spécification simple						
nbpers	1,36	0,023	58,6	0,00E+00	0,148	a
Données : EMD 1995 Périm. 1995 (c)						
Spécification étendue						
nbhom	0,99	0,077	12,9	1,19E-37	0,132	
nbfem	0,90	0,062	14,5	7,85E-47		a
presvp	0,37	0,091	4,1	4,70E-05		
presconj	0,26	0,128	2,1	4,00E-02		
presfoyer	1,31	0,152	8,6	9,58E-18		
presret	1,59	0,087	18,2	6,27E-72		f
preschom	1,72	0,187	9,2	5,77E-20		
Spécification simple						
nbpers	1,37	0,023	59,3	0,00E+00	0,146	a
Spécification étendue						
nbhom	0,84	0,073	11,5	3,77E-30	0,121	
nbfem	1,00	0,075	13,3	6,92E-40		a
presvp	0,16	0,087	1,9	6,28E-02		a
presconj	0,46	0,114	4,0	5,54E-05		a
presfoyer	0,98	0,185	5,3	1,13E-07		
presret	1,76	0,079	22,2	2,76E-105		a
preschom	1,49	0,172	8,7	6,19E-18		
Spécification simple						
nbpers	1,31	0,023	57,1	0,00E+00	0,162	a
Données : EMD 2006 Périm. 1995 (e)						
Spécification étendue						
nbhom	0,86	0,072	12,0	7,79E-33	0,118	
nbfem	1,00	0,072	13,8	1,57E-42		a
presvp	0,16	0,085	1,9	5,72E-02		a
presconj	0,43	0,111	3,8	1,27E-04		
presfoyer	0,94	0,179	5,3	1,38E-07		
presret	1,79	0,077	23,3	8,94E-116		a
preschom	1,49	0,169	8,8	1,94E-18		
Spécification simple						
nbpers	1,31	0,022	59,4	0,00E+00	0,161	a
Données : EMD 2006 Périm. AUL (f)						
Spécification étendue						
nbhom	0,94	0,062	15,1	8,34E-51	0,125	
nbfem	0,99	0,062	16,1	3,00E-57		a
presvp	0,15	0,077	1,9	5,43E-02		a
presconj	0,37	0,097	3,8	1,59E-04		
presfoyer	0,93	0,154	6,0	1,62E-09		
presret	1,80	0,067	27,0	9,23E-154		ac
preschom	1,54	0,151	10,2	1,93E-24		
Spécification simple						
nbpers	1,33	0,019	71,2	0,00E+00	0,167	a

Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006.

stabilité des paramètres ne révèle que des différences pour les variables connaissant des évolutions monotones et l'hypothèse de stabilité est rejeté dans la plupart des cas dans des comparaisons mettant en jeu les paramètres de 1985.

La comparaison des résultats donnés dans la table 4.12 à ceux de la table D.2 (voir annexe, page 222), donnant les résultats d'un calage ne prenant pas en compte le plan de sondage, montre un impact assez limité de celui-ci sur les coefficients. Si on s'intéresse à la spécification simple, le coefficient associé au nombre de personnes du ménage est toujours supérieur lorsqu'on prend en compte le plan de sondage, mais seulement de 2 à 3 %. Dans la spécification étendue, les coefficients issus de l'estimation avec prise en compte du plan de sondage sont en majorité supérieurs à ceux estimés sans sa prise en compte. Tous les coefficients estimés sans prise en compte du plan de sondage sont significativement différents de zéro, à l'exception de celui associé à la variable *presconj* pour 1985. Si l'on omet les coefficients non significatifs, l'écart le plus important entre les coefficients des deux tables est d'un peu plus de 16 % (*preschom* en 1985). L'écart le plus faible est quasiment nul (*nbfem* en 1995). Au niveau du test de stabilité, comme pour le motif travail, l'impact de la prise en compte du plan de sondage conduit au rejet de l'hypothèse de stabilité pour une partie plus importante des comparaisons à 10 ans d'écart.

### 4.4.3 Capacité prédictive des modèles par régression à l'échelle du ménage

Suivant la même approche que celle que nous avons utilisée pour les modèles de classification croisée et de régression à l'échelle zonale, nous regardons la capacité prédictive à 10 et 20 ans des modèles de régression linéaire à l'échelle des ménages. Les tables 4.13 et 4.14 donnent les résultats des prévisions des demi-sorties émises en 2006, réalisées en utilisant les paramètres estimés avec les données de 1985 et 1995. Elles incluent des prévisions réalisées avec les paramètres calibrés avec et sans prise en compte du plan de sondage. La table 4.13 donne les résultats de ces prévisions à périmètre constant alors que la table 4.14 est réalisée à partir de l'ensemble des données disponibles lors de chaque enquête (dans la limite de l'aire urbaine lyonnaise).

On observe, sur la table 4.13, que les régressions à l'échelle des ménages de 1995 produisent des prévisions meilleures que celles réalisées avec les données de 1985, en termes de mobilité totale. Pour le motif travail, l'utilisation des coefficients de 1995 conduit à des surestimations de la mobilité totale de 2006 (*EREL*) comprises entre 6,44 et 8,13 %. L'utilisation des coefficients de 1985 conduit à des surestimations plus de deux fois supérieures (comprises entre 14,54 et 19,08 %). Tant pour les coefficients calibrés en prenant en compte le plan de sondage que pour ceux calibrés sans prise en compte de celui-ci, ce sont les spécifications simples qui produisent les meilleurs résultats en prévision. Dit autrement, pour le motif travail, une connaissance plus fine des déterminants de la mobilité à un instant dans le temps ne garantit pas des prévisions meilleures. Il y a vraiment eu une évolution comportementale que les variables retenues ne permettent pas de saisir complètement. Celle-ci a également une conséquence au niveau des prévisions réalisées avec ou sans prise en compte du plan de sondage : en théorie, la prise en compte de celui-ci devrait permettre d'avoir des estimations plus justes des vraies valeurs des paramètres à un moment

donné. Nous avons vu précédemment (section 4.4.2.1) que les coefficients issus d'une estimation sans prise en compte du plan de sondage sont, dans les modèles étudiés ici, quasiment tout le temps inférieurs aux coefficients issus d'estimations prenant en compte le plan de sondage. De ce fait, ils permettent de mieux prédire une mobilité pour le motif travail par ménage qui est en baisse. Au niveau de la prévision de la mobilité pour chaque ménage (*MAE*), on observe quelque chose de surprenant : les modèles calibrés en 1985 et en 1995 semblent mieux reproduire les données que les calages réalisés en 2006. Deux éléments peuvent expliquer ce constat. D'une part, le calage se fait en minimisant la somme des erreurs au carré et non des erreurs absolues. De ce fait, les deux indicateurs peuvent donner des résultats contrastés pour un même jeu de paramètres. D'autre part, l'ajustement des modèles linéaires à l'échelle des ménages est généralement moins bon qu'il ne peut l'être à l'échelle zonale en raison des distributions des variables (l'hypothèse de normalité semble plus appropriée dans les régressions zonales). Au niveau de la prévision à l'échelle des zones d'analyse (*SRMSEG*), on observe une configuration similaire à celle observée pour la mobilité totale. Les modèles calibrés avec les données de 1995 permettent de mieux reproduire la mobilité observée en 2006 que ceux calibrés avec les données de 1985. Le modèle qui conduit aux meilleures prévisions est la spécification simple sans prise en compte du plan de sondage. Si on s'intéresse aux motifs autres que travail et enseignement, on observe, au niveau des prévisions de la mobilité totale, que les modèles de 1995 sont plus performants que ceux de 1985. L'erreur de prévision est légèrement moins importante que pour le motif travail. On remarque, dans la plupart des cas, une sous-estimation de la mobilité de 2006 sauf pour la spécification étendue calibrée avec les données de 1995, en prenant en compte le plan de sondage. Contrairement à ce qui se passe pour le motif travail, pour les motifs autres que travail et enseignement, la spécification étendue conduit à de meilleures prévisions que la spécification simple. Dans la plupart des cas, la prise en compte du plan de sondage conduit à faire de meilleures prévisions de la mobilité observée en 2006. La seule exception correspond à la spécification étendue, en 1995, pour laquelle la non prise en compte du plan de sondage conduit à des résultats légèrement plus proches des observations. Cependant, dans le cas de la prise en compte du plan de sondage on a une surestimation de la mobilité totale de 2006 alors que dans l'autre cas on a une sous-estimation. Les coûts induits dans les deux cas peuvent varier grandement (voir Flyvbjerg *et al.*, 2005). La meilleure performance de la spécification étendue peut également être observée au niveau de la prévision des comportements de mobilité à l'échelle des ménages. Comme pour le motif travail, dans certains cas la performance des modèles calibrés en 1985 et 1995 est supérieure à celle des estimations de 2006. Au niveau zonal, les résultats sont sans équivoque : la spécification étendue estimée en prenant en compte le plan de sondage donne les meilleurs résultats en prévision (tout comme en description en 2006). Pour cette spécification, l'erreur de prévision faite avec les coefficients estimés en utilisant les données de 1995 est supérieure à celle de l'estimation réalisée en 2006 d'environ 1 % de la valeur moyenne par cellule, soit encore un écart relatif d'environ 8 %. L'utilisation des données de 1985 conduit à une erreur supérieure à celle de l'estimation réalisée en 2006 d'environ 6 % de la valeur moyenne par cellule, soit encore un écart relatif de 50 %.

Lorsqu'on s'intéresse aux calibrations mettant en jeu l'ensemble des données des différentes années (table 4.14), on constate, pour le motif travail, une amélioration

TABLE 4.13 – Prévision de la mobilité de 2006 dans le périmètre de l'EMD 1985 avec les coefficients estimés sur ce même périmètre en 1985 et 1995, avec et sans prise en compte du plan de sondage

Motif	Spécification	EREL		MAE			SRMSEG		
		1985	1995	1985	1995	2006	1985	1995	2006
<i>p = Travail</i>									
Sans prise en compte du plan de sondage									
	étendue	16,48	7,49	0,76	0,77	0,81	0,20	0,12	0,09
	simple	14,54	6,44	0,72	0,75	0,79	0,18	0,11	0,09
Avec prise en compte du plan de sondage									
	étendue	19,09	8,13	0,77	0,77	0,81	0,22	0,12	0,09
	simple	16,24	6,62	0,72	0,75	0,78	0,20	0,11	0,09
<i>p = Autres que Travail et Enseignement</i>									
Sans prise en compte du plan de sondage									
	étendue	-13,63	-3,60	2,20	2,22	2,21	0,20	0,13	0,17
	simple	-19,06	-6,99	2,39	2,40	2,39	0,25	0,16	0,18
Avec prise en compte du plan de sondage									
	étendue	-11,55	3,82	2,20	2,24	2,22	0,18	0,13	0,12
	simple	-16,90	-4,31	2,39	2,41	2,40	0,23	0,14	0,16

Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006.

TABLE 4.14 – Prévision de la mobilité de 2006 dans l'AUL avec les coefficients estimés en utilisant toutes les données disponibles en 1985 et 1995, avec et sans prise en compte du plan de sondage

Motif	Spécification	EREL		MAE			SRMSEG		
		1985	1995	1985	1995	2006	1985	1995	2006
<i>p = Travail</i>									
Sans prise en compte du plan de sondage									
	étendue	14,60	5,53	0,77	0,78	0,82	0,18	0,11	0,10
	simple	12,77	4,66	0,73	0,77	0,80	0,17	0,11	0,11
Avec prise en compte du plan de sondage									
	étendue	17,15	5,57	0,78	0,78	0,82	0,21	0,11	0,10
	simple	14,45	4,37	0,73	0,77	0,80	0,18	0,11	0,10
<i>p = Autres que Travail et Enseignement</i>									
Sans prise en compte du plan de sondage									
	étendue	-14,88	-3,70	2,29	2,31	2,31	0,21	0,14	0,17
	simple	-19,48	-7,00	2,47	2,49	2,48	0,27	0,17	0,18
Avec prise en compte du plan de sondage									
	étendue	-12,84	2,55	2,28	2,32	2,31	0,20	0,13	0,13
	simple	-17,33	-4,35	2,47	2,50	2,49	0,25	0,15	0,17

Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006.

des prévisions par rapport à celles réalisées uniquement sur le périmètre de l'EMD de 1985, en ce qui concerne la mobilité totale. En ce qui concerne la mobilité à l'échelle des ménages, on observe une légère baisse tant dans la qualité de reproduction des observations de 2006 comme dans la qualité des prévisions. Il n'y a cependant pas de changement de configuration, les paramètres calibrés en 1985 donnent toujours la valeur la plus faible de l'indicateur *MAE*. Pour ce qui est de l'indicateur zonal, comme pour l'indicateur à l'échelle des ménages, on remarque une légère baisse de la qualité de la reproduction des données de 2006 par le modèle calibré en cette année

sur le périmètre de l'aire urbaine. En revanche, au niveau des prévisions réalisées avec les paramètres de calibration de 1985 et de 1995, on a, comme pour la mobilité totale, une légère amélioration. La performance relative des différentes spécifications n'a pas changé, celle qui donne les meilleures prévisions étant la spécification simple calibrée sans prendre en compte le plan de sondage. Pour les motifs autres que travail et enseignement, on observe une légère baisse de la qualité des prévisions de la mobilité totale, sauf pour la spécification étendue estimée avec les données de 1995, en prenant en compte le plan de sondage. La baisse de qualité dans les prévisions de la mobilité à l'échelle des ménages est générale, et on peut la retrouver également au niveau zonal.

## 4.5 Classification automatique : est-ce optimal de conserver les mêmes segments de population en prévision ?

Nous avons vu dans les sections précédentes de ce chapitre que le modèle de classification croisée permet d'obtenir, pour un certain choix de variables, des résultats en matière de prévision qui sont meilleurs que ceux des approches par régression, que ce soit au niveau zonal comme au niveau des ménages. Une partie de ce succès peut être expliquée par une des qualités de l'approche par classification croisée qui est de ne pas avoir de contrainte sur la forme des variables explicatives. Ainsi, il a été possible d'inclure la variable cycle de vie, qui est qualitative et qui n'a pas une relation linéaire avec la mobilité. Or, si la forme des variables ne pose pas de problème particulier pour cette méthode, la question du choix des variables à retenir reste ouverte. Nous avons appliqué l'algorithme CART aux données provenant des différentes enquêtes, en prenant en compte les changements de périmètre, pour savoir si les segments de population ainsi définis sont constants dans le temps. Plus précisément, on cherche à savoir si les variables retenues pour chaque cas, ainsi que les points de coupure de chacune de ces variables, sont les mêmes pour tous les cas considérés. Les résultats de l'application de l'algorithme CART sont donnés ci-dessous.

### 4.5.1 Motif Travail

Les figures D.3 à D.14 de l'annexe D illustrent les résultats de la procédure CART pour le motif travail, avec (figures D.3 à D.8) et sans (figures D.9 à D.14) prise en compte des probabilités de tirage liées au plan de sondage. On remarque que tous ces arbres ont une forme très proche. La variable la plus discriminante en matière des profils de mobilité est `nbperstrav`, pour les trois premiers niveaux de divisions binaires. En effet, on aurait pu s'y attendre, au moins au niveau de la première division, qui sépare les profils des ménages sans actifs employés de ceux des ménages où au moins l'un des membres travaille. C'est naturellement au niveau de cette première division que le gain en pureté est le plus important (signifié par la hauteur des branches) puisque les ménages sans actif ont théoriquement un profil identique de zéro demi-sorties réalisées pour le motif travail. Au-delà de la troisième division, qui sépare les ménages à 2 actifs des ménages à plus de 2 actifs, on commence à observer des différences dans les variables retenues par l'algorithme. Les variables qui

apparaissent le plus ont un lien direct avec la motorisation du ménage (nbvpdispo et nbpermis). Par rapport aux autres types de modèles étudiés dans ce chapitre, l'arbre de classification semble attribuer un rôle plus important que celui auquel on se serait attendu. La variable cycle de vie n'apparaît comme contribuant significativement au gain de pureté des profils de mobilité qu'en 2006, pour les ménages à un actif employé. Ceci corrobore l'idée que la coordination des déplacements pour certains types de ménages joue un rôle important, ce qui, avec l'arrivée au stade de maturité de la motorisation, conduit à faire du cycle de vie des ménages une variable intéressante pour les modèles de prévision.

#### 4.5.2 Motifs autres que Travail et Enseignement

Les figures D.15 à D.26 de l'annexe D illustrent les résultats de la procédure CART pour les motifs autres que travail et enseignement, avec (figures D.15 à D.20) et sans (figures D.21 à D.26) prise en compte des probabilités de tirage liées au plan de sondage. Par rapport au motif travail, on remarque une hétérogénéité plus importante au niveau de la forme des arbres. En effet, la première division est la même pour tous les arbres. Elle distingue les ménages uni-personnels des autres. Sur le côté « gauche » de cette division (ménages uni-personnels), on peut observer les deux mêmes variables explicatives nbpersretraite et nbperschom, mais l'apport discriminant de nbpersretraite est bien plus important en 1995 et 2006. À « droite » de la première division, l'importance de la variable nbpersretraite s'est également accrue. En 1985, sur ce côté, la première division se fait à partir de la variable presfoyer. En 1995 et 2006 elle se fait à partir de nbpersretraite, et le pouvoir discriminant de cette variable est, à ce niveau, plus important en 2006 qu'en 1995. Les ménages uni-personnels et ceux à deux personnes ou plus ayant des retraités ont les profils de mobilité les moins hétérogènes. Ceci, combiné au vieillissement de la population, explique l'importance accrue du rôle de la variable nbpersretraite dans la segmentation. Les autres ménages ont des mobilités beaucoup plus hétérogènes. Celles-ci se trouvent départagées par la taille des ménages (réapparition de la variable nbpers), la motorisation et leur structure (cycledevie, presfoyer, presconj). Cependant, compte tenu de l'hétérogénéité de la mobilité, le pouvoir discriminant de ces variables est relativement faible, ce qui conduit à avoir des structures d'arbres (en termes de niveaux de divisions et de variables utilisées à chaque niveau) très différente d'une année à une autre, pour cette population.

### 4.6 Remarques conclusives

L'analyse que nous avons menée dans ce chapitre a mis en évidence une évolution de la propension des ménages à réaliser des sorties pour le motif travail et pour les motifs autres que travail et enseignement. Ces évolutions sont de nature à biaiser toute prévision qui s'appuierait sur l'hypothèse classique de stabilité des paramètres des modèles statiques. Le biais est d'autant plus important que l'horizon de prévision est éloigné. Il semblerait que pour un horizon de prévision de 10 ans, les erreurs réalisées restent acceptables. En revanche, pour une prévision à 20 ans, elles sont trop importantes. En effet, nous avons montré que l'hypothèse de stabilité des paramètres

des modèles de régression aurait conduit en 1985 à une surestimation de la mobilité en 2006 de plus de 16 %, avec les modèles de régression zonale. Au niveau zonal l'erreur induite serait autour de 20 % des sorties.

L'erreur en prévision dépend des modèles utilisés. Les résultats les moins bons correspondent au modèle par régression à l'échelle des ménages. Les meilleurs correspondent au modèle par classification croisée. En effet, les évolutions constatées de manière agrégée au niveau du nombre de demi-sorties par personne (ou par personne employée dans le cas du motif travail) n'affectent pas tous les ménages de la même manière. Dans les cas étudiés ici, la flexibilité dans le choix des variables explicatives du modèles de classification semble l'emporter sur la possibilité d'utiliser des variables quantitatives dans les modèles de régression. En effet, l'inclusion de la variable cycle de vie dans les modèles par classification permet, à un certain degré, d'améliorer à la fois la capacité descriptive de ceux-ci, mais également leur capacité prédictive.

Toutefois, certaines évolutions restent sans explication à l'intérieur des modèles présentés ici. Afin d'améliorer la capacité prédictive des modèles, il conviendrait d'étudier plus en détail les déterminants de la complexification des sorties et de la réduction du nombre de demi-sorties pour le motif travail. Des études de ce type se multiplient dans les dernières années (Goulias, 1999 ; Harvey et Wilson, 2001 ; Raux *et al.*, 2011). Il serait ainsi possible de savoir si la tendance à la baisse de celui-ci est durable. Notons que la méthode de génération retenue dans le projet SIMBAD, de par son mécanisme de micro-simulation qui conserve la diversité des profils de mobilité, pourrait facilement être complétée par un module décrivant les évolutions dans le chaînage des activités. Par exemple, une modification des probabilités de tirage des « donneurs » de profils de mobilité pourrait être envisagée.

Nous nous sommes concentrés sur l'utilisation d'une seule enquête pour calibrer les modèles. Nous avons vu dans la section 2.2.1 que quelques travaux récents mettent en évidence que l'utilisation de données de plusieurs enquêtes dans la définition des paramètres de modèles statiques permet d'améliorer leur performance prospective. Il pourrait nous être reproché de ne pas avoir inclus dans notre étude des analyses utilisant plus d'une enquête dans le calage des modèles. En effet, il convient de signaler que, compte tenu des modèles estimés (pas de prise en compte des tendances temporelles par l'intermédiaire de variables retardées), les méthodes combinant plusieurs enquêtes ne produisent de meilleurs résultats qu'à condition d'observer un nombre de demi-sorties qui évolue de manière non monotone, autour d'une moyenne. Autrement dit, que dans les enquêtes utilisées il n'y ait pas une tendance bien définie à la hausse ou à la baisse du nombre de demi-sorties par unité décisionnelle. Si cette configuration peut avoir lieu lorsque l'on ne prend pas en compte les motifs de déplacement (ce qui est le cas des études combinant plusieurs enquêtes), on a vu dans ce chapitre que ce n'est pas le cas lorsqu'on prend en compte les motifs de déplacement. Par ailleurs, la combinaison d'enquêtes semble plus efficace dans les modèles de régression, qui comme nous l'avons vu ici, sont ceux qui font les prévisions les moins bonnes.

Il aurait également été intéressant de tester une spécification de modèle de type tobit, pour mieux prendre en compte l'impact des ménages qui n'ont pas fait de déplacement le jour précédant l'enquête.



---

## Distribution

La modélisation de la distribution peut être agrégée (comme dans les modèles de type gravitaire) ou désagrégée (comme dans les modèles combinés de choix de destination et choix de mode). Ce chapitre est consacré à l'étude de la modélisation agrégée de type gravitaire<sup>1</sup>. Il s'interroge sur la capacité du modèle de distribution gravitaire à intégrer les dynamiques du système urbain jouant dans les dix ou vingt ans que l'on cherche à simuler. En particulier, nous examinons la capacité du modèle à rendre compte de l'allongement des distances de déplacement urbains observées ces vingt dernières années.

Comme nous l'avons remarqué dans le chapitre 2, la littérature relative à la validation empirique, la performance prospective et, plus généralement, la transférabilité temporelle du modèle gravitaire est relativement réduite par rapport à celle consacrée aux autres modules de la séquence en quatre étapes (Ortúzar et Willumsen, 2001 ; Fox et Hess, 2010). Il est donc opportun de s'intéresser à l'étude de la validité de l'hypothèse de stabilité temporelle des paramètres eu celui-ci.

Notre étude de la question commence par la définition d'une méthodologie, pour ensuite présenter des résultats de la performance descriptive et prédictive du modèle gravitaire.

### 5.1 Méthodologie

L'objectif de cette étude est d'évaluer la capacité du modèle gravitaire à faire des prévisions sur un horizon temporel de moyen-long terme (10 ans et plus), et de voir s'il est capable de rendre compte de l'allongement des distances observé ces vingt dernières années. Pour ce faire, nous étudions le cas de l'agglomération lyonnaise à partir des données des enquêtes ménages déplacements (EMD) de 1985, 1995 et 2006 et d'une codification du réseau routier correspondant à ces trois dates et permettant d'avoir des temps de parcours comparables dans le temps (voir chapitre 3). Ces données permettent de tester le pouvoir prédictif du modèle *a posteriori*, autrement-dit de vérifier si les paramètres de calibration d'une année permettent de reproduire raisonnablement bien les données observées à une autre année. Cette

---

1. Le contenu de ce chapitre a été publié, avec quelques remaniements, dans Cabrera Delgado et Bonnel (2012).

section commence par une quantification de l'allongement des distances observé (5.1.1). Elle donne également les spécifications des modèles étudiés (5.1.2) et la méthode de calibration ainsi que les indicateurs permettant d'analyser les résultats (5.1.3).

### 5.1.1 Évaluation de l'allongement des distances observées

Le tableau 5.1 donne l'évolution des distances moyennes à vol d'oiseau d'une demi-sortie du domicile pour les motifs Travail et Loisirs. Il y a une distinction faite par périmètre de résidence des ménages (Cf. figure 3.1. Les deux premières lignes du tableau, pour chaque motif, donnent des évolutions de distance à périmètre de résidence constant (périmètre couvert par les EMD de 1985 et 1995). La dernière ligne donne les distances moyennes calculées à partir de l'ensemble des ménages enquêtés (en ne retenant que les sorties réalisées au sein du périmètre). L'augmentation des distances de déplacement peut se voir clairement, pour les motifs Travail et Loisirs, sur tous les périmètres. Les distances moyennes augmentent avec le temps mais également avec le périmètre choisi : les distances parcourues par les résidents des zones les plus éloignées du centre sont les plus grandes. Les distances moyennes entre l'origine et la destination pour le motif Travail sont supérieures à celles pour le motif Loisirs, en accord avec l'intuition selon laquelle on est prêt à aller plus loin lorsque le motif de la sortie est plus « contraint ».

TABLE 5.1 – Distances moyennes à vol d'oiseau en km des demi-sorties, pour les motifs Travail et Loisirs.

Date EMD	1985	1995	2006
Motif Travail			
Périmètre 1985	4,47	4,83	5,01
Périmètre 1995	-	5,17	5,47
Périmètre évolutif	4,47	5,17	7,28
Motif Loisir			
Périmètre 1985	2,88	2,97	3,03
Périmètre 1995	-	3,24	3,26
Périmètre évolutif	2,88	3,24	4,18

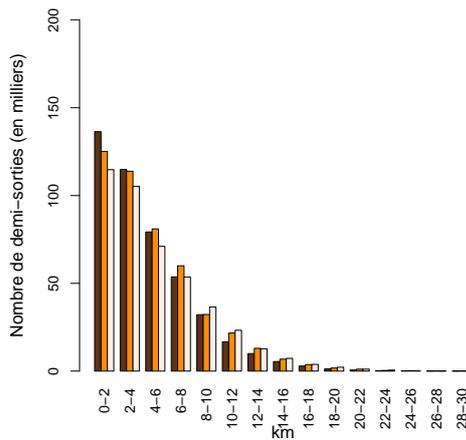
*Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006.*

*Périmètre : Cf. figure 3.1.*

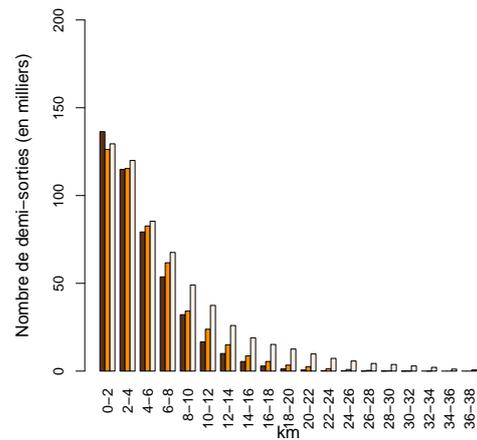
La figure 5.1 permet d'avoir le détail de la distribution des distances des demi-sorties. Pour le motif Travail, on observe que les sorties très courtes sont plus nombreuses en 1985 et, au fur et à mesure que la distance s'accroît, le nombre de sorties est plus important en 1995 et puis en 2006. Pour le motif Loisir, on observe une augmentation généralisée du nombre de demi-sorties avec le temps. En termes relatifs, elle est plus importante pour les distances plus longues, ce qui explique l'augmentation en moyenne constatée dans le tableau 5.1.

Il convient de remarquer que les deux motifs présentés ici sont ceux pour lesquels on observe un allongement des distances. Pour les autres motifs, on constate plutôt une stabilité entre 1985 et 2006. Dans un cadre d'évaluation des performance en prévision, notre présentation se focalisera sur ces deux motifs.

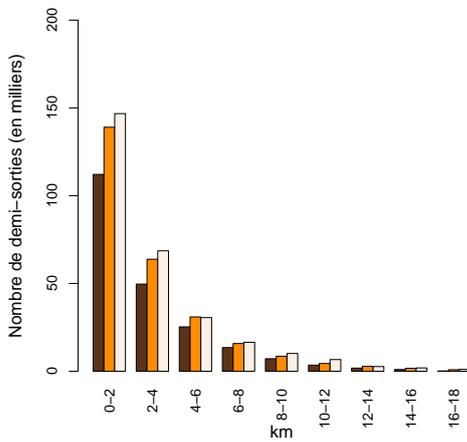
FIGURE 5.1 – Évolution de la distribution des distances à vol d’oiseau des demi-sorties, pour les motifs Travail et Loisir.



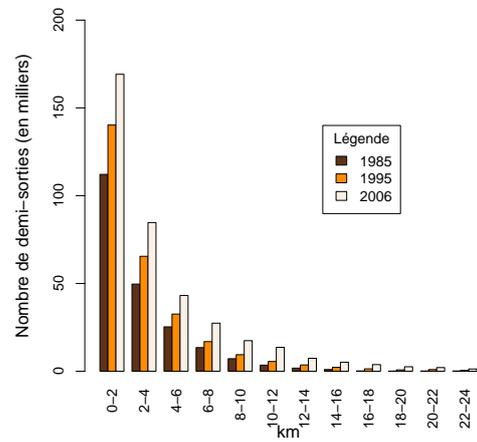
(a) Motif Travail, périmètre EMD 1985



(b) Motif Travail, périmètre évolutif



(c) Motif Loisirs, périmètre EMD 1985



(d) Motif Loisirs, périmètre évolutif

Source : EMD 1985, 1995 et 2006. Périmètre : Cf. figure 3.1.

### 5.1.2 Spécifications du modèle gravitaire segmentées par motif de déplacement

Dans la famille des modèles gravitaires, la formulation qui a la meilleure performance prospective est celle du modèle doublement contraint (Openshaw, 1979 ; Mozolin *et al.*, 2000). Cette supériorité dépend de la bonne estimation des marges de la matrice O-D dans la phase de génération. Dans la mesure où évaluer la transférabilité temporelle du modèle gravitaire revient à, dans un premier temps, tester le pouvoir prédictif du modèle *a posteriori*, c'est-à-dire en utilisant des variables en entrée effectivement observées à l'horizon de prévision, notre étude fonctionne comme si nous avons fait une prévision parfaite des émissions et attractions. L'utilisation du modèle doublement contraint est donc parfaitement adaptée et elle peut servir de benchmark pour d'autres types de modèles.

Les modèles gravitaires sont également caractérisés par la forme de la fonction de résistance retenue. Classiquement, on utilise une fonction exponentielle avec un seul paramètre. Celle-ci présente l'avantage de permettre une interprétation relativement aisée des évolutions paramétriques (voir, par exemple, les indications données par Elmi *et al.*, 1999 ; Mikkonen et Luoma, 1999). L'utilisation d'un seul paramètre peut ne pas être suffisante pour reproduire de manière acceptable la distribution des déplacements par tranche de distance ou temps. Le recours à une fonction avec deux paramètres (dite « exponentielle-puissance » ou Tanner, 1961) peut être avantageux dans un tel cas<sup>2</sup>.

Les spécifications du modèle à la Tanner sont données dans les équations (5.1) à (5.3):

$$T_{ij}^p = A_i^p O_i^p B_j^p D_j^p c_{ij}^{\alpha^p} \exp(\beta^p c_{ij}) \quad (5.1)$$

$$A_i^p = \left( \sum_j B_j^p D_j^p c_{ij}^{\alpha^p} \exp(\beta^p c_{ij}) \right)^{-1} \quad (5.2)$$

$$B_j^p = \left( \sum_i A_i^p O_i^p c_{ij}^{\alpha^p} \exp(\beta^p c_{ij}) \right)^{-1} \quad (5.3)$$

où

$T_{ij}^p$  est le nombre de demi-sorties ayant comme origine la zone  $i$  et ayant le motif principal  $p$  situé dans la zone  $j$  ;

$O_i^p$  est le nombre de demi-sorties ayant comme origine la zone  $i$  et ayant le motif principal  $p$  (il s'agit donc des émissions de la zone  $i$ ) ;

$D_j^p$  est le nombre de demi-sorties ayant la zone  $j$  comme localisation de leur motif principal  $p$  (il s'agit donc des attractions de la zone  $j$ ) ;

$c_{ij}$  est le temps généralisé pour aller de la zone  $i$  à la zone  $j$ .

$$c_{ij}^{\alpha^p} \exp(\beta^p c_{ij}) \quad (5.4)$$

est la fonction de résistance à deux paramètres,  $\alpha^p$  et  $\beta^p$ . Le modèle gravitaire exponentiel simple correspond au cas où le paramètre  $\alpha^p$  est nul. La fonction de

2. En réalité, à partir du moment où l'on a des matrices O-D tous modes (lorsque la distribution précède la répartition modale, par exemple), les demi-sorties courtes sont très nombreuses et les « pics » sont mal reproduits par la fonction de résistance exponentielle.

résistance (5.4) peut s'écrire comme indiqué en (5.5)

$$\exp(\beta^p c_{ij}) \tag{5.5}$$

Il convient d'avoir un modèle de référence par rapport auquel on puisse comparer la performance du modèle à la Tanner. On peut prendre un modèle simple à facteurs de croissance (Deming et Stephan, 1940 et Furness, 1965), qui correspond au cas où  $\alpha^p$  et  $\beta^p$  sont tous les deux nuls. Il s'agit d'un modèle purement théorique de référence dans la mesure où l'application des facteurs de croissance se fait d'habitude à partir d'une matrice O-D observée dans le passé.

### 5.1.3 Calibration et indicateurs

#### 5.1.3.1 Estimation des paramètres par la méthode du maximum de vraisemblance

Erlander et Stewart (1990) ont mis en évidence un principe d'équivalence mathématique entre les modèles dérivés suivant chacune des différentes approches théoriques évoquées dans la section 1.2.2. La traduction pratique de cette équivalence est que l'on peut supposer que l'une des approches est la plus adaptée à la modélisation, d'un point de vue théorique, et utiliser une autre approche pour effectuer les calculs. Ce principe d'équivalence sera utile au moment de proposer une méthode pour calibrer le modèle. Ainsi, nous pouvons supposer que notre modèle a été défini, de manière usuelle, à partir de la maximisation de l'entropie sous une contrainte de coût (ou de temps donnée par les équations (5.8) et (5.9)), et pour déterminer les estimateurs par maximum de vraisemblance, supposer que le nombre de demi-sorties observés entre  $i$  et  $j$ ,  $N_{ij}^p$ , suit une loi de poisson de paramètre  $T_{ij}^p$  de sorte que  $E(N_{ij}^p) = T_{ij}^p$ .

Batty et Mackie (1972) et Kirby (1974), ainsi que, plus tard, Erlander et Stewart (1990) et Sen et Smith (1995) ont montré qu'il suffit de résoudre les systèmes définis par les équations (5.6) à (5.8) et (5.6) à (5.9) pour trouver les estimateurs par maximum de vraisemblance des paramètres du modèle gravitaire doublement contraint avec une fonction de résistance de type exponentielle et Tanner<sup>3</sup>, respectivement.

$$\sum_j T_{ij}^p = O_i^p, \tag{5.6}$$

$$\sum_i T_{ij}^p = D_j^p, \tag{5.7}$$

$$\sum_{ij} c_{ij} T_{ij}^p = \sum_{ij} c_{ij} N_{ij}^p, \tag{5.8}$$

$$\sum_{ij} \log(c_{ij}) T_{ij}^p = \sum_{ij} \log(c_{ij}) N_{ij}^p \tag{5.9}$$

Les paramètres  $\alpha$  et  $\beta$  sont ainsi définis de manière unique.

Afin de résoudre les systèmes non linéaires décrits par les équations (5.6) à (5.9), on utilise la procédure MSC (modified scoring procedure) décrite dans Sen et Smith (1995, chapitre 5), que l'on a implémentée en utilisant le logiciel statistique R (R Development Core Team, 2011).

3. Ces équations sont, en effet, les conditions de premier ordre de la maximisation de la fonction de log-vraisemblance.

Il est nécessaire de préciser que les équations (5.8) et (5.9), connues comme contraintes sur les coûts (temps) moyens, jouent un rôle majeur dans la formulation théorique du modèle gravitaire (voir annexe A.1). Leur introduction donne la forme de la fonction de résistance. Malgré cela, les méthodes de calibration usuellement retenues pour le modèle ne sont pas toujours compatibles avec ces contraintes, ce qui pose des problèmes de cohérence interne du modèle. Dans le cas présent, la méthode retenue les prend en compte. Il s'agit d'une méthode de celles connues comme « ciblant le coût moyen » (Smith et Hutchinson, 1981).

### 5.1.3.2 Évaluation de la qualité des calages

La mesure de la qualité de la reproduction des matrices observées par le modèle calibré nécessite la définition de certains indicateurs pertinents. Pour ce faire, on s'est appuyé sur la littérature existante.

Le premier indicateur découle naturellement de la méthode de calage retenue : il s'agit de la log-vraisemblance.

La très grande majorité des études de validation empirique du modèle gravitaire utilisent au moins une statistique générale de comparaison « cellule à cellule » des matrices modélisées et observées, le plus souvent en rapport avec l'erreur quadratique moyenne. Nous avons retenu pour cette étude deux indicateurs de ce type. Le premier est le *SRMSE*<sup>4</sup> (standardized root mean square error), donné dans l'équation (5.10) suivante :

$$SRMSE = \frac{\sqrt{\frac{\sum_{ij} (N_{ij} - T_{ij})^2}{I \cdot J}}}{\frac{\sum_{ij} N_{ij}}{I \cdot J}} \quad (5.10)$$

où l'on a omis les motifs  $p$  pour faciliter la lecture.  $I$  et  $J$  sont respectivement le nombre d'origines et de destinations émettant ou attirant au moins un demi-sortie. Comme nous avons vu dans le chapitre 4, Fotheringham et Knudsen (1987) indiquent que l'utilisation du *SRMSE* est préférable à celle des autres types de RMSE car ces dernières sont très sensibles aux ordres de grandeur des données. Dans la mesure où nous étudions des modèles sur trois périmètres différents, cette statistique est ainsi pertinente. Sa limite inférieure est zéro — réplique parfaite des données — et elle est généralement inférieure à un — une valeur supérieure à un indique que l'erreur moyenne est plus grande que la valeur moyenne. La sensibilité du *SRMSE* aux erreurs est linéaire, ce qui lui confère, d'après Fotheringham et Knudsen, une supériorité par rapport aux autres mesures générales de distance « cellule à cellule » pour comparer un même modèle dans deux contextes différents.

La deuxième mesure de distance entre matrices est une forme particulière de *RMSE*, utilisée par Elmi *et al.* (1999). Nous l'avons introduite dans un souci de comparabilité des résultats. Elle prend la forme suivante :

$$RMSE = \left[ \frac{\sum_{ij} \frac{(N_{ij} - T_{ij})^2}{T_{ij}}}{\sum_{ij} T_{ij}} \right]^{\frac{1}{2}} \quad (5.11)$$

---

4. Racine carrée normalisée de l'erreur quadratique moyenne.

### 5.1.3.3 Évaluation de la transférabilité des modèles

L'analyse de la transférabilité prend appui en général sur les indicateurs de qualité de calage. Il s'agit de comparer les résultats des simulations faites avec les paramètres de calage correspondant à la date des données observées et ceux utilisant des paramètres d'une date antérieure.

Comme le notent McArthur *et al.* (2011), les revues récentes de la littérature relative à la transférabilité temporelle des modèles de prévision de la demande de transport (Ortúzar et Willumsen, 2001, sont évoqués explicitement mais on pourrait y ajouter Karasmaa, 2008 et Fox et Hess, 2010) retiennent deux indicateurs principaux : la statistique de test de transférabilité (*TTS*, du sigle en anglais) et l'indice de transfert (*TI*, *idem*)<sup>5</sup>. Cependant, McArthur *et al.* remarquent que l'indicateur *TTS* semble mal s'adapter à l'étude du modèle gravitaire en raison des non-linéarités inhérentes à celui-ci. On peut penser que cette raison a également conduit Elmi *et al.* (1999) à ne pas prendre en compte la statistique *TTS* et à privilégier l'utilisation d'autres indicateurs. Nous avons donc décidé de procéder de la même manière et de ne pas l'utiliser et de ne retenir que l'indice *TI*. Celui-ci (Koppelman et Wilmot, 1982) est moins une mesure d'égalité paramétrique qu'une mesure de performance relative :

$$TI_v(\theta_u) = \frac{LL_v(\theta_u) - LL_v(\theta_{REF})}{LL_v(\theta_v) - LL_v(\theta_{REF})} \quad (5.12)$$

où  $LL_u(\theta_v)$  est la valeur de la log-vraisemblance du modèle calibré à partir de données observées à la date  $v$ .  $\theta_v$  correspond aux paramètres issus de ce calage<sup>6</sup>.  $LL_v(\theta_u)$  est la valeur de la log-vraisemblance du modèle simulé à partir des paramètres de calage à une date antérieure  $u$  et des émissions, attractions et temps de la date d'application  $v$ .  $\theta_{REF}$  correspond aux paramètres d'un modèle de référence ou nul (facteurs de croissance). L'indicateur décrit le degré avec lequel la log-vraisemblance du modèle transféré améliore un modèle nul ou de référence, par rapport à l'amélioration apportée par un modèle développé dans le contexte d'application.

Les indicateurs donnés par (5.10) et (5.11) sont des mesures absolues de l'erreur produite. On peut les utiliser pour produire un indicateur permettant de comparer la performance des modèles en description et en prévision. Elmi *et al.* (1999) utilisent le *RATE* (relative aggregate transfer error) qui compare la valeur des erreurs issues d'une simulation faite avec les paramètres transférés à celle des erreurs issues du calage du modèle correspondant à la date d'application. Il est défini dans les équations (5.13) et (5.14) ci-dessous :

$$RATE_S = \frac{SRMSE_v(\theta_u)}{SRMSE_v(\theta_v)} \quad (5.13)$$

$$RATE_R = \frac{RMSE_v(\theta_u)}{RMSE_v(\theta_v)} \quad (5.14)$$

5. Les études sur la transférabilité des modèles portent très majoritairement sur les modèles désagrégés. Les exemples d'utilisation de ces indicateurs dans le cas de modèles gravitaires sont peu nombreux, à la connaissance des auteurs. On peut citer Elmi *et al.* (1999) et McArthur *et al.* (2011), notamment.

6. Soit  $\beta$ , pour le modèle avec une fonction de résistance exponentielle, et  $\alpha$  et  $\beta$ , pour celui avec une fonction de résistance de type Tanner.

correspondant aux deux indicateurs de distance entre matrices définis par (5.10) et (5.11), respectivement.

Pour l'étude de l'allongement des distances de déplacement, il convient de regarder la distribution de celles-ci. Notre approche diffère néanmoins de l'utilisation classique de cette distribution. En effet, la pratique habituelle est de regarder la distribution de la variable utilisée pour mesurer la « séparation » entre les zones de l'aire d'étude. Dans le cas présent, cela reviendrait à analyser la distribution des demi-sorties par tranche de temps ( $c_{ij}$ ). Or, l'objectif étant ici de voir si l'application du modèle permet de reproduire l'augmentation des distances moyennes de déplacement observées, il est plus intéressant de regarder la distribution des distances à vol d'oiseau entre les zones de l'aire d'étude. Il convient de rappeler que la méthode d'estimation retenue permet de reproduire le temps moyen observé, mais il n'y a aucune contrainte sur les distances à vol d'oiseau. L'indicateur qui s'impose naturellement pour étudier la proximité des distributions des distances observées et simulées est le  $\chi^2$  de Pearson donné par :

$$\chi_{dist}^2 = \sum_{d=1}^D \frac{(N_d - T_d)^2}{T_d} \quad (5.15)$$

où  $d$  est une classe de distance appartenant à  $\{1 \dots D\}$ .  $N_d$  et  $T_d$  sont le nombre de demi-sorties observées et modélisées, respectivement, appartenant à la classe de distance  $d$ .

## 5.2 Résultats

Pour présenter les résultats, il convient de procéder en deux temps. D'abord on présente les résultats de la calibration des modèles et ensuite leur performance en prévision.

Le modèle à deux paramètres donne systématiquement de meilleurs résultats que le modèle exponentiel classique. Ce résultat est conforme à ce que l'on attendait au vu des données retenues pour la constitution de la matrice O-D observée, celles-ci incluant les demi-sorties tous modes confondus. En conséquence, la présentation se focalise sur le modèle à la Tanner.

### 5.2.1 Résultats de calage

Le tableau 5.2 donne les résultats des calages du modèle gravitaire à la Tanner au niveau des IRIS2000. On y trouve les valeurs des coefficients  $\alpha$  et  $\beta$  ainsi que les indicateurs présentés dans la section 5.1.3.2.

Si l'on s'intéresse aux paramètres obtenus, le premier constat que l'on peut faire est qu'ils sont tous négatifs. Il s'agit encore de l'impact des sorties courtes. On observe également que les paramètres changent mais pas de manière uniforme. Pour le paramètre  $\alpha$ , l'écart le plus important est d'environ 30 % pour le motif Travail et 20 % pour le motif Loisir. Il est intéressant de remarquer que l'écart maximum se trouve pour des données de la même année (2006) pour le Loisir. Pour le paramètre  $\beta$ , l'écart le plus important est d'environ 25 % pour le motif Travail et 20 % pour le Loisir. Entre les motifs, les paramètres du motif Loisir sont plus grands en valeur absolue que ceux du motif Travail, ce qui, comme tous les paramètres sont négatifs, est en accord

TABLE 5.2 – Résultats des calages du modèle Tanner par année et périmètre pour les motifs Travail et Loisir (zonage IRIS2000).

EMD/Périmètre	$\alpha$	$\beta$	$LL_a(\theta_a)$	$RMSE$	$SRMSE$
<i>p</i> =Travail					
1985/P1985	-0,551	-0,085	-1,742E+04	1,07	2,52
1995/P1985	-0,468	-0,080	-1,820E+04	1,05	2,34
1995/P1995	-0,480	-0,078	-1,920E+04	1,07	2,4
2006/P1985	-0,551	-0,076	-2,020E+04	0,999	2,35
2006/P1995	-0,559	-0,075	-2,175E+04	1,03	2,37
2006/AUL	-0,622	-0,067	-2,881E+04	1,17	2,99
<i>p</i> =Loisir					
1985/P1985	-0,779	-0,134	-8,586E+03	1,9	6,85
1995/P1985	-0,735	-0,133	-1,039E+04	1,74	4,58
1995/P1995	-0,783	-0,124	-1,092E+04	1,86	4,79
2006/P1985	-0,681	-0,147	-1,253E+04	1,57	4,48
2006/P1995	-0,754	-0,136	-1,322E+04	1,85	4,65
2006/AUL	-0,831	-0,122	-1,674E+04	2,68	6,07

*Note de lecture : la première colonne du tableau spécifie les données utilisées : par exemple, à la deuxième ligne « 1995/P1985 » indique qu'il s'agit des données de l'EMD 1995, dont on n'utilise que la partie couvrant le périmètre de l'EMD 1985, dans la construction de la matrice observée pour le motif Travail.*

*Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006 et des réseaux modélisés.*

avec le fait que l'on se déplace plus longtemps pour aller travailler. Les variations à l'intérieur d'un motif sont moins importantes et souvent une augmentation en valeur absolue d'un paramètre correspond avec une variation dans le sens inverse de l'autre paramètre. Il est donc plus difficile d'en tirer une interprétation en relation avec une volonté de se déplacer plus longtemps.

La qualité de l'ajustement cellule à cellule, telle que mesurée par les indicateurs  $RMSE$  et  $SRMSE$  est moyenne. On obtient, dans la majorité des cas, des valeurs supérieures à 1, ce qui veut dire que l'erreur moyenne est supérieure à la valeur moyenne de la cellule. Cependant les valeurs de ces indicateurs ne sont pas très élevées compte tenu du fait qu'ils ont été établis pour un zonage très fin. Les valeurs obtenues sont du même ordre ou inférieures à celles qu'on a pu trouver dans la littérature (Elmi *et al.*, 1999, par exemple, obtiennent des valeurs pour le  $RMSE$  entre 7 et 15, alors qu'ici on obtient des valeurs allant de 0,999 à 2,68). Si l'on regarde motif par motif, on constate que l'ajustement cellule par cellule est bien meilleur pour le motif Travail. Les indicateurs  $RMSE$  et  $SRMSE$  sont environ deux fois plus grands pour le motif Loisir que pour le motif Travail.

## 5.2.2 Performance en prévision

Le tableau 5.3 donne des indicateurs de qualité d'ajustement permettant d'étudier la performance prospective des modèles de type Tanner calibrés en 1985 et appliqués en 1995 et 2006, sur les trois périmètres étudiés.

L'indice de transférabilité ( $TI$ ) est très élevé. Sa valeur minimale est supérieure à 0,95 et, pour certaines années/périmètres elle s'approche de 1 (les effets d'arrondis

font apparaître la valeur 1,000 dans ces cas). On peut en déduire que les modèles « transférés », c'est-à-dire ceux conservant les paramètres de calage d'une année antérieure à la date d'application, contiennent une grande partie de l'information obtenue lors du calage réalisé à la date d'application (1995 ou 2006), et ce malgré l'élargissement du périmètre d'étude.

Si on s'intéresse aux indicateurs relatifs  $RATE_R$  et  $RATE_S$ , on retrouve une particularité récurrente dans les études de transférabilité réalisées pour le modèle gravitaire : il existe certains cas pour lesquels les indicateurs sont inférieurs à 1, ce qui revient à dire que l'ajustement du modèle transféré est meilleur que celui du modèle calibré à la date d'application<sup>7</sup>. On peut expliquer ce type de configuration par le fait que les indices de type  $SRMSE$  ou  $RMSE$  accordent un poids plus important aux erreurs relatives que ne le fait la déviance normalisée (*scaled deviance*), qui est l'indicateur minimisé lors d'une estimation par maximum de vraisemblance. Autrement-dit, minimiser la déviance normalisée ne correspond pas tout à fait à minimiser le  $RMSE$ . Ceci n'est pas sans rappeler la différence entre le calage du modèle gravitaire par la méthode des moindres carrés et par maximum de vraisemblance.

Les valeurs de  $RATE_R$  et  $RATE_S$  sont proches de 1, ce qui permet de retrouver un résultat donné par Openshaw (1979) : la qualité des prévisions avec le modèle gravitaire est très proche de sa capacité descriptive, qui elle est moyenne. Dans le cas présent, on a établi que la qualité des calages pour le motif Travail était meilleure que celle pour le motif Loisirs. Il s'en suit que la qualité des prévisions avec les coefficients de 1985 est meilleure pour le motif Travail que pour le motif Loisir.

Si l'on s'intéresse aux distances, l'application des coefficients de calage de 1985 aux marges et temps de déplacement de 1995 et 2006 donne une très bonne prévision. Pour le motif Travail, on sous-estime légèrement la distance moyenne. Elle n'est sous-estimée que d'environ 8 % pour le motif Travail quand on applique les coefficients de 1985 aux données de 2006, sur un périmètre qui est bien plus large que celui sur lequel les coefficients ont été estimés. Lorsque les périmètres de calage et d'application sont plus proches, la sous-estimation est bien plus faible. Pour le motif Loisirs, on observe à la fois de légères surestimations et sous-estimations.

Pour étudier la distribution des distances, on a retenu des classes de 1 km, lorsque l'effectif le permettait. La dernière classe regroupe les demi-sorties pour les distances « très longues » (distances au-delà d'un certain seuil défini en fonction de l'effectif). Le  $\chi^2$  appliqué aux classes de distance donne des résultats très satisfaisants pour le motif Travail, tant au niveau de la matrice simulée lors du calage que de la matrice simulée à partir des coefficients de 1985. Le calage, reproduit bien la distribution des distances pour l'ensemble des années/périmètres. En prévision, on a un résultat similaire. Il faut cependant noter que, pour la distribution des distances avec les données de 2006 sur le périmètre de l'aire urbaine, l'hypothèse selon laquelle les écarts entre la distribution observée et la distribution modélisée peuvent provenir des aléas liés à l'échantillon ( $H_0$ ) est acceptée de très peu, avec un niveau de confiance de 95 %. Pour le motif Loisir, les résultats sont moins satisfaisants : l'hypothèse  $H_0$  est rejetée dans 4 cas sur 5 pour les distances simulées lors du calage et dans 3 cas sur 5 en prévision. Les cas où l'hypothèse est acceptée correspondent à ceux où le contexte

---

7. Ce type de comportement « surprenant » peut être observé dans Openshaw (1979) et McArthur *et al.* (2011), entre autres.

TABLE 5.3 – Performance prospective du modèle de type Tamer avec les coefficients du calage de 1985/P1985 (zonage IRIS2000).

EMD/Périmètre	$\overline{\Delta dist}$ (%)	TI	RATE <sub>R</sub>	RATE <sub>S</sub>	$\chi^2$ calage		$\chi^2$ prévision avec coeffs de 1985		
					Valeur	$P(X < x)$	Valeur	$P(X < x)$	
<i>p</i> =Travail									
1995/P1985	1,942	0,989	1,035	1,032	14,126	0,721	12,316	0,83	
1995/P1995	3,186	0,988	1,042	1,029	13,784	0,909	14,908	0,866	
2006/P1985	1,078	0,994	1,020	1,025	12,144	0,84	6,423	0,994	
2006/P1995	2,763	0,992	1,033	1,026	11,485	0,967	9,829	0,988	
2006/AUL	8,185	0,984	1,102	1,023	19,123	0,981	48,441	0,0516	
<i>p</i> =Loisirs									
1995/P1985	-2,781	0,999	1,028	0,999	22,776	0,064	15,912	0,319	
1995/P1995	1,967	0,998	1,090	0,996	28,958	0,0106	21,175	0,0972	
2006/P1985	-7,737	0,999	0,974	1,005	36,796	0,000792	40,989	0,000179	
2006/P1995	-4,432	1,000	0,996	1,001	45,657	3,19E-05	43,856	6,25E-05	
2006/AUL	3,092	0,999	1,305	0,997	60,419	1,94E-05	57,523	5,16E-05	

Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006 et des réseaux modélisés.

FIGURE 5.2 – Capacité des matrices simulées, à partir du modèle à la Tanner, à reproduire la distribution des distances observées, pour le motif Travail.

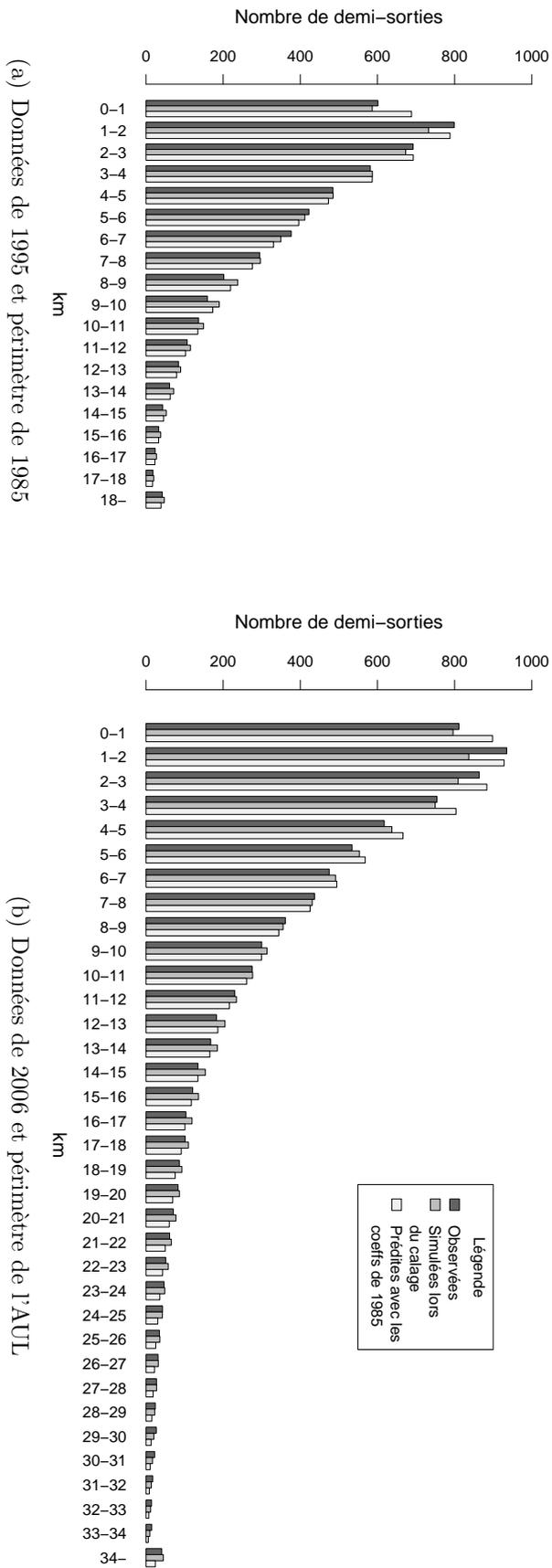
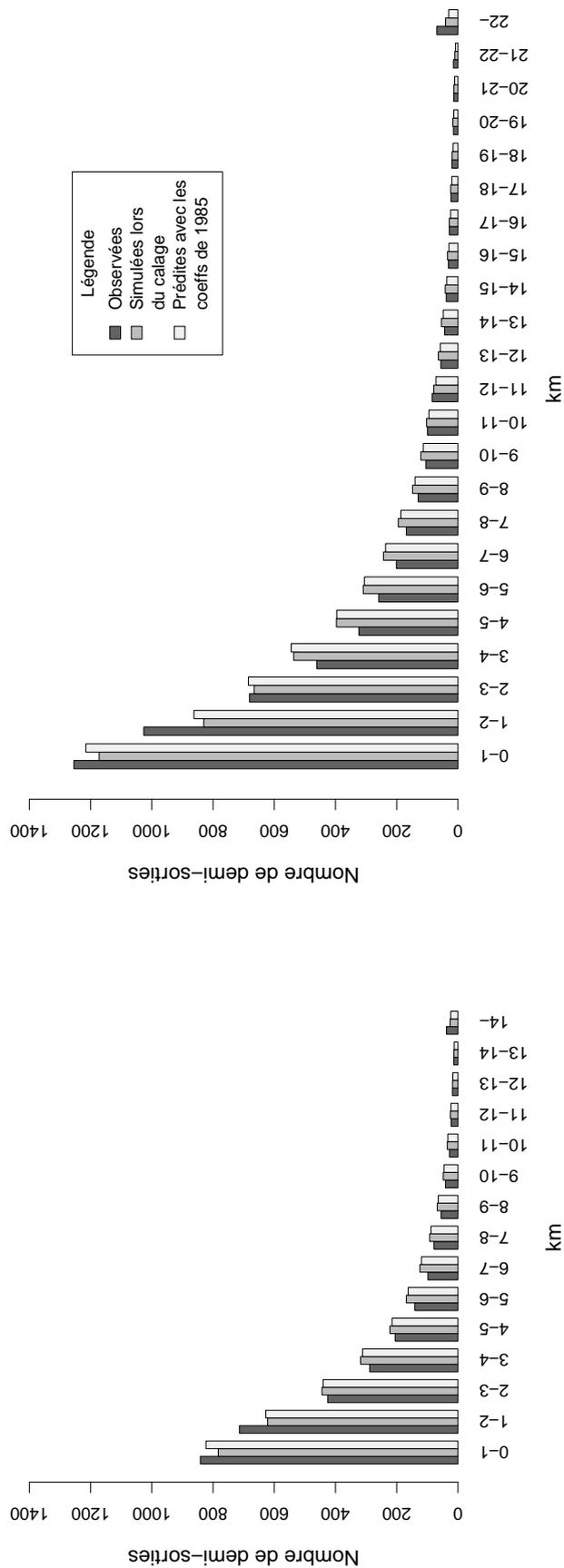


FIGURE 5.3 – Capacité des matrices simulées, à partir du modèle à la Tanner, à reproduire la distribution des distances observées, pour le motif Loisir.



(a) Données de 1995 et périmètre de 1985

(b) Données de 2006 et périmètre de l'AUL

Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006 et des réseaux modélisés.

de prévision est plus proche de 1985 (données de 1995 pour le périmètre de 1985 et, en prévision, pour le périmètre de 1995). Les figures 5.2 et 5.3 donnent les distributions des distances entre l'origine et la destination des demi-sorties observées, simulées à la date d'application et prédites à partir des coefficients de 1985, pour les motifs Travail et Loisirs, respectivement. Elles contiennent, pour les deux motifs de déplacement, le cas d'une prévision à dix ans sur le même périmètre et le cas d'une prévision à vingt ans, sur un périmètre élargi. Il s'agit des distributions qui ont permis de calculer les  $\chi^2$  correspondants, présentés dans le tableau 5.3. Les figures 5.2(a), 5.2(b) et 5.3(a) correspondent à des cas où la reproduction des distances est bonne. Les écarts entre le nombre de déplacements observés, modélisés lors du calage et prédits avec les coefficients de 1985 sont faibles. La figure 5.3(b) correspond à un cas où le test du  $\chi^2$  a rejeté  $H_0$ . On observe que les distributions sont relativement proches à l'exception de certaines classes de distance pour lesquelles les écarts sont importants. Par exemple, pour la classe de distance allant de 1 à 2 km, les demi-sorties modélisées lors du calage et prédites sont nettement inférieures aux demi-sorties observées. Inversement, pour les classes comprises entre 3 et 8 km, les demi-sorties modélisées lors du calage et prédites sont plus nombreuses que les demi-sorties observées.

La qualité du calage semble donc aller de pair avec la capacité du modèle gravitaire à reproduire les distances de déplacement. Pour le motif Travail, la fonction de résistance à la Tanner permet d'obtenir une qualité de calage relativement meilleure que pour le motif Loisir. En prévision, cette différence est retrouvée : les distances de déplacement sont bien reproduites pour le motif Travail mais moins bien pour le motif Loisir.

### 5.3 Conclusion

Le modèle gravitaire reste depuis plus de 40 ans le modèle le plus utilisé dans la modélisation transport tant pour des horizons de court/moyen terme (de l'ordre d'une dizaine d'années) que pour des horizons temporels plus lointains. Nous retrouvons les résultats déjà présentés par de nombreux auteurs à savoir que la qualité de l'estimation est souvent jugée moyenne. Ce résultat n'est pas étonnant dans la mesure où l'utilisation de zonages fins pour limiter le poids des intrazonnes conduit à utiliser des matrices observées de piètre qualité pour estimer les modèles. Les observations de chacune des cases de la matrice sont connues avec des intervalles de confiance qui sont extrêmement larges et la matrice comporte un très grand nombre de zéros pour lesquels on ne sait pas s'il n'y a effectivement aucun déplacement ou s'il y a seulement peu de déplacements. En revanche, la matrice estimée par construction ne comportera aucun zéro dès lors que les marges de la matrice n'en comporte pas non plus (Kirby, 1979). Le degré de reproduction des flux entre les différentes O-D peut donc ne pas être très bon et n'est pas forcément un indicateur de la performance prédictive. Dès lors, il ne semble pas raisonnable d'espérer que l'utilisation du modèle gravitaire en prévision puisse conduire à des résultats meilleurs que ceux obtenus lors de l'estimation des paramètres. De fait, l'analyse des résultats a montré que le degré d'erreur dans les prévisions est similaire à celui de l'estimation. Cette étude corrobore donc les résultats de travaux précédents.

C'est la raison pour laquelle la qualité de l'estimation est également mesurée à

travers d'autres indicateurs plus agrégés. En particulier, les courbes de distance de déplacements (en temps généralisé) sont fréquemment utilisées car le respect des distributions est souvent considéré comme important au regard des objectifs des modélisations transport. Compte tenu de notre problématique, qui est d'analyser la capacité du modèle gravitaire à rendre compte de l'allongement des distances (euclidiennes) de déplacement, nous avons analysé ce dernier indicateur au lieu du temps généralisé. Les résultats consignés dans la table 5.3 et les figures 5.2 et 5.3 montrent que le modèle gravitaire permet à la fois de bien reproduire une situation de référence et de proposer une projection correcte à un horizon de 20 ans. Les indicateurs que nous avons retenus montrent en effet que la qualité de reproduction de la courbe de distances de déplacements observées en 2006 est aussi bonne lorsque l'on simule la distribution avec les paramètres de calage de 1985 que lorsque l'on utilise les paramètres de 2006. Le résultat le plus intéressant et qui n'est absolument pas intuitif est que ce constat reste valable lorsqu'on considère le périmètre très élargi de l'enquête de 2006 et non plus seulement celui beaucoup plus restreint de l'enquête de 1985. La capacité prédictive reste donc valable même en cas d'élargissement du périmètre d'étude.

Ces résultats sont valables pour les deux motifs, Travail (contraint) et Loisir (non contraint). Toutefois, la qualité de l'estimation étant nettement moindre pour le motif Loisir, la qualité de la prévision des courbes de distribution des distances de déplacement l'est également moins. Il convient toutefois de nuancer ces résultats.

Ils correspondent à un exercice de prospective effectué dans un passé connu et donc de prévision parfaite des variables en entrée du modèle. Les bons résultats obtenus en prévision signifient que les variations paramétriques du modèle jouent peu et que la bonne prévision des variables en entrée de celui-ci — émissions et attractions, issues de la phase de génération, et niveau de service des réseaux — est déterminante. On peut également en tirer un autre enseignement extrêmement intéressant. L'évolution des émissions et attractions estimées lors du modèle de génération résultent de l'évolution de la localisation des ménages et des activités. L'évolution des temps généralisés résultent de l'évolution de la performance des réseaux de transport. L'allongement des distances (table 5.1) que l'on a observé au cours des 20 dernières années résulte donc non pas d'un besoin (subi ou souhaité selon les interprétations) d'aller toujours plus loin ou d'élargir son espace d'activité mais bien de l'évolution des localisations et de la performance des réseaux de transport. On aurait donc pu prévoir l'allongement des distances observé ces vingt dernières années, mais à condition de bien prévoir l'évolution des localisations et les flux émis et attirés qui en résultent, ainsi que l'évolution des niveaux de service des réseaux de transport. On retrouve donc l'enjeu des modèles d'interaction transport-urbanisme, auxquels le modèle gravitaire de distribution semble pouvoir s'intégrer.



---

## Choix modal

La détermination des matrices O-D par mode peut se faire de deux manières. La première est agrégée : elle utilise des modèles dont l'unité d'analyse est la paire de zones (origine et destination). L'application de ces modèles — dits de **répartition** modale pour insister sur leur caractère agrégé — permet de déterminer directement les parts des modes pour chaque paire d'O-D de la matrice issue de la distribution. La deuxième méthode est désagrégée : elle met à l'œuvre des modèles de choix discrets appliqués au choix du mode — on parle donc, non plus de répartition, mais de **choix** modal. Ces derniers sont calibrés à l'échelle des unités de mesure de la mobilité et des unités décisionnelles. Il est ensuite nécessaire de définir des règles d'agrégation des résultats pour définir la répartition au niveau de la matrice O-D. Ce chapitre est consacré aux approches désagrégées, qui concernent actuellement la grande majorité des modèles utilisés pour la détermination des parts modales. Cette popularité est liée à trois facteurs fondamentaux. Le premier est celui d'une plus grande pertinence théorique puisqu'on analyse le choix de mode à l'échelle de l'unité décisionnelle. Les modèles ont donc un caractère comportemental<sup>1</sup>, ce qui facilite l'interprétation des paramètres. Le deuxième facteur, évoqué principalement par les tenants de l'approche désagrégée, découle du premier. En effet, Ben-Akiva et Atherton (1977), Train (2009) et Fox et Hess (2010), entre autres, notent que les prévisions des modèles désagrégés sont meilleures que celles des approches agrégées en raison de la connaissance plus fine des causes des changements. Cet argument fait référence, plus précisément, à deux éléments. Le premier est que les modèles désagrégés sont plus sensibles aux variables influant sur le choix. Le second est la supposition — proche de la croyance — que la représentation plus fine du processus de décision conduit intrinsèquement à établir des relations plus stables dans le temps que les relations à une échelle plus agrégée<sup>2</sup>. Le troisième facteur est d'ordre pratique. Le fait de travailler à l'échelle des unités décisionnelles permet d'inclure plus de variables dans le modèle explicatif, ce qui le rend plus apte à étudier les impacts de politiques de transport variées. Dans ce qui

---

1. Notons cependant que les comportements restent inscrits dans le cadre habituel de choix rationnels d'individus optimisateurs.

2. Un des exemples de référence, avancé par les tenants des approches désagrégées, concerne la qualité supérieure des prévisions de parts modales réalisées dans le cadre des travaux préliminaires du projet BART (*Bay Area Rapid Transit*) par les modèles désagrégés (prévisions faites par des chercheurs) que par l'approche agrégée (prévisions des praticiens effectivement utilisées) dans l'aire métropolitaine de San Francisco, aux États-Unis. Voir, à ce sujet, McFadden (2001) et Train (2009).

suit, nous présentons, dans un premier temps, l'évolution des comportements de choix de mode dans le périmètre du Grand Lyon, qui correspond à l'espace desservi par les transports urbains et la méthodologie pour les analyser. Dans un second temps, nous présentons les modèles de choix modal, calibrés en coupe instantanée, pour déterminer si ceux-ci auraient été capables de prévoir les évolutions observées entre 1985 et 2006 et ainsi évaluer la pertinence de l'hypothèse de stabilité paramétrique.

## 6.1 Méthodologie pour comprendre l'évolution du choix du mode

### 6.1.1 Définition du mode de déplacement au sein d'une sortie

L'unité de mesure de la mobilité retenue lors des chapitres 4 et 5 est la même que celle qui sera retenue pour l'analyse de la répartition modale : la sortie (ou plutôt la demi-sortie). On a vu dans la section 4.1.1.1 qu'un des principaux avantages du choix de la sortie comme unité d'analyse pour la mobilité est la possibilité de conserver la cohérence modale interne à la sortie. Plus précisément, dans une sortie complexe — c'est-à-dire une sortie permettant de réaliser plus d'une activité — certaines des activités à réaliser conditionnent le choix de mode pour l'ensemble de la sortie (Kitamura, 1984 ; Bonnel, 2004). Imaginons, par exemple, le cas d'une personne seule envisageant de faire ses courses pour la semaine, lors de son retour du travail. Cette personne dispose d'un véhicule particulier mais elle a également un bon accès aux transports en commun ce qui fait qu'elle peut aussi utiliser ce mode pour se rendre au travail. Dans cette configuration, le fait de devoir faire des courses au sein de la sortie peut avoir pour conséquence que la personne privilégie l'usage de son véhicule particulier pour réaliser les trois déplacements composant la sortie. Le mode choisi pour l'aller est en général le même que celui pour le retour (notamment lorsque le mode considéré est l'automobile).

Si le choix de la sortie comme unité de mesure de la mobilité est pertinent en raison de la cohérence modale au sein des sorties, il pose une difficulté, similaire à celle rencontrée pour le traitement des motifs de déplacement (section 4.1.1), pour ce qui est de la définition d'un mode de déplacement. En réalité, dans un environnement donné, il existe un ensemble de solutions techniques permettant aux individus de se déplacer. Le mode de déplacement est une combinaison de ces solutions techniques (systèmes de transport) associée à une unité de mesure de la mobilité<sup>3</sup>. On peut utiliser un exemple simple pour illustrer ce propos. On suppose l'existence d'un environnement où il n'est possible de se déplacer que par l'intermédiaire de deux solutions techniques : à pied et en voiture. On suppose également que l'on souhaite mener, dans cet environnement, une analyse de la mobilité à l'échelle du déplacement. Celui-ci est donc l'unité de mesure de la mobilité pour laquelle on souhaite définir des modes de déplacement. On peut considérer deux cas de figure. Dans le premier, on définit trois modes. Un premier mode correspondrait au cas où la totalité du

---

3. Et, inversement, la définition d'une unité de la mesure de la mobilité fait souvent appel à la spécification d'un mode, comme, par exemple, la définition du déplacement par le CERTU (2008).

déplacement s'est effectuée en marchant. Un deuxième mode correspondrait au cas où la totalité du déplacement sur la voirie publique s'est réalisée en voiture (il s'agit du cas où le lieu d'origine et celui de destination ont tous les deux un parking intégré). Le troisième mode correspondrait aux déplacements combinant les deux systèmes (par exemple, s'il y a une phase de marche sur la rue pour pouvoir se rendre au lieu de stationnement au départ ou à l'arrivée). Dans le deuxième cas de figure, on définit seulement deux modes de déplacement. Le premier mode correspondrait au cas où la seule solution technique déployée pour se déplacer est la marche à pied. Le deuxième mode correspondrait au cas où l'on utilise la voiture pour au moins une partie du déplacement, le reste, s'il y a lieu, étant réalisé à pied. Cet exemple simple illustre à quel point la définition du mode est complexe (arbitraire) et doit s'adapter aux objectifs de l'analyse. Ainsi, le recours à la définition proposée dans le premier cas de figure serait plus approprié que l'utilisation de celle proposée dans le deuxième cas si l'on souhaitait, par exemple, évaluer l'impact, au niveau de l'usage de la voiture, d'une politique réduisant les places de stationnement dans certains quartiers et obligeant les individus à marcher un peu plus pour se rendre à leur voiture. Notons que le problème de la définition du mode gagne en complexité si l'environnement dans lequel se déroule l'analyse comprend un nombre plus important de solutions techniques de transport. Dans le cas présent, les données disponibles dans les EMD permettent d'avoir un ensemble assez large de solutions techniques et donc un ensemble encore plus large de modes possibles pour les demi-sorties. Pour garantir une comparabilité des données et pour simplifier le problème de choix modal au niveau de la modélisation, nous avons regroupé les solutions techniques de déplacement proposées dans les EMD au sein des catégories suivantes<sup>4</sup> :

TC : les transports en commun de l'agglomération lyonnaise qui, compte tenu de la codification des réseaux qui a été faite, se limitent à ceux dont la gestion dépend du SYTRAL (les trains en sont donc exclus) ;

VP : les véhicules particuliers, que ce soit en tant que conducteur ou passager ;

2RM : les 2 roues à moteur, en tant que conducteur ou passager ;

VELO : les vélos ;

AUTRE : les systèmes de transport ayant été codifiés comme « autres » ;

MAP : la marche à pied.

Ces systèmes de transport fonctionnent avec un contrôle plus ou moins important des individus sur les itinéraires, heures auxquelles les déplacements sont possibles, les temps de parcours, les coûts, etc. L'ordre selon lequel ils ont été donnés dans la liste ci-dessus correspond à peu près à un classement selon que l'individu a plus ou moins de contrôle sur les conditions de son déplacement. Le système de transport TC est, bien entendu, le plus contraint car son organisation est indépendante d'un individu isolé. Pour l'utilisation de ce système, l'individu doit organiser sa mobilité en fonction des conditions imposées par l'entité en charge de l'administration du réseau. La marche à pied est, inversement, le système de transport le moins contraint.

À partir de ces systèmes de transport, il est possible de définir des modes de déplacement pour chaque sortie. De manière analogue à celle présentée dans le deuxième

---

4. Issues du didacticiel de formation à la prévision de la demande de transport accompagnant le manuel de Bonnel (2004), à partir de la méthode de définition des modes de déplacement dans les EMD (CERTU, 2008).

cas de figure de l'exemple avec deux systèmes de transport, on peut combiner les systèmes de transport ci-dessus pour définir le mode de déplacement d'une sortie (ou demi-sortie). Le passage des systèmes de transport inclus dans une sortie au mode de cette sortie est fait en retenant le système de transport le plus contraint parmi ceux utilisés. Ainsi, par exemple, on dira qu'une sortie est réalisée en transports en commun — c'est-à-dire que le mode de déplacement de la sortie est TC — dès lors qu'au moins une partie de la sortie aura été réalisée en TC. Ceci inclut, entre autres, les sorties pour lesquelles l'accès aux stations a été réalisé en voiture (notons que ceux-ci sont très peu nombreux). Le mode d'une sortie sera VP si au moins une partie de celle-ci a été réalisée en utilisant ce mode de transport et que le système TC n'a pas été utilisé (auquel cas le mode de la sortie serait TC). Cette même logique est respectée pour les autres systèmes de transport, en suivant l'ordre de la classification présentée dans ce qui précède.

Au niveau de la définition des modes, il convient de discuter plus en détail le choix d'un système de transport VP qui regroupe à la fois les conducteurs et passagers de véhicules particuliers. En effet, les trafics routiers sont directement liés au nombre d'utilisateurs des véhicules particuliers en tant que conducteurs (un véhicule étant, trivialement, toujours conduit par une personne). Or, la définition du mode VP réalisée ici ne permet pas de distinguer entre conducteurs et passagers ; elle ne permet donc pas d'avoir directement un trafic à affecter dans un réseau. Dans le cadre d'un modèle utilisé dans la pratique, il serait nécessaire d'inclure un module supplémentaire permettant de déterminer la part de conducteurs parmi l'ensemble des usagers du mode VP. En réalité, le choix méthodologique de combiner les deux types de déplacements est nécessaire pour garder la propriété de cohérence modale de la sortie. Lorsque des individus pouvant conduire se déplacent ensemble (pour, par exemple, aller faire des courses), ils peuvent prendre le rôle de conducteur pour des parties de la sortie.

### 6.1.2 Un regain de parts de marché pour les TC

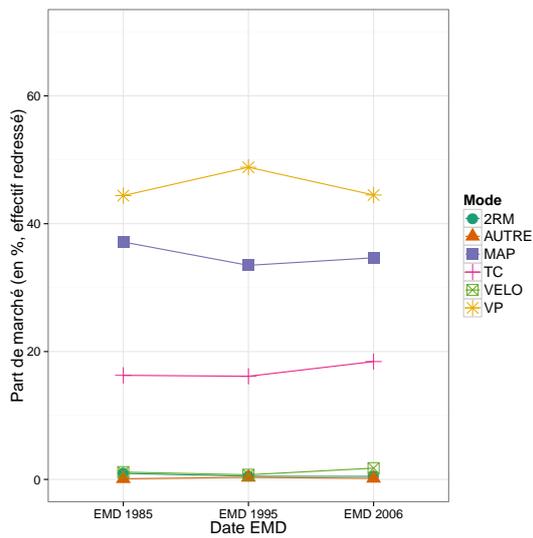
La figure 6.1 donne les évolutions des parts modales des demi-sorties des résidents du Grand Lyon — à partir de la définition des modes décrite dans le point précédent — réalisées dans ce même périmètre. La figure 6.1(a) donne les évolutions pour l'ensemble des motifs. On observe que l'ordre d'utilisation des modes n'a pas évolué entre 1985 et 2006 : VP et MAP restent les deux alternatives principales, suivies par les TC. Les parts des autres modes restent très faibles, même si on peut noter une utilisation plus fréquente du VELO en 2006. La part de la VP augmente entre 1985 et 1995 (de 44,4 % à 48,4 %, soit une augmentation relative d'environ 10 %) puis diminue entre 1995 et 2006 (pour revenir à peu près au niveau de 1985). La part de la MAP diminue entre 1985 et 1995 (de 37,1 % à 33,5 %, soit une baisse relative d'environ 10 %) puis s'accroît très légèrement (augmentation relative d'environ 3 %). Pour les TC, on observe une stabilité entre 1985 et 1995 puis une augmentation de leur part entre 1995 et 2006 (de 16,1 % à 18,4 %, soit environ 14 % en termes relatifs). Ces constats avaient été présentés comme l'un des résultats phares de l'EMD 2006 (SYTRAL, 2007a,b) car ils permettaient de montrer que les objectifs du PDU de 1997 avaient été atteints, voir

dépassés<sup>5</sup>. Il convient de remarquer, par ailleurs, que cette baisse de la part de la VP n'est pas seulement observée au niveau de l'agglomération lyonnaise mais également dans d'autres agglomérations françaises où, de manière générale, on a une baisse de l'usage de la voiture dans les quartiers centraux des villes (CERTU, 2012). Les figures 6.1(b), 6.1(c) et 6.1(d) précisent les évolutions en segmentant les demi-sorties selon trois motifs : Travail, Enseignement et Autres que Travail et Enseignement, respectivement. Pour le motif Travail, le mode prépondérant est VP (part de marché de plus de 60 % pour les trois années, ce qui représente presque 1,5 fois plus que la part moyenne de la VP pour l'ensemble des demi-sorties). Alors que la part de la VP est bien plus importante pour le motif Travail, les évolutions, évaluées en termes relatifs, sont d'ordres de grandeur similaires (augmentation relative d'environ 11 % entre 1985 et 1995 puis diminution de presque 13 % entre 1995 et 2006). Les TC sont le deuxième mode le plus utilisé devant la MAP. Pour les TC, on observe une diminution relative d'environ 10 % (de 17,4 % à 15,5 % de part de marché) entre 1985 et 1995 puis une forte augmentation entre 1995 et 2006 (passage à 21,7 % soit un gain relatif d'environ 40 %). Pour la MAP, on observe une baisse importante entre 1985 et 1995 (de 16,6 % à 12,9 % soit une baisse relative d'environ 22 %) puis une stabilisation entre 1995 et 2006. Le motif enseignement est un peu particulier car il concerne une population qui ne peut pas conduire (élèves du primaire et du secondaire) et une population faiblement motorisée (étudiants). Par ailleurs, certaines contraintes administratives (carte scolaire) limitent la distance à parcourir pour se rendre au lieu d'études. Il s'en suit que le mode le plus utilisé est la MAP suivie des TC et puis de la VP. La part de marché de la MAP en 1985 était de 52,4 %. Elle a subi une diminution relative d'environ 14 % entre 1985 et 1995 puis de seulement 2,5 % entre 1995 et 2006. L'évolution de la part des TC est quasi-symétrique puisque celle-ci s'est accrue d'environ 17 % en termes relatifs (de 27,4 % à 32,1 %) entre 1985 et 1995 puis elle est passée à 33,5 % en 2006 (augmentation relative d'environ 4 %). La part de la VP s'est accrue d'environ 13 % entre 1985 et 1995 pour ensuite rester stable autour de 21 % entre 1995 et 2006. Pour les motifs autres que Travail et Enseignement, les évolutions sont très similaires à celles correspondant à l'ensemble des motifs, à cette exception près que la part de marché de la MAP est environ 6 % plus importante pour toutes les années alors que celle des TC est moins importante d'environ 6 %.

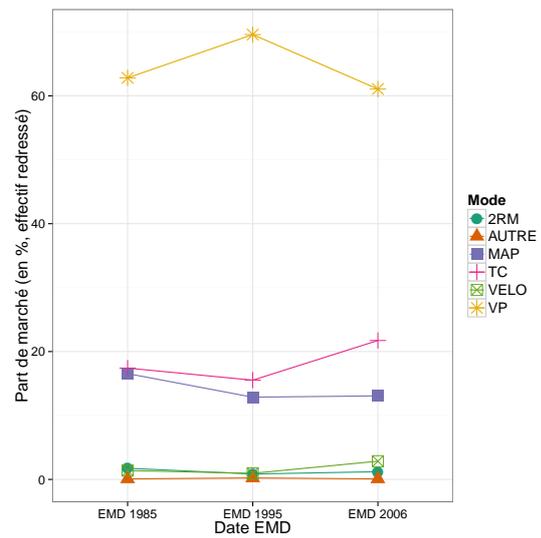
La décomposition par motifs des parts modales donnée dans la figure 6.1 laisse entrevoir un impact non négligeable, en termes de choix de mode, de la distance entre le domicile et le lieu de réalisation de l'activité principale. La figure 6.2 présente l'évolution des parts modales en fonction de la distance à vol d'oiseau entre le domicile et le lieu de réalisation de l'activité principale. Comme précédemment, la figure 6.2(a) correspond à l'ensemble des motifs, qui sont ensuite segmentés dans les figures 6.2(b) (motif Travail), 6.2(c) (motif Enseignement) et 6.2(d) (motifs autres que Travail et Enseignement). La MAP est majoritaire pour les demi-sorties courtes : elle représente près de 75 % des demi-sorties de moins de 1 km, pour l'ensemble des motifs, pour les trois années d'enquête. C'est également le cas pour les demi-sorties réalisées pour le motif Enseignement ainsi pour les motifs autres que Travail et Enseignement. Pour

5. « L'ambition affichée par le PDU de 1997 était "un renversement de tendance dans la répartition des déplacements entre les différents modes". Elle a été largement réalisée sur l'ensemble du Grand Lyon [...] » (SYTRAL, 2007a, page 12). La notion de « renversement de tendance » fait allusion principalement à deux choses : limiter le déclin de la MAP et augmenter l'attractivité des TC.

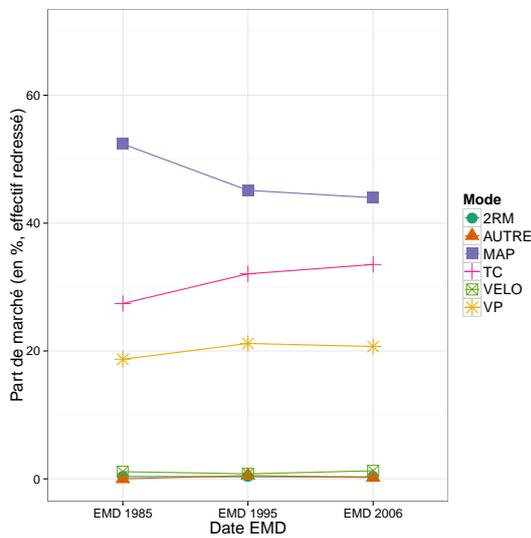
FIGURE 6.1 – Évolution des parts modales des demi-sorties des résidents du Grand Lyon, réalisées à l’intérieur de ce périmètre. Segmentation par motif.



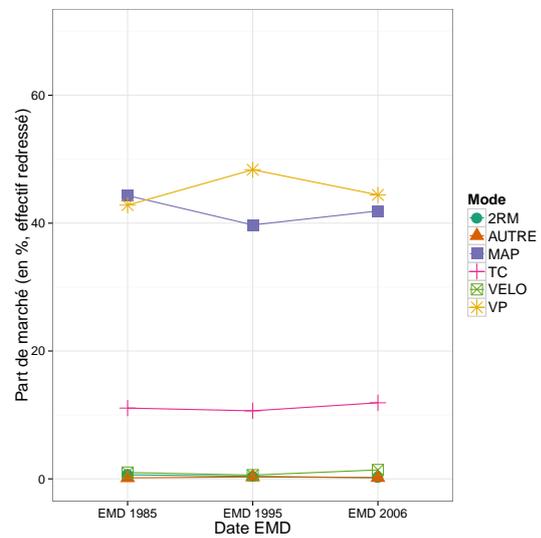
(a) Tous les motifs confondus.



(b) Travail.



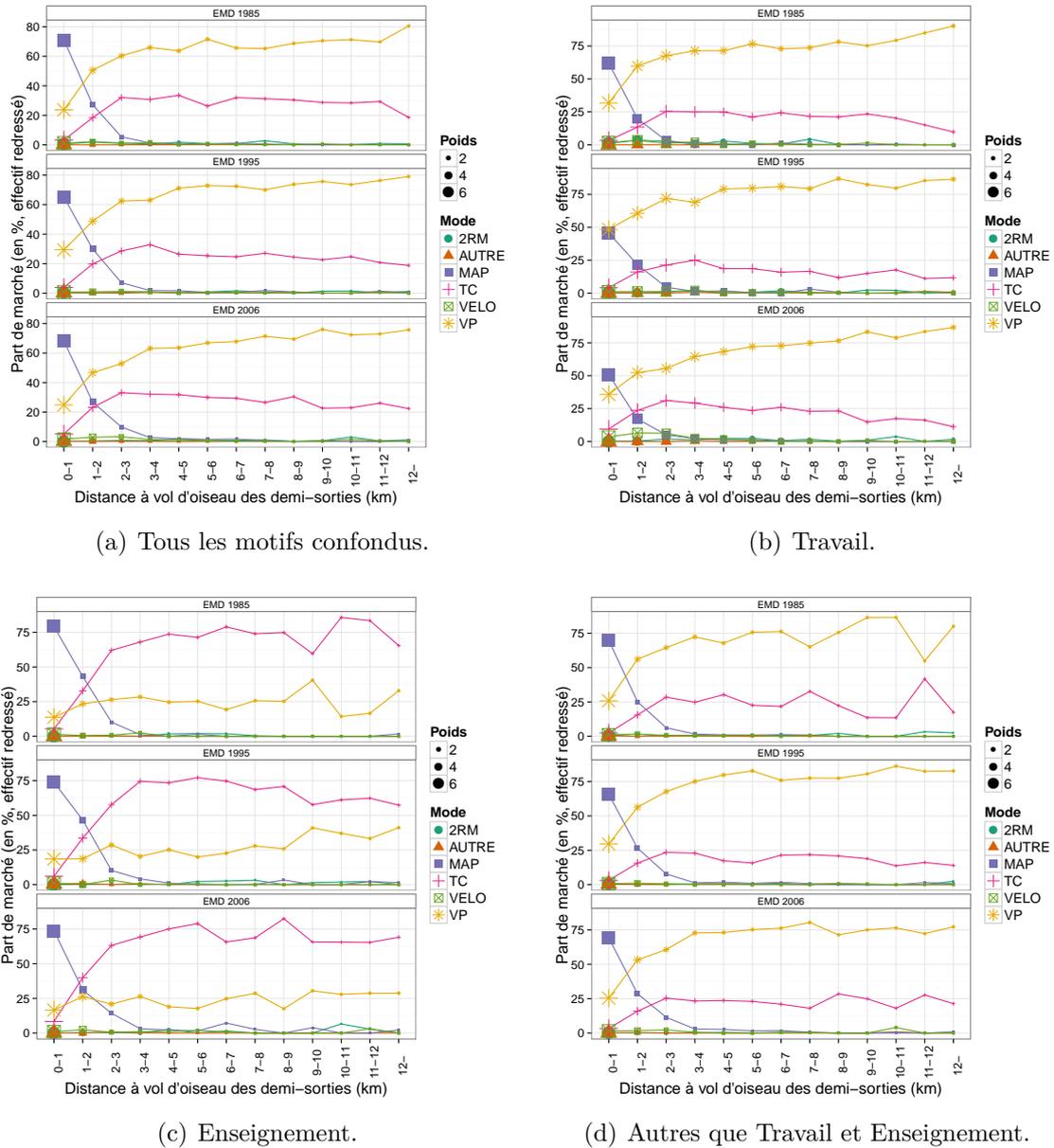
(c) Enseignement.



(d) Autres que Travail et Enseignement.

Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006.

FIGURE 6.2 – Évolution des parts modales à l’intérieur du Grand Lyon en fonction de la distance « à vol d’oiseau » des demi-sorties. Segmentation par motif.



Note de lecture : la taille des points (Poids) va de 0 à 7, 7 représentant l’effectif maximum parmi les effectifs des tranches de distance, pour chaque année d’enquête et motif.

Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006.

le motif Travail, la part de la MAP est un peu plus faible. Elle est d'environ 60 % en 1985 et se réduit à un niveau d'environ 50 % en 1995 et 2006. La part de la MAP est très faible pour les demi-sorties de plus de 3 km. Pour le motif Travail, on peut la considérer comme négligeable au-delà de 2 km. En ce qui concerne les évolutions de l'usage de la VP et des TC, on remarque que celles-ci sont plus fortes pour le motif Travail que pour les autres, notamment pour les demi-sorties de moins de 8 km. Ainsi, pour la tranche de distance de moins de 1 km, la part de la VP passe d'environ 30 % en 1985 à près de 50 % en 1995 pour ensuite baisser à 35 % en 2006. Pour la tranche de distance entre 1 et 8 km, on observe que la part de la VP augmente presque de manière linéaire jusqu'à 75 %. L'augmentation est plus drastique en 1985 et 1995, et la part de la VP dépasse la barre de 75 % dès 4 km en 1995. Pour les TC, on observe que la barre de 25 % de part de marché n'est pas dépassée en 1985 ni en 1995. En revanche, en 2006, celle-ci dépasse ce seuil pour les demi-sorties de 2 à 7 km, et s'en approche pour les demi-sorties d'entre 1 et 2 km et d'entre 7 et 9 km. Il est nécessaire de noter que, si on s'intéresse aux effectifs de chaque tranche de distance, pour les motifs Enseignement et autres que Travail et Enseignement, ceux-ci sont concentrés sur les déplacements courts (de moins de 3 km). En revanche, pour le motif Travail, les demi-sorties plus longues sont plus nombreuses, en lien avec le caractère plus « contraint » de ce motif de déplacement (voir chapitre 5).

### 6.1.3 Spécification des modèles

#### 6.1.3.1 Principes généraux

Nous avons signalé, dans ce qui précède, que les déplacements peuvent se faire par l'intermédiaire de six modes : 2RM, AUTRE, MAP, TC, VELO et VP. Les facteurs qui permettent aux agents d'arbitrer entre ceux-ci et d'en choisir un pour se déplacer sont multiples. Il est classique d'organiser les facteurs explicatifs du choix du mode en trois groupes (Bonnell, 2004) :

- l'offre de transport : les deux principaux paramètres pris en compte sont le prix ressenti par l'individu, en général différent du coût réel, et le temps perçu, qui tient compte de la perception plus ou moins négative associée aux différents moments du déplacement. Le stationnement apparaît aussi comme un facteur déterminant du choix.
- les caractéristiques du déplacement : l'origine et la destination des déplacements sont des facteurs importants du choix, car le niveau d'offre des deux modes motorisés varie selon la localisation des extrémités du déplacement. On peut également évoquer le motif (la demande de transport est une demande « dérivée » de celle de l'activité que l'on cherche à accomplir dans le lieu de destination), l'heure de la journée et le chaînage (on peut, par exemple, aller en vélo au travail mais si l'on va également faire les courses de la semaine au retour on peut préférer de prendre la voiture).
- les caractéristiques de l'individu : celles-ci peuvent conditionner l'accès aux différents modes. Par exemple, les personnes n'ayant pas le permis de conduire ne peuvent pas se déplacer en voiture en tant que conducteurs. Les personnes ayant un revenu supérieur seront plus disposées à privilégier la voiture particulière dans leurs déplacements.

Rappelons également que, d'un point de vue méthodologique, l'un des éléments centraux des modèles de choix discrets est la définition de l'ensemble constitué des alternatives parmi lesquelles le choix est réalisé, aussi connu comme ensemble de choix. On peut argumenter que la définition de celui-ci conditionne à la fois la forme des modèles — ou du moins la famille à laquelle ils appartiennent — et l'échantillon sur lequel l'estimation peut être réalisée. En effet, Ben-Akiva et Lerman (1985) expliquent que l'inclusion d'une alternative dans l'ensemble de choix doit obéir aux contraintes des individus. Ainsi, par exemple, pour modéliser le choix entre deux modes, marche à pied et voiture particulière en tant que conducteur, il est nécessaire que les individus de l'échantillon aient une voiture particulière à disposition. Dit autrement, la définition de l'ensemble de choix conditionne la population sur laquelle le modèle peut être estimé. En ce qui concerne la forme du modèle, les caractéristiques de l'ensemble de choix peuvent rendre certains types de modèles plus adaptés que d'autres. Train (2009, pp. 44–45) explique que les différences systématiques de préférences — en lien avec les variables observées, c'est-à-dire celles qui font partie de la composante déterministe de l'utilité — au sein d'une population peuvent être représentées dans le cadre d'un modèle logit. Cependant, il est nécessaire que les différences de préférences ne soient pas en partie aléatoires. Si c'est le cas, le logit ne peut être utilisé que comme une approximation des préférences moyennes, en raison de sa robustesse. En revanche, dans la mesure où le logit simple ne donne pas une indication sur la distribution des préférences, il est déconseillé de l'utiliser dans des contextes où cette distribution serait importante. Pour illustrer ce propos, Train évoque l'exemple de l'étude de pénétration d'un produit nouveau qui ne toucherait qu'une minorité de la population, plutôt que des personnes avec des goûts moyens.

### 6.1.3.2 Définition d'un ensemble de choix et d'une population de référence

Les données disponibles, ainsi que certains partis-pris méthodologiques, nous ont conduit à restreindre l'ensemble de choix. Il convient de noter d'emblée que nous souhaitons avoir des modèles opérationnels qui soient consistants avec les usages des praticiens de la modélisation.

Compte tenu des parts de marché observées dans la figure 6.1, ainsi que des données socioéconomiques disponibles dans les EMD, nous avons décidé d'éliminer de l'ensemble de choix les modes 2RM, AUTRE et VELO. En effet, dans la mesure où la part de ceux-ci est très faible, notamment celle des 2 premiers, leur inclusion dans la modélisation apporte des complications trop importantes (par exemple, la nécessité d'utiliser un logit mixte<sup>6</sup>) alors que la fiabilité des résultats risque d'être limitée par l'effectif. En ce qui concerne particulièrement le VELO, les données des EMD ne permettent pas dans tous les cas d'identifier clairement l'équipement des ménages et la capacité des individus à s'en servir.

L'utilisation de mesures de performance des réseaux issues d'une modélisation implique que nous héritons une limite de celle-ci : la fiabilité plus faible des temps intra-zones. Or, la figure 6.2 permet de voir que la plupart des demi-sorties pour les motifs autres que le motif Travail sont courtes. La moindre précision des temps

6. Ces modèles sont décrits dans les manuels opérationnels anglo-saxons mais sont, à la connaissance de l'auteur, moins présents en France.

pour les demi-sorties courtes (liée aux calculs d'intra-zones) nous à fait opérer un choix méthodologique important : nous nous concentrons sur l'étude des demi-sorties pour le motif Travail, et sur le choix entre les modes motorisés restants (VP et TC). Il convient de noter ici que ce choix peut aussi être justifié par le fait que, dans la pratique, les demi-sorties pour le motif Travail sont concentrées sur l'heure de pointe et déterminent donc les capacités des infrastructures à évaluer. Par ailleurs, les investissements à évaluer concernent en général deux types d'infrastructures : la voirie et les infrastructures spécifiques aux TC. De ce fait, se focaliser sur ces deux modes n'enlève pas trop de généralité à notre analyse.

Pour la modélisation, nous retenons donc les sous échantillons des EMD répondant aux deux conditions suivantes :

- les demi-sorties dont le motif principal est Travail ;
- les demi-sorties des ménages résidant dans le périmètre du Grand-Lyon, réalisées à l'intérieur de ce périmètre (on considère que tous les ménages résidant dans le Grand Lyon ont accès au mode TC) ;

Pour 1985, l'échantillon contient 7 052 demi-sorties pour le motif Travail (soit 689 110 demi-sorties, en effectif redressé), réalisées par 2 869 personnes (282 180, en effectif redressé) appartenant à 2 133 ménages (210 546, en effectif redressé). Pour 1995, on a 8 060 demi-sorties (726 296 en effectif redressé) réalisées par 3 397 personnes (308 338 en effectif redressé) appartenant à 2 599 ménages (236 462 en effectif redressé). Pour 2006, l'échantillon dispose de 7 393 demi-sorties (654 566 en effectif redressé) réalisées par 3 320 personnes (296 063 en effectif redressé) appartenant à 2 569 ménages (231 611 en effectif redressé).

### 6.1.3.3 Deux spécifications pour les choix entre TC et VP

Nous avons retenu deux modèles de choix entre TC et VP. La forme générale de ces modèles est donnée dans les équations (6.1) et (6.2) ci-dessous :

$$(U_{TC} - U_{VP}) = V_{TC} - V_{VP} + \epsilon = \theta X + \epsilon \quad (6.1)$$

$$\Pr(\text{TC}) = \frac{1}{1 + \exp(V_{VP} - V_{TC})} \quad (6.2)$$

où  $U_{TC}$  et  $U_{VP}$  sont les utilités aléatoires du choix de TC et VP, respectivement. De manière classique,  $\theta$  est un vecteur contenant les paramètres de la différence des parties déterministes de l'utilité ( $V_{TC}$  et  $V_{VP}$ ), qui est supposée linéaire, et  $X$  est la matrice contenant les différentes valeurs des variables explicatives<sup>7</sup>.  $\epsilon$  est un vecteur aléatoire, différence entre les parties aléatoires de  $U_{TC}$  et  $U_{VP}$ , et suit une loi logistique. Il s'agit donc de modèles logit classiques.

Après une phase exploratoire de test des spécifications, les variables retenues pour la modélisation sont les suivantes :

csttc : constante ;

rationbpermisbadultes : nombre de permis de conduire dans le ménage divisé par le nombre d'adultes du ménage ;

---

7. Notons que le choix est binaire et, de ce fait, nous avons donné directement l'expression de la différence des utilités dans l'équation (6.1). Notons par ailleurs que les résultats des estimations, donnés par la suite, correspondent au terme de droite de l'équation (6.1).

rationbvpcdisponbpcpermis : nombre de véhicules à disposition du ménage divisé par le nombre de permis de conduire dans le ménage<sup>8</sup> ;

age : âge de la personne ;

age<sup>2</sup> : âge de la personne au carré. Cette variable permet d'introduire une relation non-linéaire<sup>9</sup> ;

sexe : indicatrice égale à 1 lorsque la personne est une femme, zéro sinon ;

densactivprinc : densité (habitants+emplois)/surface (en ha) de l'IRIS2000 dont le centroïde est le plus proche de celui de la zone fine de réalisation de l'activité principale<sup>10</sup>. Cet indicateur sert de *proxy* pour la contrainte de stationnement ;

tpsgtc : temps généralisé TC entre le lieu du domicile et le lieu de réalisation de l'activité principale ;

tpsgvp : temps généralisé VP entre le lieu du domicile et le lieu de réalisation de l'activité principale ;

lyonvilleurb : indicatrice égale à 1 lorsque le domicile de l'individu est situé dans la partie centrale du périmètre (dans les communes de Lyon ou Villeurbanne).

Notons que parmi les variables retenues, on ne trouve ni une variable reflétant le coût monétaire des demi-sorties ni une variable donnant le revenu des individus. Ces deux variables constituent généralement une partie importante des modèles des choix modal. Or, nous avons vu dans le chapitre 3 que le revenu n'est pas disponible sur l'ensemble des enquêtes. En ce qui concerne le coût, son calcul pose des difficultés méthodologiques, liés en partie à la définition du mode VP (comment pondérer les déplacements des conducteurs et ceux de passagers?), ainsi que des problèmes pratiques (données, temps consacré à l'estimation). Par ailleurs, les modèles de choix modal utilisent généralement le ratio coût du déplacement sur revenu comme variable explicative, plutôt que l'un des deux ou les deux de manière séparée. Il s'en suit qu'il n'aurait pas suffi d'estimer un coût monétaire des demi-sorties, mais il aurait également été nécessaire de développer un modèle d'attribution d'un niveau de revenu aux individus. Les erreurs potentielles de ces modèles et de leur combinaison risquaient de limiter la comparabilité temporelle des paramètres associés au ratio coût/revenu.

Nous avons retenu deux spécifications de type logit binomial classique à estimer : une spécification étendue, mettant en jeu l'ensemble des variables décrites précédemment et une spécification simple, ne comportant que tpsgtc et tpsgvp. Les modèles sont estimés par la méthode du maximum de vraisemblance à l'aide de la librairie `mlogit` (Croissant, 2012) du logiciel statistique R (R Development Core Team, 2011). Notons que les coefficients de redressement des EMD ne sont pas pris en compte dans l'estimation. En effet, nous avons remarqué dans le chapitre 4 que la non prise en compte des inégalités dans les probabilités de tirage peut distordre les relations établies entre les différentes variables de l'échantillon. Dans la pratique, dans le cadre

8. Notons que cette variable n'est pas bornée à 1 (comme dans Train, 1978, par exemple). On considère ici que le fait de posséder plus de véhicules qu'il n'est possible de conduire traduit une préférence particulière pour la conduite.

9. D'autres formes fonctionnelles ont été testées, comme le logarithme de l'âge où une fonction à deux segments linéaires de pente différente (avec différents lieux de coupure). La forme quadratique est celle qui a donné les meilleurs résultats.

10. La densité des IRIS pour les dates des EMD a été calculée à partir de données de recensement et des fichiers SIRENE de l'INSEE. Les centroïdes des IRIS2000 et des zones fines EMD ont été repositionnés au barycentre de la surface bâtie à l'intérieur de l'IRIS2000.

de modèles de choix modal, ces probabilités sont rarement prises en compte. Il en sera de même ici.

### 6.1.4 Évaluer la stabilité des paramètres et la qualité des calages

#### 6.1.4.1 Tester l'égalité des paramètres

La grande majorité de la littérature relative à l'évaluation de la transférabilité temporelle des modèles de prévision des déplacements est consacrée aux modèles de choix discret. Dans la mesure où celle-ci est utilisée fréquemment dans ces études, il semble convenable de retenir la statistique donnée dans l'équation (4.1), déjà utilisée dans le chapitre consacré à la génération (voir page 105), pour évaluer la stabilité un à un des paramètres des modèles.

En prenant les données d'une année de référence, on peut établir un test d'égalité de l'ensemble des paramètres en comparant les log-vraisemblances issues de l'application à ces données d'un modèle avec deux jeux de paramètres. Le premier jeu de paramètres est issu de la calibration d'un modèle de choix modal avec les données de l'année de référence. Le second jeu de paramètres est issu, en général, d'une calibration pour une autre date. En effet, on peut voir cela comme un ensemble de restrictions sur les valeurs des paramètres du modèle. La statistique permettant de réaliser ce test est souvent appelée dans la littérature statistique de test de transfert, ou *TTS* de par les sigles de son nom en anglais (*transfer test statistic*, voir Ortúzar et Willumsen, 2001 ; Fox et Hess, 2010) et est donnée par :

$$TTS_v = -2(LL_v(\theta_v) - LL_v(\theta_u)) \quad (6.3)$$

où  $LL_v$  est la log-vraisemblance évaluée à partir des données de l'année de référence  $v$ ,  $\theta_v$  sont les paramètres du modèle calibrés en  $v$  et  $\theta_u$  sont les paramètres du même modèle calibrés avec des données de l'année  $u$ . Notons que, dans ce chapitre, nous nous limiterons à l'étude de la capacité des modèles calibrés avec des données de 1985 et 1995 à prédire le choix modal en 2006 et à la capacité du modèle calibré en 1985 à prédire le choix modal en 1995. Il s'en suit qu'on a  $u = \{85, 95\}$  et  $v = \{95, 06\}$ . Pour tester l'égalité des jeux de paramètres  $\theta_u$  et  $\theta_v$ , on compare la statistique *TTS* à la distribution d'un  $\chi^2$  avec autant de degrés de liberté qu'il y a de paramètres dans le modèle.

#### 6.1.4.2 Évaluer la capacité prédictive des modèles

Il est usuellement reproché aux tests d'égalité des paramètres présentés ci-dessus d'être beaucoup trop restrictifs dans la mesure où aucun modèle n'est jamais parfaitement spécifié (voir Karasmaa, 2008, ainsi que ce qui a été présenté dans le chapitre 2 de ce travail). Or, si l'on revient à l'objectif opérationnel principal des modèles de choix modal — qui est de prévoir des parts de marché à différentes échelles —, on ne peut que remarquer la nécessité d'inclure des indicateurs reflétant la proximité des prévisions aux parts de marché observées. Ceci peut être fait à plusieurs échelles.

À l'échelle individuelle, Ben-Akiva et Lerman (1985) suggèrent que, dans la mesure où lors de la procédure de calibration la statistique maximisée est la log-vraisemblance,

la manière de procéder la mieux adaptée pour évaluer la performance d'un modèle consiste à utiliser directement la log-vraisemblance ou une statistique dérivée de celle-ci. Dans ce travail, nous retenons la log-vraisemblance comme un indicateur de qualité des calages. Pour l'évaluation de la capacité prédictive, nous faisons appel à un indicateur souvent utilisé dans les études faisant appel à l'indicateur *TTS*, et construit presque à partir des mêmes éléments que celui-ci : l'indice de transfert ou *TI* de par les sigles de son nom en anglais (*transfer index*). Il a été conçu par Koppelman et Wilmot et est une mesure de performance relative. Il décrit le degré avec lequel le modèle transféré — c'est-à-dire celui utilisant les paramètres issus d'une calibration à une autre date ( $\theta_u$ ) — améliore un modèle nul ou de référence (dans le cas d'un modèle de choix modal, il s'agit la plupart du temps d'un modèle simple de parts modales), en termes de log-vraisemblance. Cette amélioration est rapportée à l'amélioration obtenue par l'utilisation du modèle dont les paramètres ont été estimés par la méthode du maximum de vraisemblance, à la date des données de référence ( $\theta_v$ ). De manière formelle, l'indicateur est donné par :

$$TI_v = \frac{(LL_v(\theta_u) - LL_v(\theta_{pm}))}{(LL_v(\theta_u) - LL_v(\theta_{pm}))} \quad (6.4)$$

où  $LL_v(\theta_{pm})$  est la log-vraisemblance du modèle simple de parts modales, évaluées avec les données de la date  $v$ . En termes d'interprétation, la statistique *TI* donne une indication sur la quantité d'information apportée par les paramètres utilisés pour la prévision  $\theta_u$  par rapport à celle contenue dans les paramètres correspondant à l'horizon de prévision  $\theta_v$  (Fox et Hess, 2010).

Pour évaluer la qualité des calages et la capacité prédictive, il est important de compléter les indicateurs par des mesures de performance à des niveaux plus agrégés. Dans ce sens, nous retiendrons la part des TC (en %), évaluée sur l'ensemble de l'échantillon. Nous comparerons la part observée en 2006 aux parts prédites par les autres jeux de coefficients. Il est également usuel de retenir une certaine segmentation des observations et de comparer les écarts entre observations et simulations au niveau de ces segments (voir Silman, 1981, par exemple). Dans la mesure où l'étape de répartition modale doit produire des matrices O-D par mode, nous aurions pu segmenter l'échantillon en fonction de l'origine et la destination des demi-sorties. Pour des questions d'effectif, nous avons trouvé préférable de « limiter » le nombre de segments en ne retenant que la zone de réalisation de l'activité principale, à partir du plus petit découpage commun des EMD (également utilisé pour le chapitre 4). En général, les parts modales sont comparées en termes absolus (Fox et Hess, 2010). Nous avons préféré une approche similaire à celle retenue dans les deux chapitres précédents en retenant un indicateur de type *SRMSE*, en raison des avantages de celui-ci, déjà évoqués dans les chapitres 4 et 5, en termes de linéarité par rapport aux erreurs. L'indicateur retenu ne porte que sur le mode TC. Il est donné dans l'équation (6.5) ci-dessous :

$$SRMSECM^{11} = \frac{\sqrt{\frac{\sum_j (POTC_j - PMTC_j)^2}{J}}}{\frac{\sum_i POTC_j}{J}} \quad (6.5)$$

---

11. *CM* est ajouté à *SRMSE* pour signifier que la statistique correspond aux modèles de choix

où  $POTC_j$  est la part des demi-sorties TC parmi celles observées (ayant pour mode TC ou VP), ayant la zone  $j$  comme lieu de réalisation de l'activité principale.  $PMTC_j$  est la part des demi-sorties TC parmi celles modélisées (ayant pour mode TC ou VP), ayant la zone  $j$  comme lieu de réalisation de l'activité principale.  $J$  est le nombre de zones de réalisation de l'activité principale. Notons que nous n'avons pas retenu de segmentation pour la modélisation elle-même. On applique un même modèle, que l'on évalue ensuite segment à segment.

## 6.2 L'hypothèse de stabilité des paramètres est-elle acceptable dans le cas des modèles de choix modal ?

Cette section présente, dans un premier temps, l'analyse de la qualité des calages et de l'évolution temporelle des paramètres des modèles de choix modal. Ensuite, elle se consacre à l'étude des impacts des évolutions temporelles en termes de pouvoir prédictif des modèles.

### 6.2.1 Résultats des estimations

Ce point développe l'analyse de la table 6.1, donnant les résultats des estimations des modèles présentés dans la section 6.1.3.3.

#### 6.2.1.1 Qualité des calages

L'inclusion de variables socioéconomiques caractérisant l'individu et le ménage auquel il appartient apporte un pouvoir explicatif non négligeable par rapport aux seules variables de performance des réseaux. En termes de  $\rho^2$ , l'ajustement global de la spécification simple s'améliore au cours du temps, alors celui de la spécification étendue est moins bon en 1995 qu'aux autres années. Si on s'intéresse aux valeurs des coefficients, on remarque que les signes de ceux-ci — ou, du moins, de ceux pour lesquels on a une idée *a priori* sur le sens de la relation avec la différence d'utilité — correspondent bien à ce qu'on pouvait attendre. Les ratios entre le nombre de permis et le nombre d'adultes ainsi qu'entre le nombre de véhicules disponibles et le nombre de permis ont, tous les deux, des signes négatifs. Dit autrement, ils favorisent l'usage de la voiture. Pour les variables de temps de transport, les résultats sont également conformes à ce qu'on aurait pu attendre : le temps généralisé TC a un signe négatif, et donc une augmentation de celui-ci pénalise le choix des TC. Inversement, le signe du du temps généralisé VP est positif. Une augmentation de celui-ci favorise donc l'usage des TC. La difficulté à trouver un stationnement dans le lieu de réalisation de l'activité principale pénalise en théorie l'usage de la VP pour s'y rendre. Ceci est corroboré par le signe positif de la variable *densactivprinc*. Le coefficient de la *dummy*-variable indiquant que la personne est une femme est positif pour les trois enquêtes : en moyenne, la probabilité de prendre les TC lorsqu'on est une femme

---

modal et la distinguer ainsi de la statistique correspondante appliquée aux modèles de distribution dans le chapitre précédent (5) de la présente étude.

## 6. Choix modal

TABLE 6.1 – Résultats de calage des modèles de choix modal pour le motif Travail et test de stabilité temporelle des paramètres.

Variable	Coef.	écart-type	t-value	Pr(> t )	$\rho^2$	LL	Transf. nég
Données 1985 (a)							
Spécification étendue							
csttc	3,424 2	0,385 18	8,89	0,00E+00	0,295	-2 550	bc
rationbpermisnbadultes	-3,279 6	0,152 15	-21,56	0,00E+00			bc
rationbvdpdisponbpermis	-2,822 7	0,121 88	-23,16	0,00E+00			
age	-0,089 5	0,018 87	-4,74	2,10E-06			bc
age <sup>2</sup>	0,001 1	0,000 24	4,83	1,38E-06			b
sexe	1,176 6	0,074 58	15,78	0,00E+00			c
tpsgvp	0,089 5	0,007 71	11,61	0,00E+00			
tpsgtc	-0,023 1	0,002 79	-8,30	0,00E+00			c
densactivprinc	0,000 4	0,000 05	6,96	3,38E-12			bc
lyonvilleurb	0,698 9	0,074 59	9,37	0,00E+00			
Spécification simple							
csttc	-1,275 7	0,061 51	-20,74	0,00E+00	0,029	-3 510	bc
tpsgvp	0,078 9	0,006 26	12,61	0,00E+00			
tpsgtc	-0,030 6	0,002 24	-13,65	0,00E+00			c
Données 1995 (b)							
Spécification étendue							
csttc	1,983 8	0,444 54	4,46	8,10E-06	0,273	-2 627	ac
rationbpermisnbadultes	-2,622 6	0,151 71	-17,29	0,00E+00			a
rationbvdpdisponbpermis	-2,895 7	0,116 34	-24,89	0,00E+00			
age	-0,024 4	0,021 73	-1,12	2,62E-01			ac
age <sup>2</sup>	0,000 3	0,000 27	1,21	2,27E-01			ac
sexe	1,015 2	0,074 05	13,71	0,00E+00			c
tpsgvp	0,086 2	0,006 57	13,12	0,00E+00			
tpsgtc	-0,026 9	0,002 63	-10,23	0,00E+00			c
densactivprinc	0,000 1	0,000 03	3,37	7,38E-04			ac
lyonvilleurb	0,681 3	0,074 59	9,13	0,00E+00			
Spécification simple							
csttc	-1,727 4	0,063 03	-27,40	0,00E+00	0,031	-3 502	ac
tpsgvp	0,074 1	0,005 45	13,61	0,00E+00			
tpsgtc	-0,028 9	0,002 19	-13,23	0,00E+00			c
Données 2006 (c)							
Spécification étendue							
csttc	4,587 1	0,438 58	10,46	0,00E+00	0,325	-2 718	ab
rationbpermisnbadultes	-2,311 9	0,160 85	-14,37	0,00E+00			a
rationbvdpdisponbpermis	-2,858 1	0,109 13	-26,19	0,00E+00			
age	-0,147 0	0,021 31	-6,90	5,16E-12			ab
age <sup>2</sup>	0,001 5	0,000 27	5,74	9,53E-09			b
sexe	0,687 9	0,069 45	9,90	0,00E+00			ab
tpsgvp	0,082 6	0,006 25	13,21	0,00E+00			
tpsgtc	-0,036 5	0,002 87	-12,71	0,00E+00			ab
densactivprinc	0,001 3	0,000 08	15,62	0,00E+00			ab
lyonvilleurb	0,839 6	0,071 99	11,66	0,00E+00			
Spécification simple							
csttc	-0,848 4	0,058 36	-14,54	0,00E+00	0,052	-3 819	ab
tpsgvp	0,070 7	0,004 98	14,18	0,00E+00			
tpsgtc	-0,042 4	0,002 29	-18,51	0,00E+00			ab

*Note de lecture : au niveau de la dernière colonne, les lettres a, b et c représentent les années d'estimation des modèles. a représente 1985, b représente 1995 et c représente 2006. Pour une année déterminée, chaque coefficient est comparé à son équivalent pour les deux autres années. Les lettres de ces dernières apparaissent sur la colonne « Transf. nég. » si le test d'égalité est négatif. Ainsi, sur la huitième ligne on lit que le coefficient de la variable tpsgtc évalué en 1985 est significativement différente de celui de 2006, mais non de celui de 1995 (en conséquence, seulement c apparaît sur la dernière colonne).*

*Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006 et des temps issus d'affectations dans les réseaux codifiés.*

est plus grande. Ceci peut traduire une préférence des femmes pour les TC, mais il s'agit plus vraisemblablement du résultat d'interactions au sein des ménages qui favorisent l'utilisation du mode VP par les hommes lorsqu'il y a concurrence (par exemple, lorsque le ménage n'a qu'un seul véhicule disponible et que les lieux ou les horaires de travail rendent impossible de faire le trajet à deux). Une autre explication possible est la moindre proportion de femmes ayant un permis de conduire. Le signe positif du coefficient est consistant avec d'autres travaux (Lisco, 1967, cité par Ben-Akiva et Lerman, 1985, ou encore Silman, 1981). Lorsqu'on regarde les deux variables intégrant l'âge de la personne, on constate que  $\text{age}$  a un signe négatif alors que  $\text{age}^2$  a un signe positif pour les 3 EMD : les individus ont tendance à privilégier l'usage de la VP par rapport aux TC avec l'âge, jusqu'à un point d'inflexion à partir duquel des contraintes socio-démographiques (la part des personnes ayant le permis s'accroît avec le temps, voir section 3.2.2 du chapitre 3), entre autres, font que la relation s'inverse. Ceci est à nuancer par le fait que les coefficients des variables  $\text{age}$  et  $\text{age}^2$  ne sont pas significativement différents de zéro, avec un seuil de 5 %, en 1995. La variable indiquant si le domicile se trouve dans Lyon ou Villeurbanne a un signe positif pour les trois dates. Le choix de la localisation du logement est déterminé par beaucoup de facteurs, dont notamment l'accessibilité en TC à différentes aménités. La desserte en TC (en termes de fréquence et de maillage du réseau) est meilleure dans les communes de Lyon et Villeurbanne que dans le reste du Grand Lyon. On peut penser que les personnes habitant dans un secteur où la desserte TC est meilleure ont une « prédisposition » plus importante à se déplacer en utilisant ce mode, par rapport aux personnes habitant dans les zones moins bien desservies. Lorsqu'on compare les valeurs des coefficients des deux spécifications, les écarts, pour  $\text{tpsgvp}$  et  $\text{tpsgtc}$  sont d'un ordre de grandeur similaire en 2006, mais pour les autres années on observe des écarts importants. Pour  $\text{tpsgvp}$ , ils sont de 13 % en 1985, 16 % en 1995 et 17 % en 2006. Pour  $\text{tpsgtc}$ , ils sont de 25 % en 1985, 7 % en 1995 et 14 % en 2006.

### 6.2.1.2 Évolution temporelle des paramètres

Compte tenu des écarts observés entre les coefficients des deux spécifications et du meilleur ajustement de la spécification étendue, nous nous limiterons à analyser les évolutions temporelles des coefficients de cette dernière<sup>12</sup>.

Le coefficient de  $\text{tpsgvp}$  décroît de manière continue sur l'ensemble de la période 1985–2006. La baisse entre 1985 et 1995 est de 4 %, comme celle entre 1995 et 2006. Notons que ces variations sont plutôt faibles et l'hypothèse de stabilité ne peut être rejetée sur l'ensemble de la période 1985–2006. En ce qui concerne les transports en commun, le coefficient de  $\text{tpsgtc}$ , négatif, augmente en valeur absolue sur l'ensemble de la période 1985–2006. L'augmentation est plus importante entre 1995 et 2006 (35 %) qu'entre 1985 et 1995 (16 %). Ces écarts conduisent à un rejet de l'hypothèse de stabilité entre 2006 et les autres années.

Au niveau des variables autres que les temps généralisés, on remarque des évolutions importantes. Pour la contrainte de stationnement, représentée dans la spé-

12. On peut considérer que les coefficients de la spécification simple sont biaisés par rapport à ceux d'un modèle mieux spécifié. Nous n'analysons donc que les coefficients du « meilleur » modèle disponible, en réduisant ainsi le risque de biais. Les coefficients relatifs à l'âge de la personne sont exclus en raison de leur non-significativité en 1995.

cification étendue par la variable densactivprinc, le coefficient (positif) décroît en valeur absolue entre 1985 et 1995 (70 %), la différence étant de nature à conduire au rejet de l'hypothèse de stabilité des paramètres. Entre 1995 et 2006, le coefficient associé à densactivprinc s'accroît fortement (hausse de 1 110 %), ce qui conduit au rejet de l'hypothèse d'égalité avec ceux de 1985 et 1995. L'aspect motorisation du ménage, dans sa composante rationbpermisbadultes donnant l'aptitude moyenne du ménage à la conduite, voit son coefficient (négatif) baisser en valeur absolue sur l'ensemble de la période 1985–2006. Entre 1985 et 1995, le coefficient baisse de 20 %. L'hypothèse d'égalité des paramètres est rejetée. Entre 1995 et 2006, la baisse observée est plus faible que pour la période précédente (12 %). Les paramètres ne sont pas significativement différents sur cette période. La forte baisse en valeur absolue du coefficient de rationbpermisbadultes peut s'expliquer par l'augmentation de la part des adultes ayant un permis (voir table 3.1), rendant la variable moins discriminante. L'autre composante de la motorisation, rationbvdisponbpermis, donnant la contrainte matérielle des conducteurs de chaque ménage, connaît des variations extrêmement faibles sur l'ensemble de la période d'analyse. Entre 1985 et 1995, le coefficient associé à cette variable augmente en valeur absolue (coefficient négatif) de 3 %. Entre 1995 et 2006, il diminue de 1 % en valeur absolue. Le test de transférabilité ne permet pas de rejeter l'hypothèse d'égalité des paramètres. La stabilité du coefficient de rationbvdisponbpermis peut s'expliquer par le fait que, pour les ménages ayant moins de véhicules que de conducteurs, la contrainte en matière d'organisation des déplacements au sein du ménage n'a pas vraiment changé (et qu'il n'y a pas d'effet de structure similaire à celui observé pour rationbpermisbadultes). En ce qui concerne le sexe de la personne se déplaçant, le coefficient (positif) de la variable sexe, indiquant que le déplacement a été réalisé par une femme, décroît en valeur absolue sur l'ensemble de la période. Entre 1985 et 1995, il diminue de 14 %. Entre 1995 et 2006, la baisse est de 32 %. Cette baisse plus importante conduit à ce que l'hypothèse de stabilité ne soit rejetée que pour les comparaisons entre le coefficient de 2006 et les autres. L'évolution du coefficient peut s'expliquer par l'augmentation de la motorisation des ménages, de la part des adultes ayant un permis, par la réduction de la taille de ceux-ci et surtout par le fait que les comportements des hommes et des femmes ont tendance à se rapprocher. En ce qui concerne le fait de vivre dans Lyon ou Villeurbanne (lyonvilleurb), on observe que le coefficient (positif) diminue légèrement entre 1985 et 1995 (baisse relative de 3 %) et augmente de manière plus importante entre 1995 et 2006 (augmentation relative de 23 %). Le sens de variation du coefficient est consistant avec l'évolution de la part des TC, cependant l'hypothèse de stabilité ne peut être rejetée sur l'ensemble de la période. L'augmentation entre 1995 et 2006 semble pourtant non négligeable. Le non rejet de l'hypothèse de stabilité est donc lié à une variance relativement importante du coefficient.

Enfin, en ce qui concerne la constante, on observe des écarts importants. Elle diminue 42 % entre 1985 et 1995 pour ensuite augmenter de 131 % entre 1995 et 2006. L'hypothèse de stabilité est rejetée sur l'ensemble des trois dates. Celui-ci est un résultat relativement fréquent des études de transférabilité et il est lié au fait qu'aucun modèle n'est jamais parfaitement spécifié. Les méthodes de réactualisation des modèles commencent par ajuster la constante (Karasmaa, 2008).

### 6.2.2 Aurait-on pu prévoir les parts modales en 2006 avec les modèles calibrés en 1985 et 1995 ?

La table 6.2 donne les valeurs des indicateurs  $TTS$  et  $TI$ , décrits respectivement dans les équations (6.3) et (6.4), lorsqu'on applique aux données de 2006 les coefficients estimés avec celles de 1985 et 1995. On observe que l'hypothèse d'égalité de l'ensemble des paramètres ne peut être acceptée pour aucune des années, que ce soit au niveau de la spécification simple ou de la spécification étendue.

L'information apportée par les modèles calibrés en 1985 est systématiquement supérieure à celle apportée par les modèles calibrés en 1995. La spécification étendue contient relativement plus d'information que la spécification simple dans le contexte de prévision, pour les deux années. Nous avons constaté précédemment, que la qualité de l'ajustement sur une même année est bien meilleure pour la spécification étendue que pour la spécification simple. Il s'en suit que, au niveau de l'ajustement à l'échelle des unités d'observation, la performance prédictive de la spécification étendue est bien meilleure que celle de la spécification simple. Dit autrement, la spécification étendue est celle qui permet d'approcher au mieux les comportements individuels à une certaine date et elle est également celle qui permet de faire les prévisions les plus justes. Ces résultats sont à nuancer par le fait que, dans tous les cas, les coefficients utilisés pour la prévision ne contiennent pas plus de 83 % de l'information apportée par un calage réalisé en 2006. Cette marge d'erreur semble relativement importante. Il convient de voir quel est l'impact de l'erreur au niveau de l'échelle d'observation, lorsqu'on agrège les résultats.

La table 6.3 contient les indicateurs agrégés de qualité des prévisions faites avec les deux spécifications. Ceux-ci ont été calculés avec et sans prise en compte des coefficients de redressement inclus dans les EMD. Le premier indicateur donné est la part modale des TC en 2006. Sur les deux premières lignes, où l'on ne prend pas en compte les coefficients de redressement, on constate que la part observée et la part modélisée avec les coefficients de calage de 2006 ( $\theta_{06}$ ) sont identiques (23,45 %). Ceci est une propriété du modèle logit qui, lorsqu'on inclut une constante (ici  $csttc$ ), doit reproduire les parts modales observées. Si on s'intéresse aux parts modales obtenues en utilisant les coefficients de 1985 ( $\theta_{85}$ ) et 1995 ( $\theta_{95}$ ), on constate que l'erreur de prévision est importante dans les deux cas, notamment en 1995. Avec la spécification étendue, la part de marché prévue avec les coefficients de 1985 sous-estime de 2,82 points celle effectivement observée (soit un écart relatif de 12 %). L'utilisation en prévision des coefficients de 1995 conduit également à une sous-estimation, mais l'écart est plus élevé puisqu'il est de 4,91 points (soit 21 % en termes relatifs). En ce qui concerne la spécification simple, l'utilisation des coefficients de 1985 pour la prévision conduit à une surestimation de la part de marché des TC de 3,21 points (soit 14 % en termes relatifs). L'utilisation des coefficients de 1995 conduit à une sous-estimation de 5,52 points (soit un peu moins de 24 % en termes relatifs). Lorsqu'on regarde les parts de marché des TC calculées en prenant en compte les coefficients de redressement, on constate que la part observée augmente de 2,82 points (soit 12 % en termes relatifs). Ce changement est reproduit, par les modèles calibrés en 2006, avec un degré différent de justesse selon la spécification retenue. Ainsi, alors que l'écart relatif n'est que de 1 % pour la spécification étendue, il est de 9 % pour la spécification simple. Au niveau de la prévision, la prise en compte des coefficients de

TABLE 6.2 – Tests de transférabilité globale des modèles calibrés en 1985 et 1995 pour le contexte d'application (données) de 2006.

Spécification	Coefficients 1985			Coefficients 1995		
	$TTS_{06}(\theta_{85})$	$\Pr(> \chi^2)$	$TI_{06}(\theta_{85})$	$TTS_{06}(\theta_{95})$	$\Pr(> \chi^2)$	$TI_{06}(\theta_{95})$
Étendue	455,666	$1,292E-91$	0,826	551,905	$3,506E-112$	0,789
Simple	140,793	$2,550E-30$	0,662	265,506	$2,896E-57$	0,363

*Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006 et des temps issus d'affectations dans les réseaux codifiés.*

TABLE 6.3 – Préviation de la part des TC en 2006 avec les coefficients estimés en 1985 et 1995, avec et sans prise en compte des coefficients de redressement.

Spécification	Part (%) des TC en 2006				<i>SRMSECM</i>		
	observée	avec $\theta_{85}$	avec $\theta_{95}$	avec $\theta_{06}$	1985	1995	2006
Sans appliquer les coefficients de redressement							
Étendue	23,45	20,63	18,54	23,45	0,491	0,546	0,377
Simple	23,45	25,66	17,93	23,45	0,663	0,702	0,648
En appliquant les coefficients de redressement							
Étendue	26,27	22,10	20,03	25,99	0,504	0,548	0,385
Simple	26,27	25,95	18,12	24,02	0,696	0,746	0,670

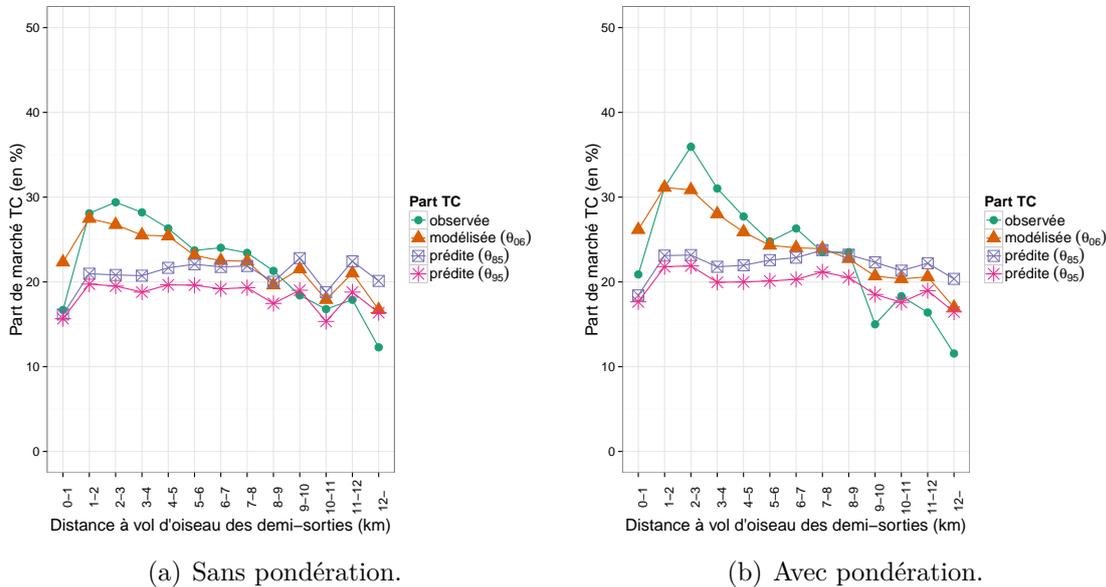
*Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006 et des temps issus d'affectations dans les réseaux codifiés.*

redressement a pour conséquence de réduire la qualité des prévisions réalisées avec la spécification étendue, à la fois en 1985 et en 1995 (les écarts relatifs par rapport à la part observée sont de 16 % et 24 %, respectivement). Pour la spécification simple, la prise en compte des coefficients de redressement a un impact différent pour les coefficients de 1985 et 1995. Elle améliore grandement la qualité des prévisions des premiers puisque l'écart relatif par rapport à la part observée n'est que d'un peu plus de 1 % (il ne s'agit plus d'une surestimation mais d'une sous-estimation). En revanche, elle détériore la qualité des prévisions faites avec les seconds puisque l'écart relatif par rapport à la part observée passe à 31 %.

Si on s'intéresse à l'indicateur *SRMSECM*, on observe que la plupart des observations faites à l'échelle de l'ensemble des demi-sorties sont également valables à l'échelle de la zone. Ainsi, pour les deux spécifications de modèles, que l'on prenne ou pas en compte les coefficients de redressement, la qualité des prévisions réalisées avec les coefficients de 1985 est meilleure que celle des prévisions utilisant les coefficients de 1995. La prise en compte des coefficients de redressement rend les modèles calibrés en 2006 ainsi que les prévisions avec les coefficients de 1985 et 1995 légèrement moins bons en termes d'ajustement. La spécification étendue donne, à l'échelle zonale, des résultats biens meilleurs que la spécification simple, même si la qualité générale de l'ajustement est moyenne puisque l'erreur moyenne est au mieux égale à environ 38 % de la part de marché des TC, pour chaque zone.

La figure 6.3 présente une agrégation des résultats de prévision de la spécification

FIGURE 6.3 – Modélisation et prévision de la part modale des TC à l’intérieur du Grand Lyon en 2006, pour le motif Travail, avec la spécification étendue. Segmentation selon la distance « à vol d’oiseau » des demi-sorties.



*Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006 et des temps issus d’affectations dans les réseaux codifiés.*

étendue par tranche de distance, avec (figure 6.3(b)) et sans (figure 6.3(a)) prise en compte des coefficients de redressement. Si on s’intéresse à la reproduction de la part des TC observée par le modèle calibré en 2006, on observe que la prise en compte des coefficients de redressement ne semble pas distordre de manière importante la capacité de la spécification étendue à reproduire les parts de marché par tranche de distance, même si la pondération génère un pic pour la tranche 2–3 km. Si on s’intéresse à la performance prédictive des modèles de 1985 et 1995, on remarque qu’elle est relativement bonne pour les tranches de distance supérieure à 4 km, notamment pour le modèle de 1985. Cette remarque est valable tant dans le cas sans pondération que dans le cas avec pondération, même si, pour ce dernier, la qualité de reproduction est un peu moins bonne. Pour les tranches de distance inférieure à 4 km, les prévisions de la spécification étendue ne sont pas proches de l’observé (le modèle de 1985 donne des résultats un peu meilleurs que celui de 1995, mais les deux sont plutôt proches entre eux et éloignés des parts observées).

### 6.2.3 Peut-on parler d’un effet de rupture entre 1995 et 2006 ?

Le changement de tendance dans l’évolution des parts modales des TC et VP observé à partir du début des années 2000 (CERTU, 2012) peut être envisagé comme pouvant résulter en partie d’une rupture, liée à des changements de perception par rapport à ces modes, et invalidant partiellement les modèles de prévision estimés (et

TABLE 6.4 – Tests de transférabilité globale de la spécification étendue calibrée en 1985 pour le contexte d’application (données) de 1995.

Spécification	Coefficients 1985		
	$TTS_{95}(\theta_{85})$	$\Pr(> \chi^2)$	$TI_{95}(\theta_{85})$
Étendue	150,078	3,592E-27	0,924

*Source : traitement Cabrera Delgado à partir des EMD 1985 et 1995 et des temps issus d’affectations dans les réseaux codifiés.*

validés) avant la rupture. Dans le point précédent 6.2.2, nous avons vu que le modèle de 1985 permet de faire de meilleures prévisions que celui de 1995 quant aux parts modales de 2006. Il est donc légitime de se demander si on aurait pu prévoir des parts modales dans une situation de prolongement de tendance ou, dit autrement, avant l’arrivée de ce qui pourrait être vu comme la rupture de tendance. Pour ce faire, on peut prendre l’année 1995 comme référence et étudier la qualité des prévisions faites à partir des calibrations en 1985. On se limitera à l’analyse de la spécification étendue qui, comme on l’a vu dans la section 6.2.2, donne généralement les meilleures prévisions des parts modales de 2006.

La table 6.4 donne les résultats du test de transférabilité et l’indice de transfert pour la spécification étendue en prenant 1995 comme année de référence. On remarque que l’information apportée par les coefficients du modèle calibré en 1985 est élevée puisque  $TI$  est supérieur à 0,9. Si on compare ce résultat à l’information apportée par les coefficients estimés en 1985 et 1995 par rapport au modèle estimé en 2006, en prenant 2006 comme année de référence (table 6.2), on constate que  $TI_{95}(\theta_{85})$  est nettement plus élevé que  $TI_{06}(\theta_{85})$  et que  $TI_{06}(\theta_{95})$ . Ceci suggère une plus grande proximité entre les conditions d’estimation de 1985 et 1995 qu’entre celles de 1985 et 2006 ou celles de 1995 et 2006. Cependant, le test d’égalité de l’ensemble des paramètres est rejeté.

La table 6.5 contient les indicateurs agrégés de qualité des prévisions obtenues à partir de la spécification étendue estimée en 1985. Si on s’intéresse à la part de marché globale des TC, on remarque que l’évolution des variables explicatives conduit à ce que le modèle estimé en 1985 prédise une diminution de la part de marché des TC de 1,92 points de pourcentage (0,97 points, si l’on prend en compte les pondérations). En termes relatifs, il s’agit d’une baisse de 9 % (4 %, si l’on prend en compte les pondérations). Le sens de variation a donc bien été prédit, cependant l’amplitude de celle-ci ne l’a pas été. En effet, l’écart entre la part des TC observée en 1995 et celle prédite par le modèle de 1985 est de 2,47 points (2,54 en redressé), soit encore une surestimation de 15 % (14 % en redressé). L’évolution de la structure de la population, en termes de variables explicatives retenues, rend donc compte seulement de 44 % (28 % en redressé) de la variation de part des TC entre 1985 et 1995. On peut comparer ces résultats, d’une prévision à un horizon d’environ 10 ans, avec ceux de la prévision faite par le modèle de 1995 de la part des TC en 2006 (table 6.3). Dans ce dernier cas, le sens de variation de la part des TC est également prédit à condition de prévoir l’évolution de la structure de la population. Cependant, l’évolution de cette dernière ne rend compte que de 29 % (22 % en redressé) de la variation de la part des TC entre

TABLE 6.5 – Prévision de la part des TC en 1995 avec les coefficients estimés en 1985 et 1995, avec et sans prise en compte des coefficients de redressement.

Spécification	Part (%) des TC observée en 1985	Part (%) des TC en 1995			<i>SRMSECM</i>	
		observée	avec $\theta_{85}$	avec $\theta_{95}$	1985	1995
Sans appliquer les coefficients de redressement						
Étendue	20,92	16,53	19,00	16,53	0,552	0,482
En appliquant les coefficients de redressement						
Étendue	21,71	18,24	20,74	18,20	0,585	0,517

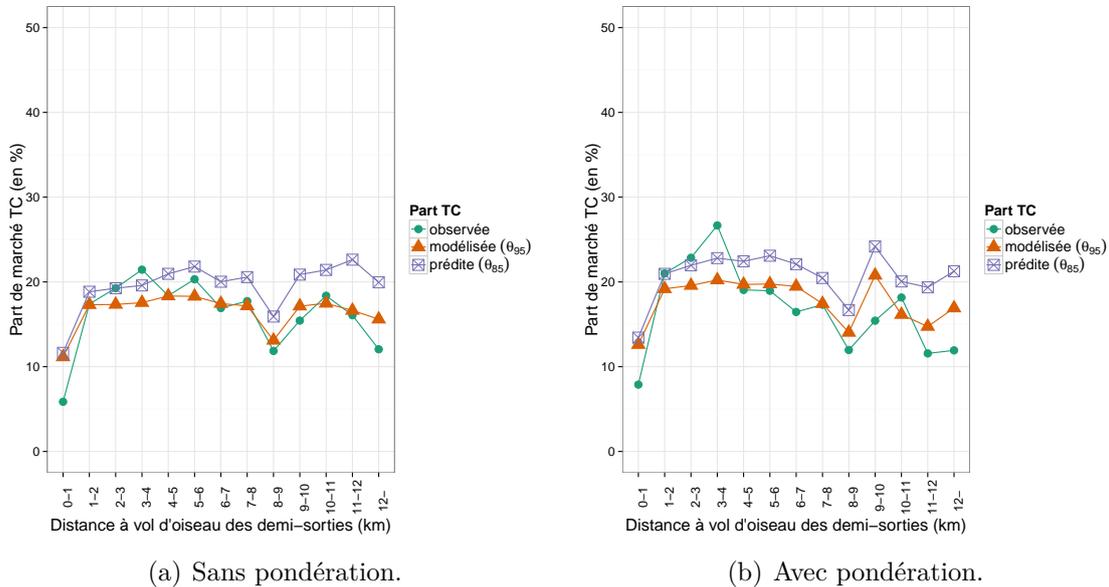
*Source : traitement Cabrera Delgado à partir des EMD 1985 et 1995 et des temps issus d'affectations dans les réseaux codifiés.*

1995 et 2006. Si on s'intéresse à la prévision de la part des TC à l'échelle zonale, on constate que le modèle de 2006 reproduit nettement mieux les observations de 2006 que le modèle de 1995 celles de 1995. L'écart relatif de l'indicateur *SRMSECM* est de 28 % (34 % en redressé). En prévision, les résultats contrastent avec ceux obtenus pour la part globale des TC puisque les *SRMSECM* des prévisions de la part des TC en 2006 faites par le modèle de 1995 sont légèrement plus faibles que celles de la part des TC en 1995 faites par le modèle de 1985. L'écart relatif est de 1 %, sans prendre en compte les pondérations, et de 7 %, si celles-ci sont prises en compte. Il convient, cependant, de noter que les écarts entre les *SRMSECM* des modélisations à la date de référence et ceux des prévisions sont plus faibles pour le couple 1985–1995 que pour 1995–2006. Les écarts relatifs sont de 15 % (sans pondération) et de 13 % dans le premier cas, alors qu'ils sont de 45 % (sans pondération) et de 42 % dans le second. Ceci peut suggérer une plus grande proximité des modèles entre 1985 et 1995 qu'entre 1995 et 2006.

De manière analogue à celle de la figure 6.3, la figure 6.4 présente une agrégation des résultats de prévision de la spécification étendue par tranche de distance, avec (figure 6.4(b)) et sans (figure 6.4(a)) prise en compte des coefficients de redressement. En comparant 6.3(a) et 6.4(a), on constate que, si dans les deux cas la part des TC par tranche de distance est bien reproduite par la modélisation à la date de référence, la prévision semble être meilleure entre 1985 et 1995 qu'entre 1995 et 2006. En effet, les déplacements dont la distance à vol d'oiseau est inférieure à 8 km sont nettement mieux reproduits. Ce constat est également valable pour les données pondérées (figure 6.4(b)), pour lesquelles on remarque néanmoins, comme dans la prévision entre 1995 et 2006 (figure 6.3(b)), une dégradation de qualité de la reproduction par rapport à celle des données non pondérées.

Si ces résultats ne permettent pas d'inférer l'existence d'une rupture liée à des changements de perception des modes de transport, ils sont consistants avec cette hypothèse puisqu'ils révèlent que les modèles estimés sont plus proches entre 1985 et 1995 qu'entre 1995 et 2006.

FIGURE 6.4 – Modélisation et prévision de la part modale des TC à l'intérieur du Grand Lyon en 1985, pour le motif Travail, avec la spécification étendue. Segmentation selon la distance « à vol d'oiseau » des demi-sorties.



Source : traitement Cabrera Delgado à partir des EMD 1985 et 1995 et des temps issus d'affectations dans les réseaux codifiés.

### 6.3 Conclusion

Ce chapitre a permis de corroborer, au niveau des demi-sorties, des évolutions des parts modales bien connues au niveau des déplacements (SYTRAL, 2007a,b). Entre les années 1990 et 2000, les différentes politiques mises en place, visant à limiter l'usage de la voiture, ont accompli leur objectif.

En ce qui concerne l'aptitude des modèles de choix modal à anticiper l'impact de ces politiques, les résultats obtenus par ce travail sont plutôt négatifs. En ce qui concerne la prévision à un horizon de dix ans, les écarts entre, d'une part, les parts de TC observées en 1995 et 2006 et, d'autre part, les prévisions pour ces années faites par les modèles calibrés en 1985 et 1995, respectivement, sont importants. Si les sens de variation des parts modales sont bien reproduits, leur amplitude ne l'est pas. Les modèles ne prédisent que, dans le meilleur des cas, moins de 45 % de la variation totale observée entre deux dates d'enquête. Dans le cas de la prévision à un horizon de vingt ans, contrairement à ce qu'on a pu observer pour les modèles de génération et distribution (chapitres 4 et 5), ce sont les coefficients de 1985 et non ceux de 1995 qui permettent de faire la prévision la plus juste des parts modales en 2006, à la fois au niveau désagrégé comme suivant des agrégations par zone de destination ou tranche de distance. On aurait donc pu faire une meilleure évaluation à 20 ans d'écart qu'à 10 ans. Ce propos doit être nuancé par le fait que les prévisions ne sont satisfaisantes dans aucun des cas.

On peut considérer que les performances explicatives et prédictives des modèles

pourraient être améliorées par une meilleure prise en compte de l'évolution des coûts monétaires des transports et, en particulier, d'usage de la voiture. Cependant, il semble tout aussi plausible que les changements dans les mentalités et la perception des différents modes de transport jouent un rôle non négligeable en faveur des TC. Sur ce dernier point, la comparaison des prévisions à dix ans met en évidence une plus grande part de variation non prédite entre 1995 et 2006 qu'entre 1985 et 2006. Si celle-ci ne montre pas explicitement un effet de rupture, elle reste néanmoins compatible avec celui-ci.

L'analyse présentée ici n'est qu'une première approche d'un sujet complexe de par la disponibilité limitée de données et la simplicité des modèles étudiés. Pour la compléter, il serait souhaitable de compiler des données fiables sur l'évolution des coûts d'usage de la voiture et des revenus des ménages. Également, si par souci de garder une proximité avec la pratique de la modélisation, les résultats présentés correspondent à une spécification de modèle simple, il serait souhaitable, une fois que les données seront disponibles, de tester le comportement dans le temps de spécifications plus complexes. Deux spécifications apparaissent comme étant prioritaires. La première est la spécification principale parmi celles de type *GEV*<sup>13</sup> : le logit hiérarchique. Celui-ci permettrait d'inclure la marche à pied dans l'ensemble de choix modélisé en limitant les effets potentiels d'un non respect de la propriété iid des erreurs<sup>14</sup>. La seconde concerne les modèles à variable latente, pour tenter de prendre en compte l'évolution de la perception des modes de transport.

---

13. *Generalized Extreme Value*, voir Train (2009), chapitre 4.

14. On rappelle que celle-ci donne lieu à la propriété IIA (voir chapitre 1).

---

## Conclusion

La séquence de modélisation à quatre étapes a pour objectif principal de produire des prévisions de trafic, nécessaires à la prise de décision en matière d'investissements en infrastructures de transport. La capacité de la séquence de modélisation à accomplir cette tâche de la manière la plus juste possible est un enjeu d'autant plus important que son utilisation est répandue et que les montants investis dans les infrastructures de transport sont élevés (centaines de millions d'euros) et irrécupérables. Compte tenu du caractère statique des modèles qui la composent, la question de la validité de ces modèles à plusieurs dates est cruciale. La vérification de la stabilité temporelle des modèles permettrait de garantir que les prévisions faites par ces derniers sont pertinentes au regard du processus de prise de décision. Le cas contraire conduirait à devoir intégrer dans les évaluations une part d'incertitude plus importante.

Cette recherche a entrepris la tâche de tester l'hypothèse de stabilité des modèles de génération, de distribution et de répartition modale, en utilisant les données des enquêtes ménages déplacements, réalisées en 1985, 1995 et 2006 sur l'agglomération lyonnaise. Elle a montré que celles-ci peuvent servir à une étude de stabilité temporelle du modèle à quatre étapes (ou, du moins, des trois premières étapes), à condition que certaines perturbations survenues lors des enquêtes, et surtout la baisse du taux de réponse, aient un impact négligeable sur la comparabilité des enquêtes dans le temps. Par ailleurs, elle a permis de constater que l'homogénéisation des données relatives à l'enseignement secondaire et supérieur est problématique, ce qui compliquerait une étude dans le temps de ces motifs de déplacement.

Les données d'enquête permettent d'alimenter la composante « demande », des modèles de prévision. Pour la composante « offre », ce travail a dû surmonter la difficulté du manque de données disponibles. En effet, si des réseaux de voirie et de transports en commun, codifiés à des fins de modélisation, sont disponibles auprès de différents acteurs et institutions de l'agglomération lyonnaise, deux obstacles majeurs à leur utilisation peuvent être trouvés. Le premier est d'ordre technique : il n'y a pas un modèle multimodal officiel sur l'agglomération lyonnaise<sup>15</sup>. Il y a des modèles monomodaux routiers (Grand Lyon et CETE) et de transports en commun (TCL, Egis Rail). De ce fait, il aurait été nécessaire de les retravailler pour les rendre compatibles, sans garantie de succès. Le deuxième obstacle est d'ordre institutionnel :

---

15. Le Grand Lyon a récemment passé un appel d'offre à cette fin. Cette affirmation ne devrait donc plus être valable en 2016.

les prérogatives et intérêts de chaque organisme peuvent entrer en contradiction ce qui rend difficile la mise en place d'une coopération. Par ailleurs, la disponibilité de réseaux codifiés actuels n'implique pas un archivage — sauvegardes systématiques d'états du réseau dans le passé. Une partie importante de la durée du projet a été destinée à la production des données relatives aux réseaux, c'est-à-dire à la codification de ces derniers. Nous avons défini une méthodologie reproductible dans le temps, permettant d'assurer une comparabilité des indicateurs issus des affectations dans les réseaux, à partir des principes de codification retenus pour le projet SIMBAD. Cette méthodologie repose elle-même, néanmoins, sur une hypothèse de stabilité des comportements de conduite, qui semble plausible dans le cas d'études réalisées pour l'heure de pointe, mais dont la validation est confrontée au manque de données.

L'utilisation des données des enquêtes ménages déplacements a permis de tester la capacité de trois approches de modélisation de la génération de sorties, calibrées avec les données de 1985 et 1995, à reproduire les données de 2006. Globalement, toutes les approches permettent de faire de bonnes prévisions de la mobilité à 10 ans d'écart (l'erreur en prévision est similaire à l'erreur des modèles estimés en 2006). Les prévisions à 20 ans sont moins bonnes, produisant des erreurs trop importantes pour être acceptables, notamment pour le motif travail. La qualité des prévisions varie en fonction du choix des modèles. Les modèles de classification croisée sont ceux qui donnent les meilleurs résultats, mais les différences sont très faibles. Les modèles par régression à l'échelle des ménages peuvent bien reproduire les données de calage, mais être moins bons en prévision. La segmentation de la mobilité par motif montre une baisse continue de la mobilité pour le motif travail, alors que celle pour les motifs autres que travail et enseignement n'a pas de tendance marquée (hausse puis baisse). Ces évolutions ne sont pas complètement prises en compte par les variables explicatives retenues. Ainsi, les coefficients des variables explicatives changent dans le temps et les prévisions réalisées sont moins bonnes lorsque l'horizon de prévision est plus éloigné, notamment pour le motif travail.

En ce qui concerne la distribution des demi-sorties, la nécessité d'un zonage fin pour produire les temps de parcours<sup>16</sup> et pour limiter les intra-zones est contraire au besoin de zones étendues pour améliorer la précision des cases de la matrice O-D. Les calages sur un zonage fin sont de qualité moyenne lorsqu'il s'agit de reproduire une matrice O-D, ce qui est une contrainte connue des modèles gravitaires. Cela étant, quand on s'intéresse à la distribution des distances, on constate que le modèle gravitaire permet de les reproduire de manière acceptable, à la fois pour l'année de calage mais aussi dans les prévisions à 10 ans. Pour les prévisions à plus long terme, les distances des demi-sorties pour le motif travail ont été assez bien reproduites alors que celles pour le motif loisirs l'ont été moins bien. La capacité du modèle gravitaire à bien reproduire les distances dépend de la qualité des prévisions des « marges » de la matrice et des temps de parcours.

Pour ce qui est du choix modal, l'analyse des données d'enquête a permis de montrer que la variation des parts modales observée entre 1985 et 2006 est en grande partie concentrée dans la mobilité pour le motif travail. Les estimations, pour ce motif, d'un modèle de type logit binomial appliqué au choix entre les modes motorisés

---

16. Le zonage fin est nécessaire pour que les temps de parcours calculés « aient un sens » pour la pratique.

principaux, voiture particulière et transports en commun, ont permis de mettre en évidence des évolutions (ou stabilités) paramétriques, dont deux sont particulièrement intéressantes du point de vue des dynamiques urbaines. La première concerne la motorisation. Nous avons montré une stabilité du coefficient d'une variable donnant la contrainte effective de partage d'un véhicule dans le ménage : le ratio entre le nombre de véhicules disponibles et le nombre de permis. Cette stabilité peut être interprétée comme reflétant le fait que la coordination interne de la mobilité des ménages, à contrainte de motorisation donnée, n'a pas changé. Le deuxième concerne la réduction des différences de comportements de mobilité, en termes de choix modal, entre hommes et femmes, qui est liée à une convergence des modes de vie. En ce qui concerne l'analyse de la capacité prédictive, les résultats obtenus sont moins encourageants que ceux des autres étapes. À un horizon de 10 ans, les modèles estimés permettent de prévoir le sens de variation de la part modale globale des TC, alors que celui-ci n'est pas le même entre 1985 et 1995 et entre 1995 et 2006. Cependant, les magnitudes des variations ne sont pas prédites de manière satisfaisante. Entre 1985 et 1995, l'erreur de prévision est d'environ 2,5 points (soit 15 % en termes relatifs). Entre 1995 et 2006, l'erreur de prévision est de 4,91 points (soit près de 24 % en termes relatifs). À un horizon de 20 ans, le modèle de 1985 n'a pas été en mesure de prévoir le sens de variation de la part des TC. En revanche, la prévision qu'il fait de celle-ci est nettement meilleure que celle faite par le modèle de 1995 (toutes les deux étant, néanmoins, insuffisantes). En effet, il est plausible que, derrière le renversement de tendance observé au niveau de la répartition modale, il y ait un effet de rupture.

Bien qu'il faille nuancer les résultats décevants du modèle de choix modal, en termes de performance prospective, par le fait que le coût des déplacements en voiture et une mesure fine de la contrainte de stationnement n'ont pas pu être intégrés, les résultats présentés dans ce qui précède, sont plutôt négatifs en relation à l'utilisation de l'ensemble de la chaîne de modélisation à des fins de prévision. En effet, même si les deux premières composantes de la séquence de modélisation à quatre étapes permettent de faire des prévisions acceptables, à un horizon de moyen terme (10 ans), les erreurs issues de la troisième étape sont de nature à rendre les prévisions de l'ensemble de la chaîne peu fiables. La prise en compte des effets d'accumulation des erreurs ne peut qu'aggraver la situation. Pour faire des prévisions à un horizon plus éloigné, il convient donc de mettre en place des stratégies permettant d'intégrer une partie des évolutions des comportements. Une première piste à explorer, en ce sens, est l'étude des déterminants du chaînage des activités, qui peuvent être responsables de la baisse de la mobilité pour le motif travail, ce qui permettrait de savoir si cette tendance risque de se prolonger.

Les perspectives de recherche qui se présentent à la suite de cette thèse peuvent être abordées en fonction de leur proximité thématique au travail accompli, mais aussi selon les possibilités ouvertes par la disponibilité des données sur les réseaux de transport. Dans la première catégorie, on peut évoquer le modèle de distribution gravitaire. En effet, le travail sur celui-ci a donné lieu à une première publication (Cabrerá Delgado et Bonnel, 2012). Partant des traitements réalisés, il est envisageable de travailler sur la stabilité paramétrique, en prenant en compte les aspects liés à la taille des zones (MAUP). On peut également évoquer l'exploration de la performance prospective des méthodes de microsimulation aléatoire, appliquée à la génération, à

partir des algorithmes de classification automatique. Dans la deuxième catégorie, on peut exploiter la donnée que constituent les réseaux modélisés pour étudier les impacts des changements d'accessibilité. En ce sens, une première contribution sur l'évolution des inégalités d'accessibilité aux emplois, pour les habitants des quartiers défavorisés de l'agglomération lyonnaise, a été soumise à publication et acceptée (Bouzouina *et al.*, 2012), sous réserve de modifications. Il serait également souhaitable d'explorer les possibilités d'amélioration de notre méthodologie de codification des réseaux, pour pouvoir exploiter les nouvelles sources de données libres<sup>17</sup> et celles qui devraient être disponibles avec la popularité croissante de l'*open data*. Notons, sur ce dernier point, que l'utilisation de données libres a un avantage non négligeable dans un contexte de recherche : celui de faciliter la reproductibilité des méthodologies.

---

17. Parmi les données libres (libres, en ce qui concerne les droits d'utilisation, et gratuites) disponibles actuellement on peut citer notamment le projet *OpenStreetMap* (<http://www.openstreetmap.org/>, consulté le 24 avril 2013).

---

## Bibliographie

- AMPT, E. S. (1997). Response rates — do they matter? *Dans* BONNEL, P., CHAPLEAU, R., LEE-GOSSELIN, M. E. H. et RAUX, C., éditeurs : *Les Enquêtes de Déplacements Urbains : Mesurer le Présent, Simuler le Futur*, pages 115–125. Éditions du Programme Rhône-Alpes Recherches en Sciences Humaines, Lyon. ISBN 2-909604-31-4.
- AMRHEIN, C. G. (1995). Searching for the elusive aggregation effect: Evidence from statistical simulations. *Environment and Planning A*, Vol. 27 (No. 1), pp.105–119.
- ANAS, A. (1983). Discrete choice theory, information theory and the multinomial logit and gravity models. *Transportation Research Part B: Methodological*, Vol. 17 (No. 1), pp. 13–23.
- ANDERSON, T. W. et HSIAO, C. (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association*, Vol. 76 (No. 375), pp. 598–606.
- ANDERSON, T. W. et HSIAO, C. (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics*, Vol. 18 (No. 1), pp. 47–82.
- ANTONI, J.-P., éditeur (2011). *Modéliser la Ville*. Economica (avec le soutien du PREDIT), Paris, 1<sup>re</sup> édition. ISBN 978-2-7178-5971-3.
- ARELLANO, M. (1989). A note on the Anderson-Hsiao estimator for panel data. *Economic Letters*, Vol. 31 (No. 4), pp. 337–341.
- ARMOOGUM, J. et MADRE, J.-L. (1997). Correction de la non-réponse dans l'Enquête Transports 1993–94 : L'exemple des déplacements quotidiens. *Dans* BONNEL, P., CHAPLEAU, R., LEE-GOSSELIN, M. E. H. et RAUX, C., éditeurs : *Les Enquêtes de Déplacements Urbains : Mesurer le Présent, Simuler le Futur*, pages 115–125. Éditions du Programme Rhône-Alpes Recherches en Sciences Humaines, Lyon. ISBN 2-909604-31-4.
- ASHFORD, N. et HOLLOWAY, F. M. (1972). Time stability of zonal trip production models. *Transportation Engineering Journal*, Vol. 98 (No. 4), pp. 799–806.

- ATROSTIC, B. K., BATES, N., BURT, G. et SILBERSTEIN, A. (2001). Nonresponse in U.S. government household surveys: Consistent measures, recent trends and new insights. *Journal of Official Statistics*, Vol. 17 (No. 2), pp. 209–226.
- AXHAUSEN, K. W. (2000). Definition of movement and activity for transport modelling. Dans HENSHER, D. A. et BUTTON, K. J., éditeurs : *Handbook of Transport Modelling*, chapitre 16, pages 271–284. Pergamon, Amsterdam, 1<sup>re</sup> édition. ISBN 0-08-043594-7.
- BADDOE, D. A. et CHEN, C.-C. (2004a). Modeling trip generation with data from single and two independent cross-sectional travel surveys. *Journal of Urban Planning and Development*, Vol. 130 (No. 4), pp. 167–174.
- BADDOE, D. A. et CHEN, C.-C. (2004b). Unit of analysis in conventional trip generation modelling: An investigation. *Canadian Journal of Civil Engineering*, Vol. 31 (No. 2), pp. 272–280.
- BADDOE, D. A. et MILLER, E. J. (1995a). Analysis of the temporal transferability of disaggregate work trip mode choice models. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1493, pp. 1–11.
- BADDOE, D. A. et MILLER, E. J. (1995b). Comparison of alternative methods for updating disaggregate logit mode choice models. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1493, pp. 90–100.
- BADDOE, D. A. et MWAKALONGE, J. L. (2011). Estimating household trip rates for cross-classification cells with no data: Alternative methods and their performance in prediction of travel. *Journal of Urban Planning and Development*, Vol. 137 (No. 3), pp. 262–271.
- BADDOE, D. A. et STEUART, G. N. (1997). Urban and travel changes in the Greater Toronto Area and the transferability of trip-generation models. *Transportation Planning and Technology*, Vol. 20 (No. 4), pp. 267–189.
- BALTAGI, B. H. (2007). Forecasting with panel data. Center for Policy Research. Paper 74. Disponible en ligne (<http://surface.syr.edu/cgi/viewcontent.cgi?article=1073&context=cpr>, consulté le 24 avril 2013).
- BATTY, M. et MACKIE, S. N. H. (1972). The calibration of gravity, entropy, and related models of spatial interaction. *Environment and Planning*, Vol. 4 (No. 2), pp. 205–233.
- BATTY, M. et SIKDAR, P. K. (1982a). Spatial aggregation in gravity models. 1. An information-theoretic framework. *Environment and Planning A*, Vol. 14 (No. 3), pp. 377–405.
- BATTY, M. et SIKDAR, P. K. (1982b). Spatial aggregation in gravity models. 2. One-dimensional population density models. *Environment and Planning A*, Vol. 14 (No. 4), pp. 525–553.

- BATTY, M. et SIKDAR, P. K. (1982c). Spatial aggregation in gravity models. 3. Two-dimensional trip distribution and location models. *Environment and Planning A*, Vol. 14 (No. 5), pp. 629–658.
- BATTY, M. et SIKDAR, P. K. (1982d). Spatial aggregation in gravity models. 4. Generalisations and large-scale applications. *Environment and Planning A*, Vol. 14 (No. 6), pp. 795–822.
- BAYART, C. (2009). *Le Potentiel du Web pour les Enquêtes de Mobilité Urbaine*. Thèse pour le doctorat de Sciences Économiques option Économie des Transports, Université Lumière Lyon 2, Lyon. vi–361 p.
- BEARDWOOD, J. E. et KIRBY, H. R. (1975). Zone definition and the gravity model: The separability, excludability and compressibility properties. *Transportation Research*, Vol. 9 (No. 6), pp. 363–369.
- BECKER, G. S. (1965). A theory of the allocation of time. *The Economic Journal*, Vol. 75 (No. 299), pp. 493–517.
- BEN-AKIVA, M. et ATHERTON, T. J. (1977). Methodology for short-range travel demand predictions: Analysis of carpooling incentives. *Journal of Transport Economics and Policy*, Vol. 11 (No. 3), pp. 224–261.
- BEN-AKIVA, M. et LERMAN, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press, Cambridge, Massachusetts. ISBN 0-262-02217-6.
- BHOURI, N. (2002). Intermodalité : Bilan et perspectives des systèmes informatiques. Rapport INRETS/GRETIA de convention DTT.
- BIEBER, A. (1985). Le rôle des transports en commun dans la planification de l'agglomération lyonnaise. *Cahier / Groupe Réseaux*, Vol. 1 (No. 2), pp. 5–20.
- BLACK, W. R. (1995). Spatial interaction modeling using artificial neural networks. *Journal of Transport Geography*, Vol. 3 (No. 3), pp. 159–166.
- BOITEUX, M. (2001). Introduction. *Dans Transports : choix des investissements et coût des nuisances*. Rapport du Commissariat Général du Plan. Rapporteur Général : Luc Baumstark. 323p.
- BOND, S. R. (2002). Dynamic panel data models: A guide to micro data methods and practice. *Portuguese Economic Journal*, Vol. 1 (No. 2), pp. 141–162.
- BONNAFOUS, A. (1989). *Le Siècle des Ténèbres de l'Économie*. Economica, Paris. ISBN 2-7178-1797-2.
- BONNAFOUS, A. (2004). Le choix entre voiture et transport collectif. *Dans INSTITUT DES VILLES*, éditeur : *Villes et Économie*, pages 185–206. La Documentation Française, Paris. ISBN 2-11-005569-3.
- BONNAFOUS, A. et PUEL, H. (1983). *Physionomies de la Ville*. Éditions ouvrières, Paris. ISBN 2-7082-2304-6.

- BONNEL, P. (2003). Postal, telephone and face-to-face surveys: How comparable are they? Dans STOPHER, P. R. et JONES, P., éditeurs : *Transport Survey Quality and Innovation*, chapitre 12, pages 215–237. Pergamon. Elsevier Science Ltd., Oxford, 1<sup>re</sup> édition. ISBN 0 08 044096 7.
- BONNEL, P. (2004). *Prévoir la Demande de Transport*. Presses de l'École Nationale des Ponts et Chaussées, Paris, 1<sup>re</sup> édition. ISBN 2-85978-395-4.
- BONNEL, P. et LE NIR, M. (1998). The quality of survey data: Telephone versus face-to-face interviews. *Transportation*, Vol. 25 (No. 2), pp. 147–167.
- BOUZOUINA, L., CABRERA DELGADO, J. E. et ÉCHARD, S. (2011). Quartiers pauvres, quartiers aisés : À qui bénéficient les évolutions récentes de l'accessibilité urbaine ? Application à l'agglomération lyonnaise. Working Paper. Laboratoire d'Économie des Transports. Lyon.
- BOUZOUINA, L., CABRERA DELGADO, J. E. et EMMERICH, G. (2012). Inégalités spatiales d'accessibilité à l'emploi en transport collectif : Deux décennies d'évolution du réseau et de dynamiques urbaines dans l'agglomération lyonnaise. *Revue d'Économie Régionale et Urbaine*. Accepté sous réserve de modifications.
- BOYCE, D. (2002). Is the sequential travel forecasting paradigm counterproductive? *Journal of Urban Planning and Development*, Vol. 128 (No. 4), pp.169–183.
- BRADLEY, M. (2006). Process data for understanding and modelling travel behaviour. Dans STOPHER, P. R. et STETCHER, C., éditeurs : *Travel Survey Methods: Quality And Future Directions*, chapitre 27, pages 491–510. Elsevier, Oxford. ISBN 0-08-044662-0.
- BREIMAN, L., FRIEDMAN, J., OLSHEN, R. A. et STONE, C. J. (1984). *Classification and Regression Trees*. Chapman and Hall/CRC, London. ISBN 0-412-04841-8.
- BRIANT, A., COMBES, P.-P. et LAFOURCADE, M. (2010). Dots to boxes: Do the size and shape of geographical units jeopardize economic geography estimations? *Journal of Urban Economics*, Vol. 67 (No. 3), pp. 287–302.
- CABANNE, I. (2005). *Modélisation à Long Terme de l'Évolution des Trafics Voyageurs à Longue Distance en France*. Thèse pour le doctorat de Sciences Économiques option Économie des Transports, Université Lumière Lyon 2, Lyon. 356 p.
- CABRERA DELGADO, J. E. et BONNEL, P. (2012). Aurait-on pu prévoir l'allongement des distances des déplacements urbains observé ces vingt dernières années avec le modèle de distribution gravitaire? *Cahiers Scientifiques du Transport*, Vol. 62, pp. 33–64.
- CAMBRIDGE SYSTEMATICS, INC (2009). A recommended approach to delineating traffic analysis zones in Florida. Rapport technique, Florida Department of Transportation Systems Planning Office. Préparé en association avec AECOM Consult.
- CAMBRIDGE SYSTEMATICS, INC (2010). *Travel Model Validation and Reasonableness Checking Manual Second Edition*. Travel Model Improvement Program. Federal Highway Administration, Washington D.C.

- CATS (1952–1962). Chicago Area Transportation Study. Final report in three parts. Study conducted under the sponsorship of the State of Illinois, Department of Public Works and Buildings, County of Cook, Board of Commissioners of Cook County. CONTENTS: v. 1, Survey Findings, December 1959 ; v. 2, Data projections, July 1960 ; Transportation Plan, April 1962. Chicago.
- CAUBEL, D. (2006). *Politique de Transports et Accès à la Ville pour Tous ? Une Méthode d'Évaluation Appliquée à l'Agglomération Lyonnaise*. Thèse pour le doctorat de Sciences Économiques option Économie des Transports, Université Lumière Lyon 2, Lyon. 444 p.
- CERTU (2002). *Modélisation des déplacements urbains de voyageurs : Guide des Pratiques*. Centre d'Études sur les Réseaux, les Transports, l'Urbanisme et les constructions publiques. Éditions du Certu, Lyon. ISBN 2-11-093140-X.
- CERTU (2008). *L'enquête ménages déplacements « standard Certu » : Guide méthodologique*. Centre d'Études sur les Réseaux, les Transports, l'Urbanisme et les constructions publiques. Éditions du Certu, Lyon. ISBN 978-2-11-097161-6.
- CERTU (2012). *La Mobilité Urbaine en France. Enseignements des années 2000–2010*. Centre d'Études sur les Réseaux, les Transports, l'Urbanisme et les constructions publiques. Éditions du Certu, Lyon. ISBN 978-2-11-099597-1.
- CETE DE LYON (1986). Enquête ménages Lyon 1985 : Agglomération lyonnaise 1985–1986. Document Technique.
- CETE DE LYON (2006). Enquête ménages déplacements de l'aire métropolitaine lyonnaise 2005–2006. Rapport de Mission d'Assistance à Maîtrise d'Ouvrage du CETE de Lyon. Suivi d'Enquête.
- CLARK, W. A. V. et AVERY, K. L. (1976). The effects of data aggregation in statistical analysis. *Geographical Analysis*, Vol. 8 (No. 4), pp. 428–438.
- CLARKE, M. I., DIX, M. C. et GOODWIN, P. (1982). Some issues of dynamics in forecasting travel behaviour — a discussion paper. *Transportation*, Vol. 11 (No. 2), pp. 153–172.
- COCHRANE, R. A. (1975). A possible economic basis for the gravity model. *Journal of Transport Economics and Policy*, Vol. 9 (No. 1), pp. 34–49.
- COTRUS, A. V., PRASHKER, J. N. et SHIFTAN, Y. (2005). Spatial and temporal transferability of trip generation demand models in Israel. *Journal of Transportation and Statistics*, Vol. 8 (No. 1), pp. 37–56.
- CROISSANT, Y. (2012). Estimation of multinomial logit models in R: The mlogit packages. Disponible en ligne (<http://cran.r-project.org/web/packages/mlogit/vignettes/mlogit.pdf>, consulté le 24 avril 2013).
- CROISSANT, Y. et MILLO, G. (2008). Panel data econometrics in R: The plm package. *Journal of Statistical Software*, Vol. 27 (No. 2), pp. 1–43. Disponible en ligne (<http://www.jstatsoft.org/v27/i02/paper>, consulté le 24 avril 2013).

- CROZET, Y. (2005). Le temps et les transports de voyageurs. *Dans CONFÉRENCE EUROPÉENNE DES MINISTRES DES TRANSPORTS*, éditeur : *Le Temps et les Transports : Rapport de la 127<sup>e</sup> Table Ronde d'Économie des Transports*, Tables Rondes du CEMT, pages 27–69. Service des Publications de l'OCDE, Paris. ISBN 92-821-2332-4.
- DEMING, W. E. et STEPHAN, F. F. (1940). On least squares adjustment of a sampled frequency table when the expected marginal totals are known. *The Annals of Mathematical Statistics*, Vol. 11 (No. 4), pp. 427–444.
- DILLMAN, D. A. et CHRISTIAN, L. M. (2005). Survey mode as a source of instability in responses across surveys. *Field Methods*, Vol. 17 (No. 1), pp. 30–52.
- DIMOU, M. (2003). Vers la construction d'un cadre d'analyse de la croissance endogène localisée : Une revue de littérature sur les nouveaux modèles de représentation des phénomènes d'agglomération et de leur évolution. *Revue d'Économie Régionale et Urbaine*, Vol. 2003 (No. 5), pp. 755–772.
- DODD, N. (1966). Studies of trends of travel between 1954 and 1964 in a large metropolitan area. *Highway Research Record*, Vol. 141, pp. 1–23.
- DOMENCICH, T. A. et MCFADDEN, D. (1975). *Urban Travel Demand: A Behavioral Analysis*. North-Holland Publishing, Amsterdam, Oxford. ISBN 0-7204-3196-4.
- DOUBLEDAY, C. (1977). Some studies of the temporal stability of person trip generation models. *Transportation Research*, Vol. 11 (No. 4), pp. 255–264.
- DOWNES, J. D. et GYENES, L. (1976). Temporal stability and forecasting ability of trip generation models in reading. Transportation and Road Research Laboratory, Report 726, 40 p. Crowthorne, UK.
- DREIF (2008). *MODUS v2.1 : Documentation détaillée du modèle de déplacements de la DREIF*. Direction Régionale de l'Équipement Île-de-France, Paris. Disponible en ligne sur le site de la DRIEA Île-de-France (<http://www.driea.ile-de-france.developpement-durable.gouv.fr/modus-documentation-detailllee-du-a1459.html>, consulté le 24 avril 2013).
- DU CREST, T. et CLÉMENT, L. (1999). *Les Logiciels de Planification des Déplacements Urbains : Catalogue*. Centre d'Études sur les Réseaux, les Transports, l'Urbanisme et les constructions publiques. Disponible en ligne (<http://lara.inist.fr/handle/2332/1158>, consulté le 24 avril 2013).
- DUFFUS, L. N., ALFA, A. S. et SOLIMAN, A. H. (1987). The reliability of using the gravity model for forecasting trip distribution. *Transportation*, Vol. 14 (No. 3), pp. 175–192.
- DUMOUCHEL, W. H. et DUNCAN, G. J. (1983). Using sample survey weights in multiple regression analyses of stratified samples. *Journal of the American Statistical Association*, Vol.78 (No. 383), pp. 535–543.

- DURANTON, G. et TURNER, M. A. (2011). The fundamental law of road congestion: Evidence from US cities. *American Economic Review*, Vol. 101 (No. 6), pp. 2616–2652.
- ELMI, A. M., BADOE, D. A. et MILLER, E. J. (1999). Transferability analysis of work-trip-distribution models. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1676, pp. 169–176.
- ERLANDER, S. (1977). Accessibility, entropy and the distribution and assignment of traffic. *Transportation Research*, Vol. 11 (No. 3), pp. 149–153.
- ERLANDER, S. et STEWART, N. F. (1990). *The Gravity Model in Transportation Analysis: Theory and Extensions*. Topics in Transportation. VSP, Utrecht, the Netherlands. ISBN 90-6764-089-1.
- ETTEMA, D. F. et TIMMERMANS, H. J. P., éditeurs (1997). *Activity-Based Approaches to Travel Analysis*. Pergamon Press, Oxford. ISBN 0080425844.
- EVANS, A. W. (1971). The calibration of trip distribution models with exponential or similar cost functions. *Transportation Research*, Vol. 5 (No. 1), pp. 15–38.
- EVANS, S. P. et KIRBY, H. R. (1974). A three-dimensional Furness procedure for calibrating gravity models. *Transportation Research*, Vol. 8 (No. 2), pp. 105–122.
- FLYVBJERG, B. (2005). Measuring inaccuracy in travel demand forecasting: Methodological considerations regarding ramp up and sampling. *Transportation Research Part A: Policy and Practice*, Vol. 39 (No. 6), pp. 522–530.
- FLYVBJERG, B., HOLM, M. K. S. et BUHL, S. (2005). How (in)accurate are demand forecasts in public works projects. *Journal of the American Planning Association*, Vol. 71 (No. 2), pp. 131–146.
- FOTHERINGHAM, A. S. et KNUDSEN, D. C. (1987). *Goodness of Fit Statistics*. Numéro 46 de *Catmog*. Geo Books, Norwich. ISBN 0306-614-2.
- FOTHERINGHAM, A. S. et WONG, D. W. S. (1991). The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A*, Vol. 23 (No. 7), pp. 1025–1044.
- FOX, J. et HESS, S. (2010). Review of evidence for temporal transferability of mode-destination models. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2175, pp. 74–83.
- FRIEDRICH, M. et GALSTER, M. (2009). Methods for generating connectors in transport planning models. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2132, pp. 133–142.
- FURNESS, K. P. (1965). Time function iteration. *Traffic Engineering and Control*, Vol. 7 (No. 7), pp. 458–460.

- GALBRAITH, R. A. et HENSHER, D. A. (1982). Intra-metropolitan transferability of mode choice models. *Journal of Transport Economics and Policy*, Vol. 16 (No. 1), pp. 7–29.
- GAUDRY, M. (2007). Structure de la modélisation du trafic et théorie économique. Dans MAURICE, J. et CROZET, Y., éditeurs : *Le Calcul Économique*, chapitre 1, pages 6–97. Economica (avec le soutien du PRÉDIT), Paris, 1<sup>re</sup> édition. ISBN 978-2-7178-5425-1.
- GESMAD (2000). Évaluation des modèles de prévision de trafic, note de synthèse. Rapport pour le PRÉDIT. Paris.
- GITTON, F. (2006). Comportements de mobilité et simulation des déplacements sur l'agglomération lyonnaise. Rapport de stage. ENSAE 2<sup>e</sup> année.
- GODINOT, C. et BONNEL, P. (2008). Phase d'affectation : mise en forme du réseau routier sur l'aire urbaine de Lyon. Laboratoire d'Économie des Transports. Rapport intermédiaire du projet SIMBAD pour le compte de la DRI (MEEDDAT) et de l'ADEME dans le cadre du groupe 11 du PRÉDIT 3, 173 p., Lyon.
- GOULIAS, K. G. (1999). Longitudinal analysis of activity and travel pattern dynamics using generalized mixed Markov latent class models. *Transportation Research Part B: Methodological*, Vol. 33 (No. 8), pp. 535–558.
- GREENE, W. H. (2011). *Econometric Analysis*. Prentice Hall, Boston, 7<sup>e</sup> édition. ISBN 0-13-139538-6.
- GUISTON, H. (1963). *Statistique et Économétrie*. Dalloz, Paris, 2<sup>e</sup> édition.
- HANSEN, S. (1981). In favor of cross-cultural transferability of travel-demand models. Dans STOPHER, P. R., MEYBURG, A. H. et BRÖG, W., éditeurs : *New Horizons in Travel-Behavior Research*, pages 637–651. Lexington Books, Lexington, Massachusetts. ISBN 0669028509.
- HARVEY, A. S. et WILSON, C. (2001). Evolution of daily activity pattern from 1971 to 1981: A study of the Halifax Activity Panel Survey. *Canadian Studies in Population*, Vol. 28 (No. 2), pp. 459–489.
- HEGGIE, I. G. (1978). Putting behaviour into behavioural models of travel choice. *The Journal of the Operational Research Society*, Vol. 29 (No. 6), pp. 541–550.
- HENSHER, D. A. et BUTTON, K. J., éditeurs (2005). *Handbook of Transport Modelling*. Pergamon, Amsterdam, 3<sup>e</sup> édition. ISBN 0-08-043594-7.
- HOROWITZ, A. J. (2001). Computational issues in increasing spatial precision of traffic assignments. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1777, pp. 68–74.
- HYMAN, G. M. (1969). The calibration of trip distribution models. *Environment and Planning*, Vol. 1 (No. 1), pp. 105–112.

- INSEE, SYTRAL et CETE DE LYON (1995). Enquête déplacements auprès des ménages de l'agglomération lyonnaise 1994–1995. Document Technique.
- ISMART, D. (1990). Calibration and adjustment of system planning models. Rapport technique FHWA-ED-90-015, U.S. Department of Transportation. Federal Highway Administration, Washington D.C.
- KALFS, N. et VAN EVERT, H. (2003). Nonresponse and travel surveys. Dans STOPHER, P. R. et JONES, P., éditeurs : *Transport Survey Quality and Innovation*, chapitre 35, pages 567–585. Pergamon. Elsevier Science Ltd., Oxford, 1<sup>re</sup> édition. ISBN 0 08 044096 7.
- KANNEL, E. J. et HEATHINGTON, K. W. (1973). Temporal stability of of trip generation relationships. *Highway Research Record*, Vol. 472, pp. 17–27.
- KARASMAA, N. (2008). *The Transferability of Travel Demand Models: An Analysis of Transfer Methods, Data Quality and Model Estimation*. VDM Verlag Dr. Müller Aktiengesellschaft & Co. KG, Saarbrücken, Allemagne. ISBN 978-3-639-09759-7.
- KARASMAA, N. et PURSULA, M. (1997). Empirical studies of transferability of Helsinki metropolitan area travel forecasting models. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1607, pp. 38–44.
- KASS, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Journal of Applied Statistics*, Vol. 29 (No. 2), pp. 119–127.
- KIRBY, H. R. (1974). Theoretical requirements for calibrating gravity models. *Transportation Research*, Vol. 8 (No. 2), pp. 97–104.
- KIRBY, H. R. (1979). Partial matrix techniques. *Traffic Engineering and Control*, Vol. 20 (No. 8–9), pp. 422–428.
- KITAMURA, R. (1984). Incorporating trip chaining into analysis of destination choice. *Transportation Research Part B: Methodological*, Vol. 18 (No. 1), pp. 67–81.
- KITAMURA, R. (1990). Panel analysis in transportation planning: An overview. *Transportation Research Part A: General*, Vol. 24 (No. 6), pp. 401–415.
- KITAMURA, R. et KOSTYNIUK, L. P. (1986). Maturing motorization and household travel: The case of nuclear-family households. *Transportation Research Part A: General*, Vol. 20 (No. 3), pp. 245–260.
- KOCKELMAN, K. M. (1998). *A Utility-Theory-Consistent System-of-Demand-Equations Approach to Household Travel Choice*. A dissertation presented in partial satisfaction of the requirements for the degree of Doctor of Philosophy in Engineering, University of California at Berkley, Berkley. 165 p.
- KOLLO, H. P. H. et PURVIS, C. L. (1984). Changes in regional travel characteristics in the San Francisco Bay Area: 1960–1981. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 987, pp. 57–66.

- KOPPELMAN, F. et WILMOT, C. (1982). Transferability analysis of disaggregate choice models. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 895, pp. 18–24.
- KOSTYNIUK, L. P. et KITAMURA, R. (1984). Trip chains and activity sequences: Test of temporal stability. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 987, pp. 29–39.
- KUMAR, A. et LEVINSON, D. (1993). Specifying, estimating, and validating a new trip generation model: A case study of Montgomery County, Maryland. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1413, pp. 107–113.
- LAMB, G. M. et HAVERS, G. E. (1970). Introduction to transportation planning: Treatment of networks. *Traffic Engineering and Control*, Vol. 11 (No. 11), pp. 486–489.
- LANCASTER, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy*, Vol. 74 (No. 2), pp. 132–157.
- LEONTIEF, W. W. (1941). *The Structure of American Economy, 1919–1929: An Empirical Application of Equilibrium Analysis*. Harvard University Press, Cambridge, Massachusetts.
- LICHÈRE, V. et RAUX, C. (1997a). Développement d’un modèle stratégique de simulation des déplacements : Guide de l’utilisateur. SEMALY, LET. Lyon.
- LICHÈRE, V. et RAUX, C. (1997b). Développement d’un modèle stratégique de simulation des déplacements : Présentation générale. SEMALY, LET. Lyon.
- LISCO, T. E. (1967). *The Value of Commuter’s Travel Time: A Study in Urban Transportation*. PhD dissertation. Department of Economics, University of Chicago, Chicago.
- LUCE, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. Wiley and Sons, New York.
- LUMLEY, T. (2010). *Complex Surveys: A Guide to Analysis Using R*. John Wiley & Sons, Hoboken, New Jersey. ISBN 978-0-470-28430-8.
- MANDEL, J. (1995). *The Analysis of Two-Way Layouts*. Chapman and Hall/CRC, New York, 1<sup>re</sup> édition. ISBN 0-412-98611-6.
- MANHEIM, M. L. (1979). *Fundamentals of Transportation Systems Analysis: Basic Concepts (Volume 1)*. MIT Press, Cambridge, Massachusetts. ISBN 0-262-63289-6.
- MANHEIM, M. L. (1980). Understanding “supply” in transportation systems. *Transportation Research Part A: General*, Vol. 14 (No. 2), pp. 119–135.
- MASSON, S. (2000). *Les Interaction entre Système de Transport et Système de Localisation en Milieu Urbain et leur Modélisation*. Thèse pour le doctorat de Sciences Économiques option Économie des Transports, Université Lumière Lyon 2, Lyon. 542 p.

- MCARTHUR, D. P., KLEPPE, G., THORSEN, I. et UBØE, J. (2011). The spatial transferability of parameters in a gravity model of commuting flows. *Journal of Transport Geography*, Vol. 19 (No. 4), pp. 596–605.
- MCCARTHY, P. S. (1982). Further evidence on the temporal stability of disaggregate travel demand models. *Transportation Research Part B: Methodological*, Vol. 16 (No. 4), pp. 263–278.
- MCFADDEN, D. (1974). The measurement of urban travel demand. *Journal of Public Economics*, Vol. 3 (No. 4), pp. 303–328.
- MCFADDEN, D. (2001). Economic choices. *American Economic Review*, Vol. 91 (No. 3), pp. 351–378.
- MCNALLY, M. G. (2000a). The activity-based approach. Dans HENSHER, D. A. et BUTTON, K. J., éditeurs : *Handbook of Transport Modelling*, chapitre 4, pages 53–69. Pergamon, Amsterdam, 1<sup>re</sup> édition. ISBN 0-08-043594-7.
- MCNALLY, M. G. (2000b). The four-step model. Dans HENSHER, D. A. et BUTTON, K. J., éditeurs : *Handbook of Transport Modelling*, chapitre 3, pages 35–52. Pergamon, Amsterdam, 1<sup>re</sup> édition. ISBN 0-08-043594-7.
- MEYER, J. R. et STRASZHEIN, M. R. (1971). *Pricing and Project Evaluation*, volume 1 de *Techniques of Transport Planning*. The Brookings Institution, Washington D.C. ISBN 0-8157-5690-9.
- MIKKONEN, K. et LUOMA, M. (1999). The parameters of the gravity model are changing — how and why? *Journal of Transport Geography*, Vol. 7 (No. 4), pp. 277–283.
- MILLER, H. J. (1999). Potential contributions of spatial analysis to geographic information systems for transportation (GIS-T). *Geographical Analysis*, Vol. 31 (No. 4), pp. 373–399.
- MONTÈS, C. (2003). *Les Transports Dans l'Aménagement Urbain à Lyon*. Revue de Géographie de Lyon : Géocarrefour, Lyon, 1<sup>re</sup> édition. ISBN 2-9519647-0-6.
- MONZON, J., GOULIAS, K. G. et KITAMURA, R. (1989). Trip generation models for infrequent trips. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1220, pp. 40–46.
- MOZOLIN, M., THILL, J. C. et LYNN USERY, E. (2000). Trip distribution forecasting with multilayer perceptron neural networks: A critical evaluation. *Transportation Research Part B: Methodological*, Vol. 34 (No. 1), pp. 53–73.
- MWAKALONGE, J. L. (2010). *Econometric Modeling of Total Urban Travel Demand Using Data Collected in Single and Repeated Cross-Sectional Surveys*. A dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Engineering, Tennessee Technological University, Cookeville, Tennessee. 155 p.

- MWAKALONGE, J. L. et BADOE, D. A. (2012). Trip generation modeling using data collected in single and repeated cross-sectional surveys. *Journal of Advanced Transportation*. Publié en ligne dans Wiley Online Library (<http://onlinelibrary.wiley.com/doi/10.1002/atr.217/abstract>, consulté le 24 avril 2013).
- NICOLAS, J.-P. (2011). SIMBAD : Un outil pour intégrer le développement durable dans les politiques publiques. Dans ANTONI, J.-P., éditeur : *Modéliser la Ville : Formes Urbaines et Politiques de Transport*, chapitre 4, pages 119–164. Economica (avec le soutien du PRÉDIT), Paris, 1<sup>re</sup> édition. ISBN 978-2-7178-5971-3.
- NICOLAS, J.-P., BONNEL, P., CABRERA DELGADO, J. E., GODINOT, C., HOMO-CIANU, M., ROUTHIER, J.-L., TOILIER, F. et ZUCCARELLO, P. (2009). Simuler les mobilités pour une agglomération durable. rapport final du projet simbad. Rapport de recherche du Laboratoire d'Économie des Transports, pour le compte de la DRI (MEEDDAT) et de l'ADEME dans le cadre du groupe 11 du PRÉDIT 3. Avec le soutien de l'Agence d'Urbanisme pour le développement de l'agglomération lyonnaise.
- NOUREDDINE, M. (2011). Finalisation du réseau de transports collectifs lyonnais à ces différents horizons : 1985, 1995 et 2006, sous DAVISUM. Projet de Fin de Formation. École Hassania des Travaux Publics.
- OPENSHAW, S. (1976). An empirical study of some spatial interaction models. *Environment and Planning A*, Vol. 8 (No. 1), pp. 23–41.
- OPENSHAW, S. (1977). Optimal zoning systems for spatial interaction models. *Environment and Planning A*, Vol. 9 (No. 2), pp. 169–184.
- OPENSHAW, S. (1979). Alternative methods of estimating spatial interaction models and their performance in short-term forecasting. Dans BARTELS, C. P. A. et KETELLAPPER, R. H., éditeurs : *Exploratory and Explanatory Statistical Analysis of Spatial Data*, chapitre 8, pages 201–225. Martinus Nijhoff Publishing, Boston, The Hague, London. ISBN 0-89838-004-9.
- OPENSHAW, S. (1984). *The Modifiable Areal Unit Problem: Concepts and Techniques in Modern Geography*. Geo Books, Norwich, Royaume-Uni. ISBN 0-860-94134-5.
- OPENSHAW, S. (1998). Neural network, genetic, and fuzzy logic models of spatial interaction. *Environment and Planning A*, Vol. 30 (No. 10), pp. 1857–1872.
- ORFEUIL, J.-P. (2002). État des lieux des recherches sur la mobilité quotidienne en France. Dans LÉVY, J.-P. et DUREAU, F., éditeurs : *L'Accès à la Ville : Les Mobilités Spatiales en Questions*, pages 65–98. L'Harmattan, Paris, 1<sup>re</sup> édition. ISBN 2-7475-2806-5.
- ORTÚZAR, J. de D. et WILLUMSEN, L. G. (2001). *Modelling Transport*. John Wiley & Sons, Chichester, Royaume-Uni, 3<sup>e</sup> édition. ISBN 0-471-861103.
- PARODY, T. E. (1977). Analysis of predictive qualities of disaggregate modal-choice models. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 637, pp. 51–57.

- PAS, E. I. (1978). An empirical comparison of zonal, household and personal models of home-based trip generation. *Traffic Engineering and Control*, Vol. 19 (No. 2), pp. 64–68.
- PITTION, J. (2006). Module d'affectation TC du projet SIMBAD. Travail de Fin d'Études. École Nationale des Travaux Publics de l'État.
- QIAN, Z. et ZANG, H. M. (2012). On centroid connectors in static traffic assignment: Their effects on flow patterns and how to optimize their selections. *Transportation Research Part B: Methodological*, Vol. 46 (No. 10), pp. 1489–1503.
- R DEVELOPMENT CORE TEAM (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- RAUX, C., MA, T.-Y., JOLY, I., KAUFMANN, V., CORNELIS, E. et OVTRACHT, N. (2011). Travel and activity time allocation: An empirical comparison between eight cities in Europe. *Transport Policy*, Vol. 18 (No. 2), pp. 401–412.
- RICHARDSON, A. J. (2003). Behavioral mechanisms of nonresponse in mail-back travel surveys. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1493, pp. 191–199.
- SAMMER, G. (1997). Problems and solutions in urban travel survey. Dans BONNEL, P., CHAPLEAU, R., LEE-GOSSELIN, M. E. H. et RAUX, C., éditeurs : *Les Enquêtes de Déplacements Urbains : Mesurer le Présent, Simuler le Futur*, pages 145–159. Éditions du Programme Rhône-Alpes Recherches en Sciences Humaines, Lyon. ISBN 2-909604-31-4.
- SEMALY (2000). TERESE, modèle d'affectation de voyageurs dans les études de transport collectif. Université Lumière Lyon 2 et ENTPE. Document pédagogique du cours Analyse et Prévision de la Demande de Transport du DESS Transports Urbains et Régionaux de Personnes. Lyon.
- SEN, A. et SMITH, T. E. (1995). *Gravity Models of Spatial Interaction Behavior*. Springer, Berlin. ISBN 3-540-60026-4.
- SILMAN, L. A. (1981). The time stability of a modal-split model for Tel-Aviv. *Environment and Planning A*, Vol. 13 (No. 6), pp. 751–762.
- SKALLI, M. (2006). Test d'un modèle d'affectation des transports collectifs : élaboration du réseau des transports collectifs lyonnais sous DAVISUM sur la base de données disponibles, et comparaison des résultats d'affectation issus de cette approche avec ceux de TERESE. Projet de Fin de Formation. École Hassania des Travaux Publics.
- SMITH, D. P. et HUTCHINSON, B. G. (1981). Goodness of fit statistics for trip distribution models. *Transportation Research Part A: General*, Vol. 15 (No. 4), pp. 295–303.

- SMITH, Jr., R. L. et CLEVELAND, D. E. (1976). Time-stability analysis of trip-generation and predistribution modal-choice models. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 569, pp. 76–86.
- SNEDECOR, G. W. et COCHRAN, W. G. (1989). *Statistical Methods*. Iowa State University Press, Ames, IA. États-Unis, 8<sup>e</sup> édition. ISBN 0 8138 1561 6.
- SNICKARS, F. et WEIBULL, J. W. (1977). A minimum information principle: Theory and practice. *Regional Science and Urban Economics*, Vol. 7 (No. 1-2), pp. 137–168.
- SOUTHWORTH, F. (1979). Spatial structure and parameter disaggregation in trip distribution models. *Regional Studies*, Vol. 13 (No. 4), pp. 381–394.
- SOUTHWORTH, F. (1983). Temporal versus other impacts upon trip distribution model parameter values. *Regional Studies*, Vol. 17 (No. 1), pp. 41–47.
- STOPHER, P. R. et McDONALD, K. G. (1983). Trip generation by cross-classification: An alternative methodology. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 944, pp. 84–91.
- STRAMBI, O. et VAN DE BILT, K.-A. (1998). Trip generation modeling using CHAID, a criterion-based segmentation modeling tool. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1645, pp. 24–31.
- SUN, Y. (2009). *Lifecycle Stage, Automobility Cohort and Travel: Probing into Structural Change in Urban Travel*. A dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Engineering, Kyoto University, Department of Urban Management, Kyoto. 139 p.
- SUPERNAK, J., TALVITIE, A. et DEJOHN, A. (1983). Person-category trip-generation model. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 944, pp. 74–83.
- SYTRAL (2007a). Enquête ménages déplacement 2006 : Les principaux enseignements pour l'agglomération lyonnaise. Disponible en ligne ([http://www.sytral.fr/include/viewFile.php?idtf=801&path=51%2F801\\_384\\_EMD-Enseignements-principaux.pdf](http://www.sytral.fr/include/viewFile.php?idtf=801&path=51%2F801_384_EMD-Enseignements-principaux.pdf), consulté le 24 avril 2013).
- SYTRAL (2007b). Enquête ménages déplacement 2006 de l'aire métropolitaine lyonnaise : Principaux résultats. Disponible en ligne ([http://www.sytral.fr/include/viewFile.php?idtf=810&path=83%2F810\\_385\\_EMD-Plaqueette.pdf](http://www.sytral.fr/include/viewFile.php?idtf=810&path=83%2F810_385_EMD-Plaqueette.pdf), consulté le 24 avril 2013).
- TALVITIE, A. et KIRSHNER, D. (1978). Specification, transferability and the effect of data outliers in modeling the choice of mode in urban travel. *Transportation*, Vol. 7 (No. 3), pp. 311–331.
- TANNER, J. C. (1961). Factors affecting the amount of travel. Road Research Technical Paper No. 5. Her Majesty's Stationery Office, London.

- THERNEAU, T. M. et ATKINSON, E. J. (2011). An introduction to recursive partitioning using rpart routines. Rapport technique 61, Mayo Foundation, Rochester. Disponible en ligne (<http://tolstoy.newcastle.edu.au/R/e13/help/att-0860/zed.pdf>), consulté le 24 avril 2013).
- TIPPING, D. G. (1968). Time savings in transport studies. *The Economic Journal*, Vol. 78 (No. 312), pp. 843–854.
- TRAIN, K. E. (1978). A validation test of a disaggregate mode choice model. *Transportation Research*, Vol. 12 (No. 3), pp. 164–174.
- TRAIN, K. E. (1979). A comparison of the predictive ability of mode choice models with various levels of complexity. *Transportation Research Part A: General*, Vol. 13 (No. 1), pp. 11–16.
- TRAIN, K. E. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press, New York, 2<sup>e</sup> édition. ISBN 978-0-521-76655-5.
- TRB (2007). *Metropolitan Travel Forecasting: Current Practice and Future Direction*. Transportation Research Board. Special Report 288. National Academy of Sciences, Washington D.C. ISBN 978-0-309-10417-3.
- VAN EVERT, H. et MORITZ, G. (2000). The new dutch travel survey. Paper for the 9th International Association for Travel Behaviour Conference. Gold Coast, Queensland, Australia. 2–7 juillet.
- VIEGAS, J. M., MARTÍNEZ, L. M. et SILVA, E. A. (2009). Effects of the modifiable areal unit problem on the delineation of traffic analysis zones. *Environment and Planning B: Planning and Design*, Vol. 36 (No. 4), pp. 625–643.
- VOLET, P. et HUTCHINSON, B. G. (1986). Explanatory and forecasting capabilities of trip-distribution models. *Canadian Journal of Civil Engineering*, Vol. 13 (No. 6), pp. 666–673.
- WALKER, W. T. et PENG, H. (1991). Long-range temporal stability of trip generation rates based on selected cross-classification models in the Delaware Valley Region. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1305, pp. 61–71.
- WHITE, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, Vol. 48 (No. 4), pp. 817–838.
- WILLIAMS, I. (1976). A comparison of some calibration techniques for doubly constrained models with an exponential cost function. *Transportation Research*, Vol. 10 (No. 2), pp. 91–104.
- WILSON, A. G. (1967). A statistical theory of spatial distribution models. *Transportation Research*, Vol. 1 (No. 3), pp. 253–269.
- YUNKER, K. R. (1976). Test of temporal stability of travel simulation models in Southeastern Wisconsin. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 610, pp. 1–5.

ZHAO, Y. et KOCKELMAN, K. M. (2002). The propagation of uncertainty through travel demand models: An exploratory analysis. *The Annals of Regional Science*, Vol. 36 (No. 1), pp. 145–163.

ZIMMERMAN, C. A. (1982). The life cycle concept as a tool for travel research. *Transportation*, Vol. 11 (No. 1), pp. 51–69.

# Annexes



# A Compléments méthodologiques

## A.1 Dérivation du modèle gravitaire selon l'approche classique dite de maximisation de l'entropie

Nous faisons ici une présentation de la méthode de dérivation classique — celle de l'état le plus probable encore dite de maximisation de l'entropie (Wilson, 1967) — du modèle gravitaire avec une fonction de résistance de type exponentielle puissance ou Tanner (1961), en empruntant à Ortúzar et Willumsen (2001).

On considère un système constitué par un grand nombre d'éléments distincts. En fonction des besoins de l'analyste, ce système peut être décrit à plusieurs échelles. La description du système consiste en la caractérisation de ses différents « états ». Une description complète (exhaustive) à l'échelle des différents éléments constitutifs du système correspond ainsi à la spécification de ce qu'on peut désigner comme ses « micro-états ». Ceci reviendrait, par exemple, à identifier, pour chaque personne de l'aire d'étude, l'origine, la destination, le mode, le temps de déplacement, etc. de chacun de ses déplacements. Dans la pratique, une connaissance aussi fine du système n'est pas indispensable et il peut suffire de le décrire à une échelle plus agrégée et spécifier ce qu'on pourrait appeler des « meso-états ». On se contenterait alors, pour le système d'exemple évoqué précédemment, d'identifier le nombre de déplacements pour tous les couples de zones d'origine et de destination. Il convient de noter que, en général, un méso-état peut correspondre à plusieurs micro-états possibles. Par exemple, si le méso-état est un déplacement entre une origine et une destination déterminées, celui-ci a pu être réalisé par n'importe quelle personne de l'aire étudiée (et donc correspondre à plusieurs micro-états du système). Si l'on s'intéresse à une échelle d'analyse encore plus agrégée la description du système passera, à ce moment là, par la caractérisation de ce que l'on peut appeler des « macro-états ». Ceux-ci sont caractérisés par le fait que leur spécification est relativement aisée compte tenu des informations disponibles. Ils peuvent correspondre, par exemple, au nombre d'émissions ou d'attractions par zone.

Le principe fondamental de la méthode de maximisation de l'entropie consiste en ceci : sauf indication du contraire, tous les micro-états du système pouvant correspondre à l'information que l'on a sur les macro-états de celui-ci sont équiprobables. À partir de cette hypothèse, on détermine le méso-état le plus probable, satisfaisant les informations disponibles sur les macro-états du système. Notons que, d'un point de vue pratique, dans la distinction entre meso et macro-états est de caractère fonctionnelle. Le méso-état du système est la description que l'analyste entreprend de réaliser.

Les macro-états sont des contraintes à respecter dans la description du système.

Wilson (1967) montre que le nombre de micro-états  $W(T_{ij})$  correspondant au meso-état  $T_{ij}$ , donnant le nombre de déplacements (ou de demi-sorties<sup>1</sup>) modélisés entre la zone d'origine  $i$  et la zone de destination  $j$ , est donné par :

$$W(T_{ij}) = \frac{T!}{\prod_{ij} T_{ij}!} \quad (\text{A.1})$$

où  $T = \sum_{ij} T_{ij}$ . comme tous les micro-états sont équiprobables, le meso-état le plus probable est celui qui peut être réalisé par le plus grand nombre de micro-états différents. On cherche donc à maximiser l'équation (A.1) ce qui peut être accompli en maximisant son logarithme (fonction monotone croissante), ce qui simplifie le problème. On cherche alors le maximum de

$$\log W(T_{ij}) = \log \frac{T!}{\prod_{ij} T_{ij}!} = \log T! - \sum_{ij} \log T_{ij}!. \quad (\text{A.2})$$

L'équation (A.2) peut encore être simplifiée à l'aide de l'approximation de Stirling pour  $\log X! = X \log X - X$  de telle sorte que la maximisation s'effectue sur

$$\log W(T_{ij}) = \log T! - \sum_{ij} (T_{ij} \log T_{ij} - T_{ij}). \quad (\text{A.3})$$

Dans la mesure où  $\log T!$  est généralement une constante, la maximisation du terme de droite de (A.3) peut se faire simplement sur

$$\log W'(T_{ij}) = - \sum_{ij} (T_{ij} \log T_{ij} - T_{ij}). \quad (\text{A.4})$$

connue comme la fonction d'entropie.

La dérivation de la forme gravitaire passe par l'ajout de contraintes au problème d'optimisation. Dans ce qui suit nous allons donner les contraintes nécessaires à la dérivation du modèle de distribution « doublement contraint » avec une fonction de résistance exponentielle-puissance, encore connue sous le nom de Tanner (1961), qui est utilisée dans le chapitre 5. Les deux premières contraintes à introduire correspondent au fait que le nombre de déplacements émis et attirés par chacune des différentes zones est connu (il est déterminé dans l'étape de génération). Les contraintes peuvent s'écrire :

$$\sum_j T_{ij} - O_i = 0 \quad (\text{A.5})$$

$$\sum_i T_{ij} - D_j = 0 \quad (\text{A.6})$$

où  $O_i$  est le nombre de déplacements (demi-sorties) émis par la zone  $i$  et  $D_j$  est le nombre de déplacements (demi-sorties) attirés par la zone  $j$ . Les équations (A.5) et

---

1. L'unité d'analyse de la mobilité retenue dans les applications des modèles du présent travail est bien entendu la sortie, ou plus précisément, la demi-sortie. Voir le chapitre 4 pour la définition d'une demi-sortie et la justification du choix.

(A.6) sont ce que l'on appelle les « contraintes aux marges » de la matrice origine-destination. La dénomination « doublement contraint » provient de leur utilisation dans la construction du modèle. Ensuite, on introduit des contraintes liées au système de transport prenant la forme de contraintes sur le budget temps (ou plutôt temps généralisé) total consacré aux déplacements sur l'aire étudiée :

$$\sum_{ij} T_{ij} \log c_{ij} - C' = 0 \quad (\text{A.7})$$

$$\sum_{ij} T_{ij} c_{ij} - C = 0 \quad (\text{A.8})$$

où  $c_{ij}$  est le temps généralisé pour aller de la zone  $i$  à la zone  $j$ .  $C$  et  $C'$  sont des dépenses totale et log-somme, respectivement, en unités de temps généralisé.

La maximisation de  $\log W'(T_{ij})$  sous les contraintes (A.5), (A.6), (A.7) et (A.8) peut se faire par la méthode du lagrangien de sorte qu'elle équivaut à maximiser sans contraintes la fonction  $L$  prenant la forme

$$\begin{aligned} L(T_{ij}, \theta'_i, \theta''_j, \alpha, \beta) &= \log W'(T_{ij}) + \sum_i \theta'_i \left( \sum_j T_{ij} - O_i \right) + \sum_j \theta''_j \left( \sum_i T_{ij} - D_j \right) \\ &+ \alpha \left( \sum_{ij} T_{ij} \log c_{ij} - C' \right) + \beta \left( \sum_{ij} T_{ij} c_{ij} - C \right). \end{aligned} \quad (\text{A.9})$$

La dérivée partielle de  $L$  par rapport à  $T_{ij}$  est égale à<sup>2</sup>

$$\frac{\partial L}{\partial T_{ij}} = -\log T_{ij} + \theta'_i + \theta''_j + \alpha \log c_{ij} + \beta c_{ij} = 0, \quad (\text{A.10})$$

ce qui permet d'écrire

$$\begin{aligned} T_{ij} &= \exp(\theta'_i + \theta''_j + \alpha \log c_{ij} + \beta c_{ij}) \\ &= \exp(\theta'_i) \exp(\theta''_j) c_{ij}^\alpha \exp(\beta c_{ij}). \end{aligned} \quad (\text{A.11})$$

On peut finalement réécrire l'équation (A.11) de manière à obtenir les équations caractérisant le modèle gravitaire

$$T_{ij} = A_i O_i B_j D_j c_{ij}^\alpha \exp(\beta c_{ij}) \quad (\text{A.12})$$

en appliquant le changement de variables suivant pour déterminer les multiplicateurs de Lagrange relatifs aux contraintes (A.5) et (A.6) :

$$A_i O_i = \exp(\theta'_i) \quad \text{et} \quad B_j D_j = \exp(\theta''_j)$$

avec

$$A_i = \left( \sum_j B_j D_j c_{ij}^\alpha \exp(\beta c_{ij}) \right)^{-1} \quad (\text{A.13})$$

---

2. Les autres conditions du premier ordre, complémentaires à celle donnée dans l'équation (A.10), sont évidemment égales aux contraintes (A.5), (A.6), (A.7) et (A.8).

et

$$B_j = \left( \sum_i A_i O_i c_{ij}^\alpha \exp(\beta c_{ij}) \right)^{-1}. \quad (\text{A.14})$$

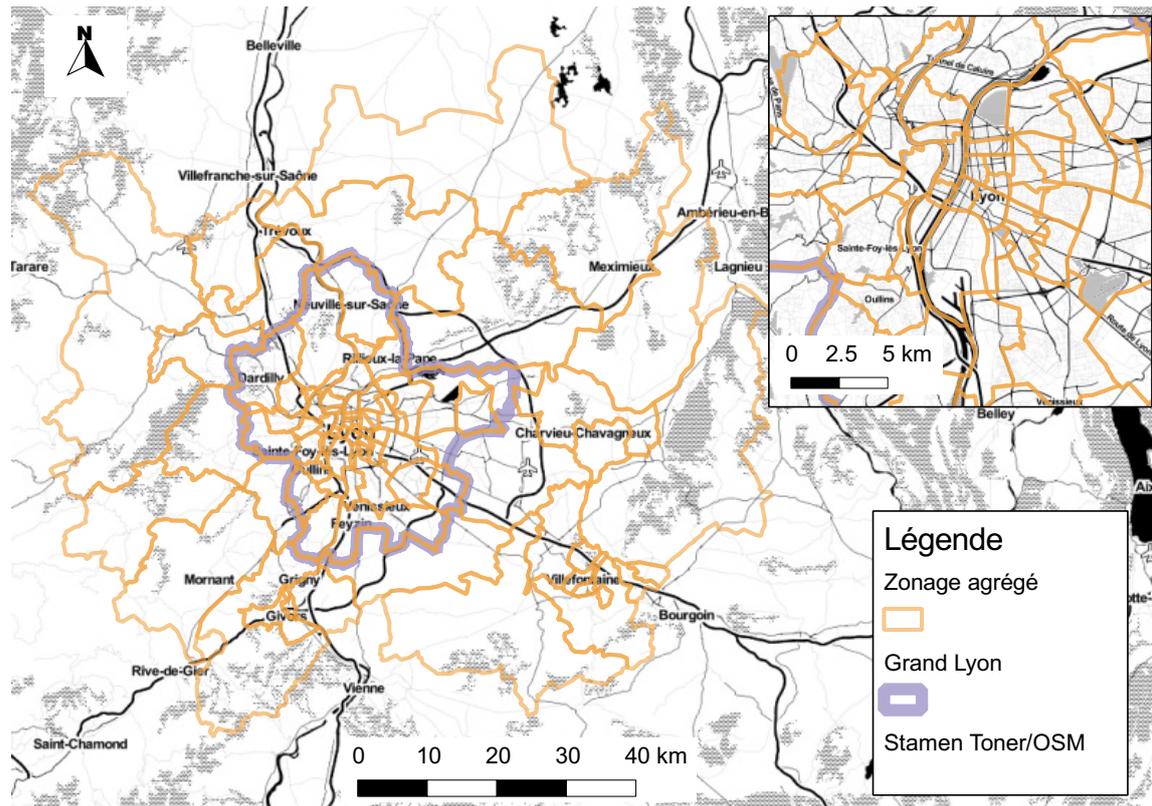
Le modèle gravitaire doublement contraint avec une fonction de résistance de type exponentielle-puissance est donné par les équations (A.12), (A.13), et (A.14). Notons que les paramètres  $\alpha$  et  $\beta$ , caractérisant la fonction de résistance sont les multiplicateurs de Lagrange des contraintes (A.7) et (A.8). Ortúzar et Willumsen (2001) remarquent que  $C$  et  $C'$  ne peuvent, en général, seulement être estimés et, de ce fait, les paramètres  $\alpha$  et  $\beta$  sont des paramètres à calibrer. Pour que l'éloignement ait un impact négatif sur le nombre de déplacement entre deux zones, il faut que  $\beta$  soit négatif;  $\alpha$  peut être positif ou négatif. Kirby (1974) montre que les contraintes (A.5), (A.6), (A.7) et (A.8) permettent de définir de manière unique les paramètres de la fonction de résistance, alors que les  $A_i$  et  $B_j$  ne sont uniques qu'à un coefficient près (il faut fixer la valeur de l'un d'entre eux pour une zone et alors les autres sont déterminés de manière unique).

La maximisation de l'entropie est une approche avantageuse pour la génération de modèles de par la flexibilité qu'elle apporte. Son utilisation permet de spécifier de manière plus rigoureuse les propriétés du modèle, ce qui facilite l'analyse de faisabilité des solutions (Ortúzar et Willumsen, 2001). En revanche, il s'agit bien d'une approche descriptive, permettant de caractériser un état du système en fonction de l'information disponible. De ce fait, l'utilisation d'un modèle dérivé de cette manière pour la prévision pose des questions de cohérence de l'analyse (Erlander et Stewart, 1990).

## A.2 Zonage agrégé commun pour la génération et le choix modal

La figure A.1 montre le niveau d'agrégation du zonage retenu pour les calculs des indicateurs réalisés dans les chapitres sur la génération (chapitre 4) et le choix modal (chapitre 6).

FIGURE A.1 – Carte du zonage commun pour l'analyse des modèles de génération et choix modal.



Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006. Fond de carte « Toner » conçu par Stamen Design à partir des données d'OpenStreetMap.

## B Définition de la variable « cycle de vie »

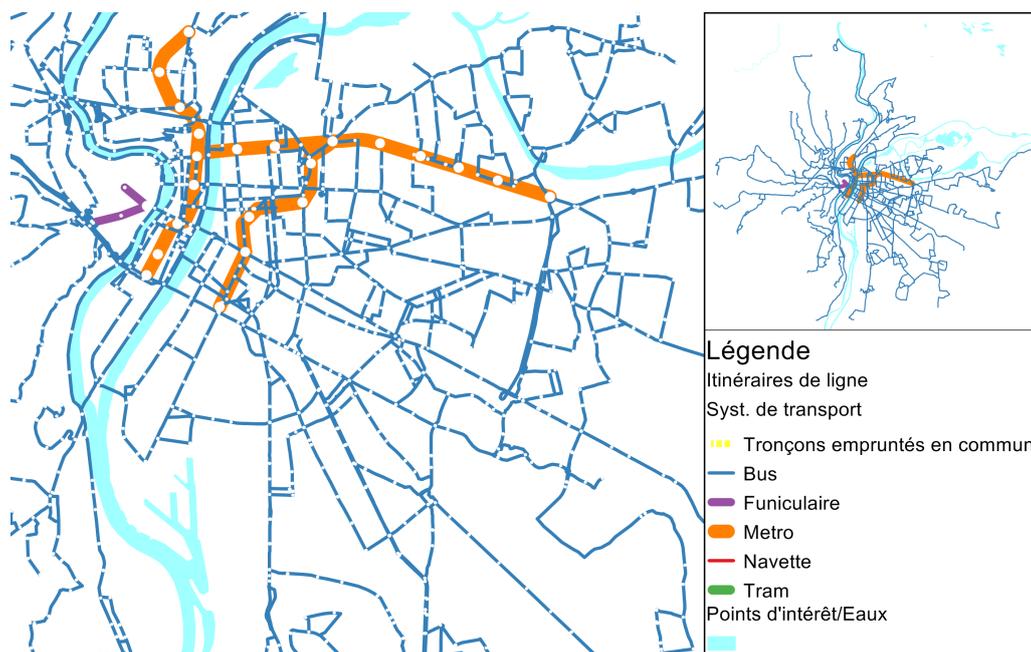
La variable « cycle de vie » a été introduite dans l'analyse en raison de l'évidence croissante de l'impact fort qu'ont les interactions au sein du ménage sur les activités — et, en conséquence, sur la mobilité — de ses membres. Pour caractériser la structure interne du ménage en termes de possibilités de coordination, nous avons retenu une définition du cycle de vie du ménage à partir de l'âge de certains de ses membres et de certains liens de filiation. Nous nous intéressons notamment à l'existence d'un lien conjugal et à la présence d'enfants. En effet, nous supposons que l'existence du lien conjugal augmente la probabilité de mutualisation des ressources et de répartition des tâches au sein du ménage (par rapport à simplement deux adultes qui cohabitent dans un même logement). La présence d'enfants peut suggérer l'existence de déplacements supplémentaires à effectuer par leurs parents, afin de satisfaire à certains des besoins des premiers. Elle suggère également une prise de décision centralisée au niveau du couple ou de la personne de référence du ménage. Nous avons retenu finalement une variable comprenant 7 stades, donnés ci-dessous :

1. le premier stade est celui des ménages dont la personne de référence a moins de 45 ans, pour lesquels il n'y a pas de lien conjugal entre les différents membres du ménage et où il n'y a pas d'enfants. Le lien conjugal et la présence d'enfants sont définis à partir de la variable « position dans le ménage » donnée dans la table « personnes », des différentes EMD ;
2. le deuxième stade est celui des ménages dont la personne de référence a moins de 45 ans, où il y a un lien conjugal mais pas d'enfants ;
3. le troisième stade dans le cycle de vie est caractérisé uniquement par la présence d'enfants et l'âge du benjamin : inférieur à 5 ans ;
4. le quatrième stade dans le cycle de vie est caractérisé par la présence d'enfants et l'âge du benjamin : supérieur ou égal à 5 ans et inférieur à 15 ans ;
5. le cinquième stade dans le cycle de vie est caractérisé par la présence d'enfants et l'âge du benjamin : supérieur ou égal 15 ans ;
6. le sixième stade est celui des ménages dont la personne de référence a au moins 45 ans, où il y a un lien conjugal mais pas d'enfants ;
7. le septième stade est celui des ménages dont la personne de référence a au moins 45 ans, où il n'y a pas de lien conjugal ni de présence d'enfants.

Une telle définition de la variable « cycle de vie » reproduit à peu près celle de Kitamura et Kostyniuk (1986).

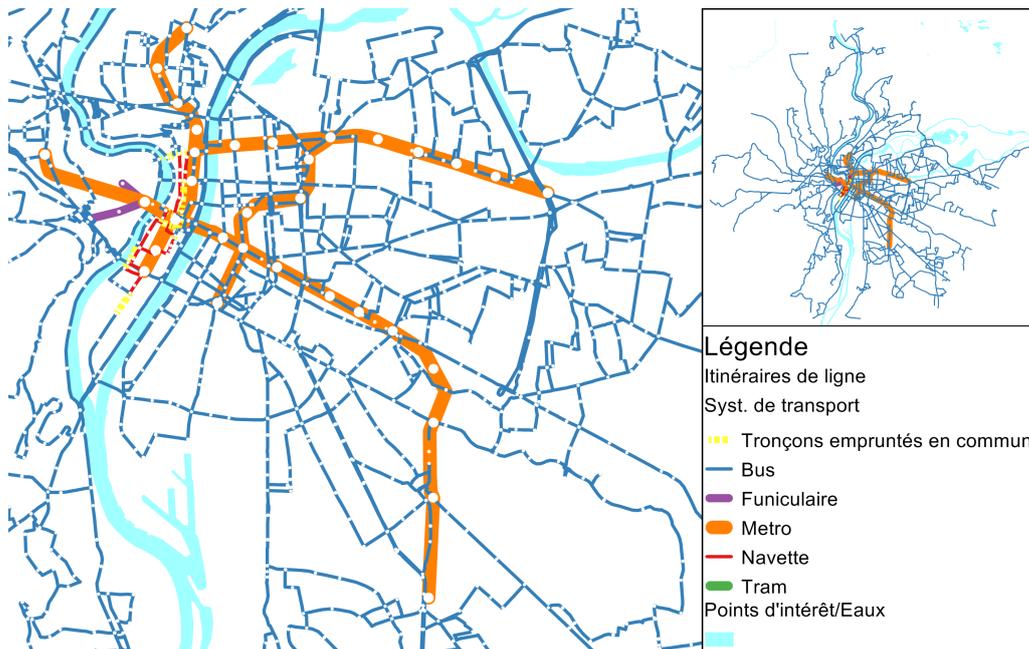
## C Réseau TC modélisé

FIGURE C.1 – Capture d'écran du réseau TC de 1985 modélisé sous DAVISUM.



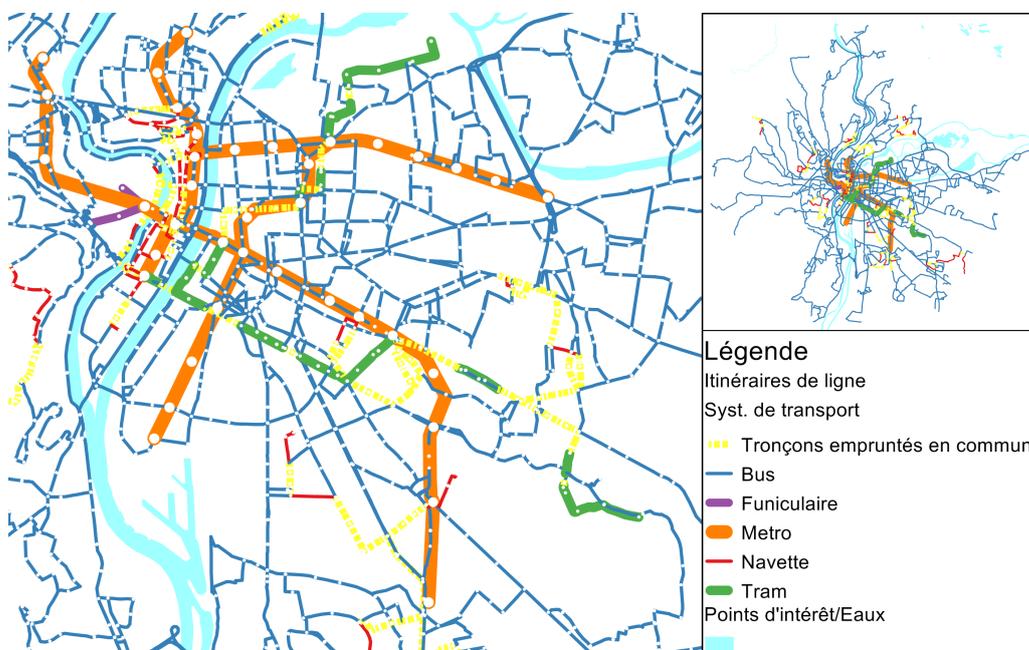
*Source : Reconstitution à partir des plans et des guides du réseau TCL de 1985, consultés au fonds Rhône-Alpes de la Bibliothèque Municipale de Lyon.*

FIGURE C.2 – Capture d'écran du réseau TC de 1995 modélisé sous DAVISUM.



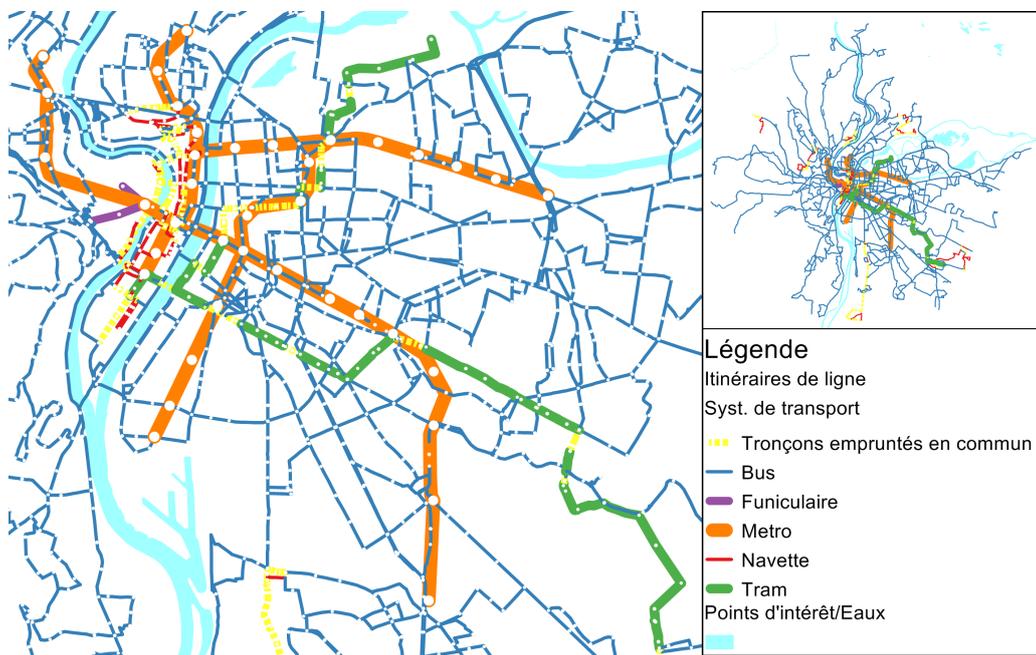
Source : Reconstitution à partir des plans et des guides du réseau TCL de 1995, consultés au fonds Rhône-Alpes de la Bibliothèque Municipale de Lyon.

FIGURE C.3 – Capture d'écran du réseau TC de 2001 modélisé sous DAVISUM.



Source : Reconstitution à partir des plans et des guides du réseau TCL de 2001, consultés au fonds Rhône-Alpes de la Bibliothèque Municipale de Lyon.

FIGURE C.4 – Capture d'écran du réseau TC de 2006 modélisé sous DAVISUM.



*Source : Reconstitution à partir des plans et des guides du réseau TCL de 2006, consultés au fonds Rhône-Alpes de la Bibliothèque Municipale de Lyon.*

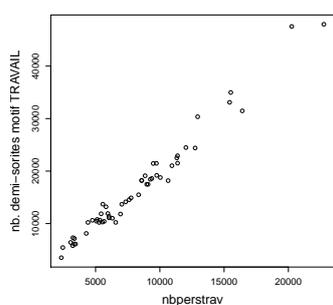
# D Compléments sur les estimations des modèles de génération

## D.1 Modèle de génération du projet SIMBAD

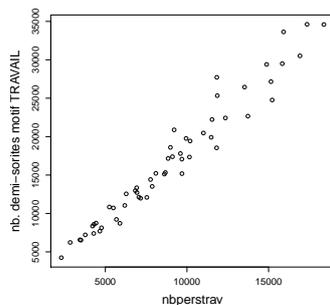
Le module de génération des déplacements du projet SIMBAD, auquel le présent travail est lié, fait appel à une procédure de segmentation par la méthode CART. Celle-ci fait partie d'une nouvelle approche de la génération combinant une segmentation des profils de mobilité des EMD avec une procédure de micro-simulation aléatoire inspirée de la méthode *hot-deck* de traitement de la non-réponse dans les enquêtes par sondage (Gitton, 2006 ; Nicolas *et al.*, 2009). On peut résumer la procédure de la manière suivante : on dispose d'une population simulée, sur l'agglomération lyonnaise, à laquelle on souhaite donner une mobilité en accord avec ce qui a pu être observé dans les EMD. On définit des segments de population (en termes de caractéristiques des ménages) en appliquant l'algorithme CART aux données des EMD et en utilisant un profil de mobilité comme variable à expliquer. Le profil de mobilité est défini en termes de nombre de sorties d'un certain type (voir la section 4.1.1.2.1) réalisées dans la journée (un profil peut être, par exemple, « 2 sorties pour le motif travail et une sortie pour le motif loisirs »). Le fait de donner à la variable à expliquer une forme qualitative permet de conserver l'information de chainage des déplacements au sein des sorties. Une fois que les segments de population ont été définis, on connaît le segment auquel chaque ménage de la population simulée appartient. On peut donc lui associer un comportement vraisemblable tiré au sort parmi les comportements observés sur le segment. Le tirage au sort se fait selon la loi de probabilité empirique observée. Ainsi, à l'aléa de la simulation près, la loi de probabilité des comportements de déplacements simulés sur ce segment sera identique à celle observée sur le segment de référence. (Gitton, 2006) note que l'avantage de ce type de modélisation est qu'il permet de reproduire une certaine variabilité des comportements en n'écartant pas les comportements rares et en leur accordant toujours la même place.

## D.2 Représentation graphique des données pour la régression zonale

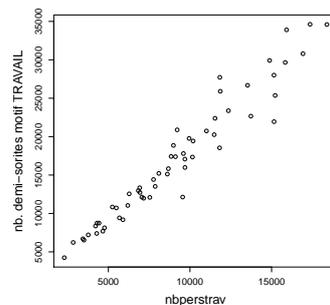
FIGURE D.1 – Relation entre le nombre de personnes ayant un emploi et le nombre de demi-sorties pour le motif Travail.



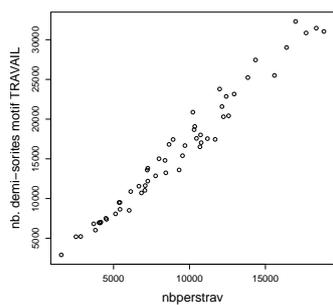
(a) EMD 1985



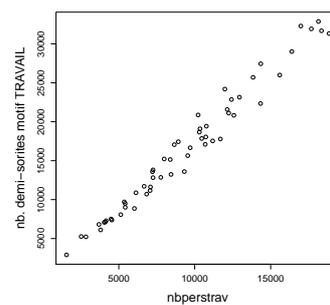
(b) EMD 1995 Périmètre 1985



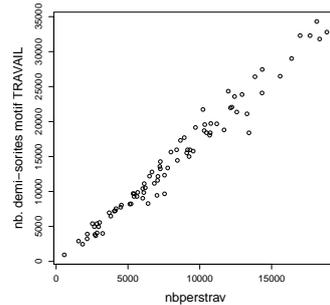
(c) EMD 1995 Périmètre 1995



(d) EMD 2006 Périmètre 1985



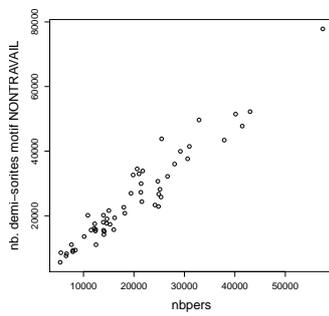
(e) EMD 2006 Périmètre 1995



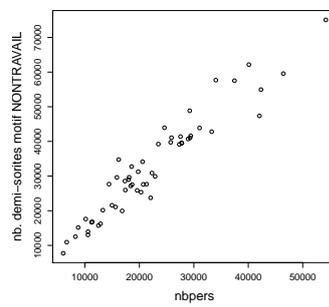
(f) EMD 2006 Périmètre AUL

Source : EMD 1985, 1995 et 2006.

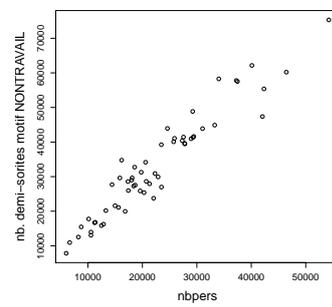
FIGURE D.2 – Relation entre le nombre de personnes de la zone et le nombre de demi-sorties pour les motifs autres que Travail et Enseignement.



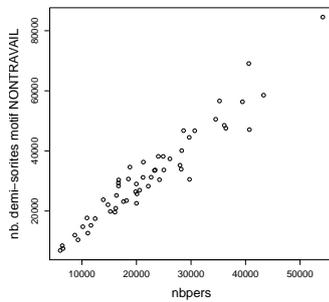
(a) EMD 1985



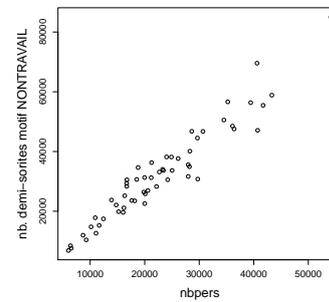
(b) EMD 1995 Périmètre 1985



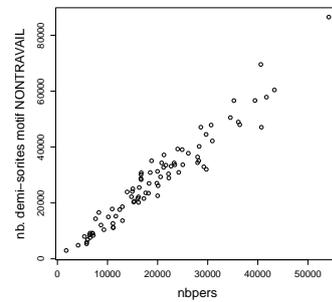
(c) EMD 1995 Périmètre 1995



(d) EMD 2006 Périmètre 1985



(e) EMD 2006 Périmètre 1995



(f) EMD 2006 Périmètre AUL

Source : EMD 1985, 1995 et 2006.

### D.3 Résultats de la régression au niveau des ménages sans la prise en compte du plan de sondage

TABLE D.1 – Résultats de calage des modèles de génération pour le motif Travail à l'échelle des ménages (sans prise en compte explicite du plan de sondage) et test de stabilité temporelle des paramètres.

Variable	Coef.	écart-type	t-value	Pr(> t )	R <sup>2</sup> ajust.	SRMSEG	Transf. nég
Données : EMD 1985 (a)							
Spécification étendue							
nbhomtrav	1,90	0,048	40,0	4,50E-301	0,756	0,105	def
nbfemtrav	1,90	0,049	38,7	1,02E-284			bcdef
nbpermis	0,06	0,029	2,2	2,97E-02			
presfoyer	0,13	0,058	2,2	3,02E-02			
Spécification simple							
nbperstrav	1,98	0,016	120,1	0,00E+00	0,755	0,109	bcdef
Données : EMD 1995 Périm. 1985 (b)							
Spécification étendue							
nbhomtrav	1,84	0,038	48,7	0,00E+00	0,754	0,100	def
nbfemtrav	1,75	0,039	44,6	0,00E+00			adef
nbpermis	0,03	0,020	1,6	1,15E-01			
presfoyer	0,12	0,049	2,5	1,34E-02			
Spécification simple							
nbperstrav	1,84	0,014	133,5	0,00E+00	0,753	0,103	adef
Données : EMD 1995 Périm. 1995 (c)							
Spécification étendue							
nbhomtrav	1,83	0,037	48,9	0,00E+00	0,754	0,119	def
nbfemtrav	1,76	0,039	45,3	0,00E+00			adef
nbpermis	0,03	0,020	1,4	1,66E-01			
presfoyer	0,14	0,048	2,8	4,73E-03			
Spécification simple							
nbperstrav	1,84	0,014	135,3	0,00E+00	0,753	0,119	adef
Données : EMD 2006 Périm. 1985 (d)							
Spécification étendue							
nbhomtrav	1,66	0,035	47,0	0,00E+00	0,733	0,086	abc
nbfemtrav	1,58	0,035	44,7	0,00E+00			abc
nbpermis	0,06	0,018	3,7	2,50E-04			
presfoyer	0,03	0,060	0,5	5,91E-01			
Spécification simple							
nbperstrav	1,69	0,013	129,8	0,00E+00	0,732	0,089	abc
Données : EMD 2006 Périm. 1995 (e)							
Spécification étendue							
nbhomtrav	1,67	0,034	48,6	0,00E+00	0,741	0,081	abc
nbfemtrav	1,60	0,035	46,2	0,00E+00			abc
nbpermis	0,07	0,017	3,9	8,57E-05			
presfoyer	0,05	0,058	0,9	3,92E-01			
Spécification simple							
nbperstrav	1,71	0,013	136,4	0,00E+00	0,740	0,085	abc
Données : EMD 2006 Périm. AUL (f)							
Spécification étendue							
nbhomtrav	1,70	0,030	56,3	0,00E+00	0,744	0,101	abc
nbfemtrav	1,59	0,031	51,9	0,00E+00			abc
nbpermis	0,05	0,015	3,4	6,18E-04			
presfoyer	0,06	0,051	1,1	2,81E-01			
Spécification simple							
nbperstrav	1,70	0,011	160,4	0,00E+00	0,744	0,107	abc

Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006.

TABLE D.2 – Résultats de calage des modèles de génération pour les motifs autres que Travail et Enseignement à l'échelle des ménages (sans prise en compte explicite du plan de sondage) et test de stabilité temporelle des paramètres.

Variable	Coef.	écart-type	t-value	Pr(> t )	R <sup>2</sup> ajust.	SRMSEG	Transf. nég
Données : EMD 1985 (a)							
Spécification étendue							
nbhom	0,91	0,057	15,9	2,07E-55	0,575	0,140	
nbfem	0,74	0,051	14,4	3,88E-46			bcdef
presvp	0,42	0,106	3,9	8,05E-05			
presconj	-0,05	0,129	-0,4	6,95E-01			def
presfoyer	1,26	0,111	11,4	1,58E-29			f
presret	1,30	0,096	13,6	4,29E-41			def
preschom	1,30	0,166	7,9	4,40E-15			
Spécification simple							
nbpers	1,15	0,016	74,2	0,00E+00	0,540	0,191	bcdef
Données : EMD 1995 Périm. 1985 (b)							
Spécification étendue							
nbhom	0,90	0,053	16,8	3,69E-62	0,600	0,140	
nbfem	0,90	0,049	18,5	1,91E-74			a
presvp	0,47	0,095	5,0	6,12E-07			
presconj	0,29	0,117	2,5	1,42E-02			
presfoyer	1,09	0,114	9,5	3,15E-21			
presret	1,47	0,090	16,3	1,84E-58			f
Données : EMD 2006 Périm. 1985 (d)							
preschom	1,50	0,137	10,9	1,68E-27			
Spécification simple							
nbpers	1,32	0,015	86,3	0,00E+00	0,560	0,162	ade
Données : EMD 1995 Périm. 1995 (c)							
Spécification étendue							
nbhom	0,91	0,053	17,1	3,99E-64	0,600	0,136	
nbfem	0,90	0,048	18,7	6,14E-76			a
presvp	0,47	0,095	5,0	6,09E-07			
presconj	0,29	0,116	2,5	1,33E-02			
presfoyer	1,12	0,113	9,8	1,10E-22			
presret	1,44	0,090	16,1	5,33E-57			ef
preschom	1,49	0,137	10,9	2,75E-27			
Spécification simple							
nbpers	1,33	0,015	87,6	0,00E+00	0,561	0,160	ade
Spécification étendue							
nbhom	0,79	0,053	14,9	1,30E-49	0,588	0,168	
nbfem	0,92	0,048	19,3	1,29E-80			a
presvp	0,32	0,089	3,6	3,68E-04			
presconj	0,42	0,106	4,0	6,49E-05			a
presfoyer	0,95	0,140	6,8	1,46E-11			
presret	1,65	0,077	21,5	1,55E-98			a
preschom	1,45	0,137	10,6	5,96E-26			
Spécification simple							
nbpers	1,28	0,015	84,7	0,00E+00	0,538	0,180	abc
Données : EMD 2006 Périm. 1995 (e)							
Spécification étendue							
nbhom	0,82	0,052	15,8	2,82E-55	0,587	0,170	
nbfem	0,92	0,047	19,8	9,52E-85			a
presvp	0,31	0,087	3,6	3,60E-04			
presconj	0,39	0,104	3,7	2,14E-04			a
presfoyer	0,91	0,137	6,6	3,40E-11			
presret	1,69	0,075	22,4	1,73E-107			ac
preschom	1,43	0,135	10,6	4,26E-26			
Spécification simple							
nbpers	1,28	0,015	87,1	0,00E+00	0,537	0,178	abc
Données : EMD 2006 Périm. AUL (f)							
Spécification étendue							
nbhom	0,91	0,046	19,9	2,17E-86	0,590	0,174	
nbfem	0,94	0,041	22,7	1,78E-110			a
presvp	0,26	0,080	3,3	9,69E-04			
presconj	0,34	0,093	3,7	2,17E-04			a
presfoyer	0,90	0,119	7,5	5,00E-14			a
presret	1,71	0,068	25,1	2,44E-134			abc
preschom	1,44	0,120	11,9	1,40E-32			
Spécification simple							
nbpers	1,30	0,013	102,7	0,00E+00	0,544	0,181	a

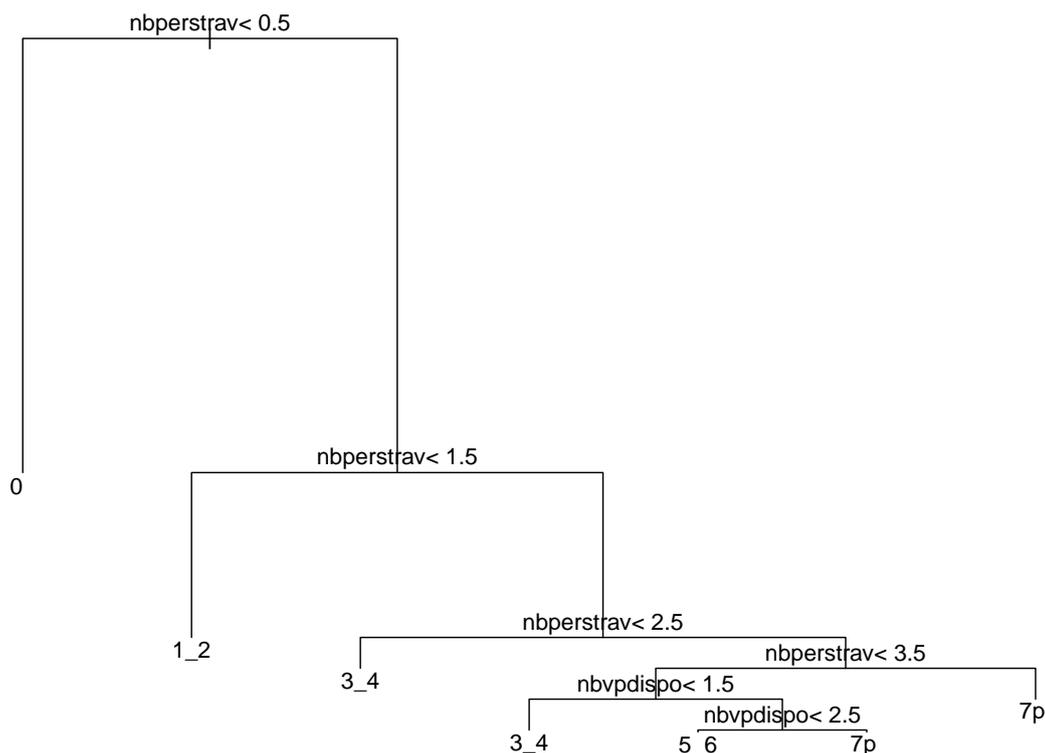
Source : traitement Cabrera Delgado à partir des EMD 1985, 1995 et 2006.

## D.4 Arbres de classification

### D.4.1 Motif Travail

#### D.4.1.1 Avec prise en compte des probabilités de tirage

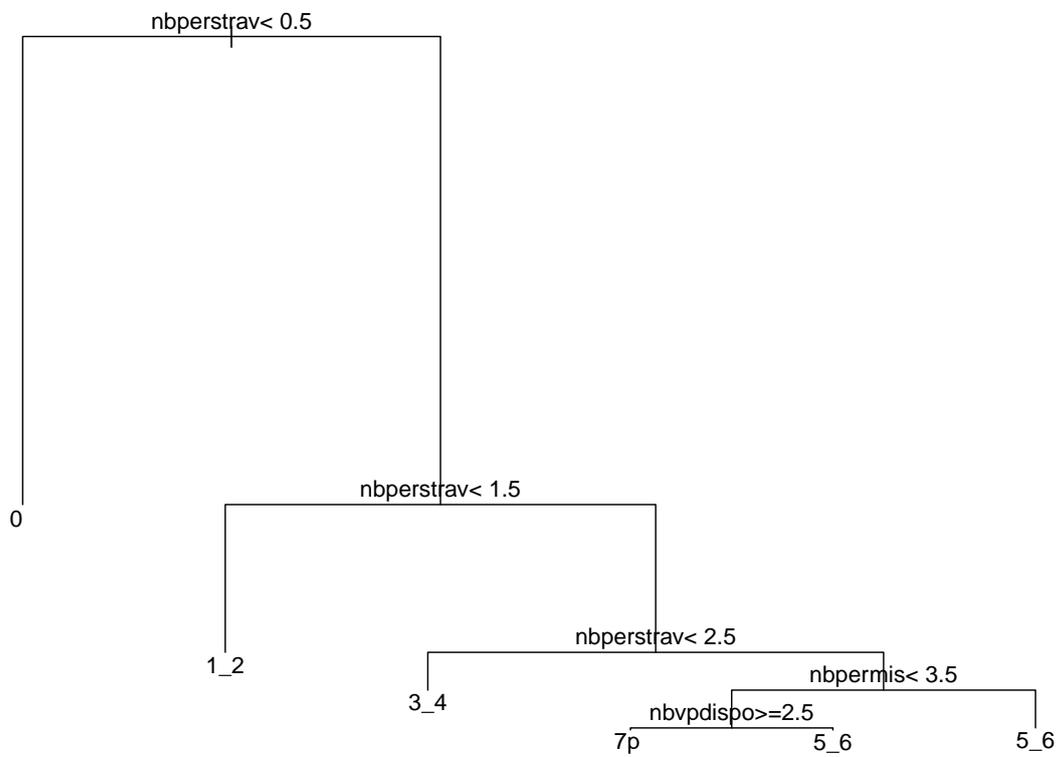
FIGURE D.3 – Arbre de classification pour le motif Travail construit à partir des données de l'EMD de 1985 (avec prise en compte des probabilités de tirage).



*Note de lecture : Sur chaque nœud intermédiaire l'indication relative à la segmentation concerne la branche de gauche. La longueur des branches est proportionnelle au pouvoir discriminant de la variable retenue pour la segmentation. Les nœuds terminaux indiquent le mode (la modalité la plus fréquente) de la variable à expliquer, correspondant aux segmentations successives. Ainsi, la branche la plus à gauche indique que le segment qui correspond aux ménages avec un nombre de personnes employées inférieur à 0,5 (soit les ménages n'ayant pas d'actifs employés) a un mode de 0 demi-sorties.*

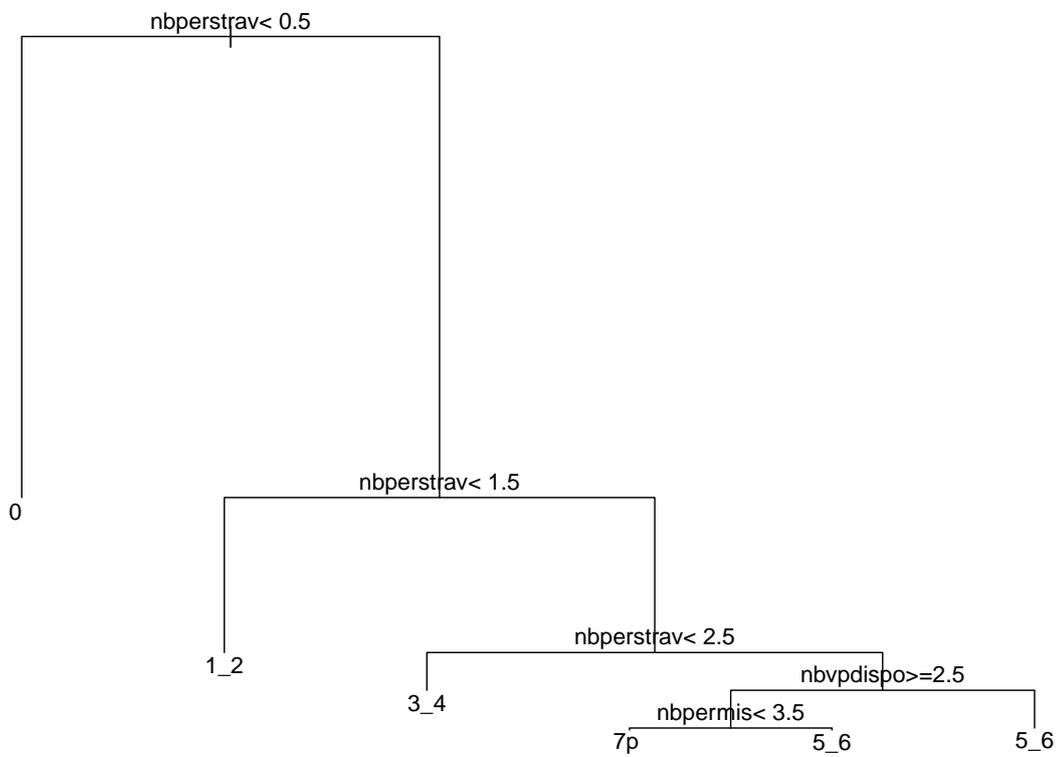
*Source : traitement Cabrera Delgado à partir de l'EMD 1985.*

FIGURE D.4 – Arbre de classification pour le motif Travail construit à partir des données de l'EMD de 1995 sur le périmètre de 1985 (avec prise en compte des probabilités de tirage).



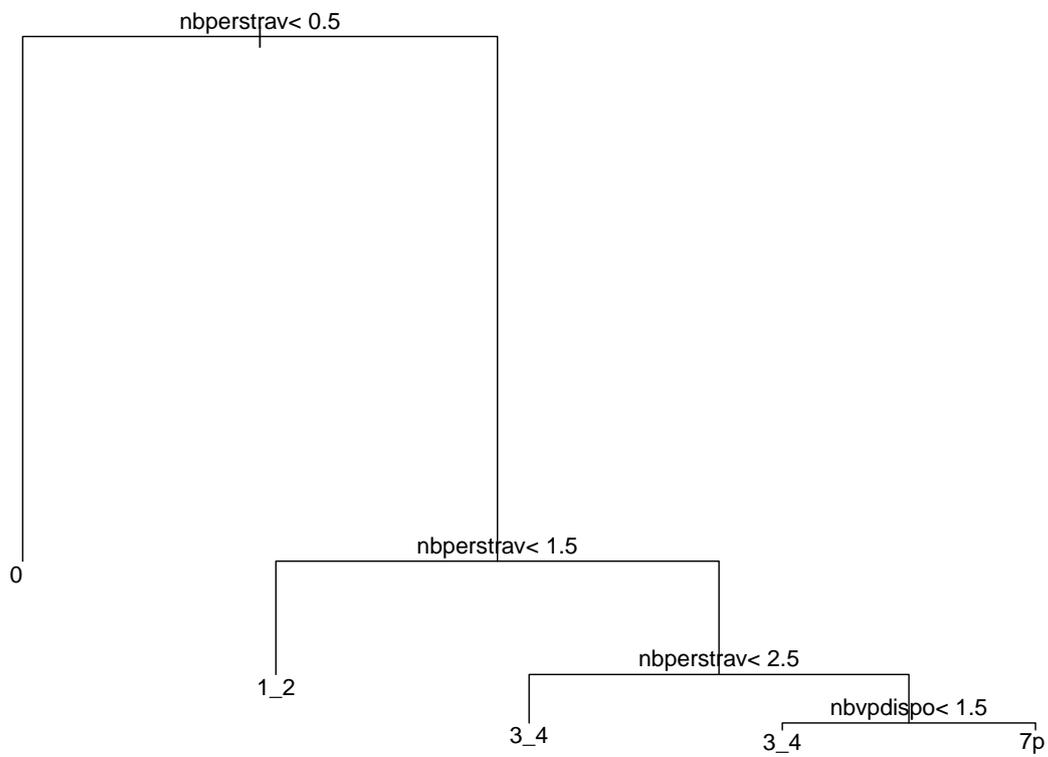
Source : traitement Cabrera Delgado à partir de l'EMD 1995.

FIGURE D.5 – Arbre de classification pour le motif Travail construit à partir des données de l'EMD de 1995 sur le périmètre de 1995 (avec prise en compte des probabilités de tirage).



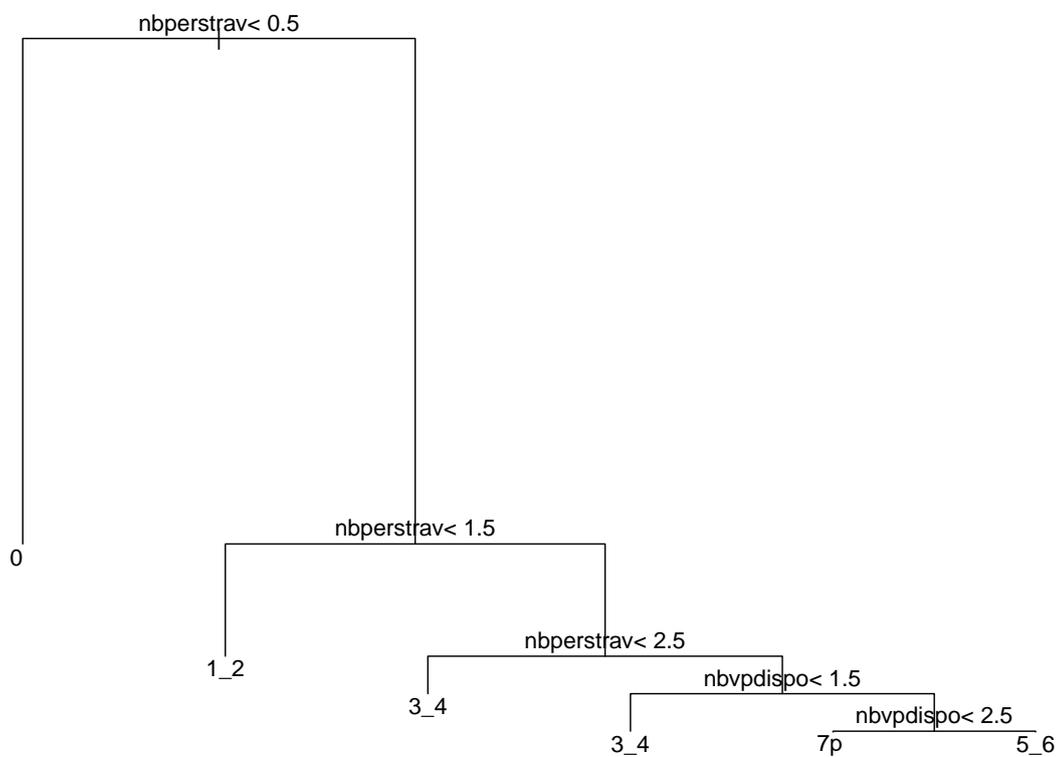
Source : traitement Cabrera Delgado à partir de l'EMD 1995.

FIGURE D.6 – Arbre de classification pour le motif Travail construit à partir des données de l'EMD de 2006 sur le périmètre de 1985 (avec prise en compte des probabilités de tirage).



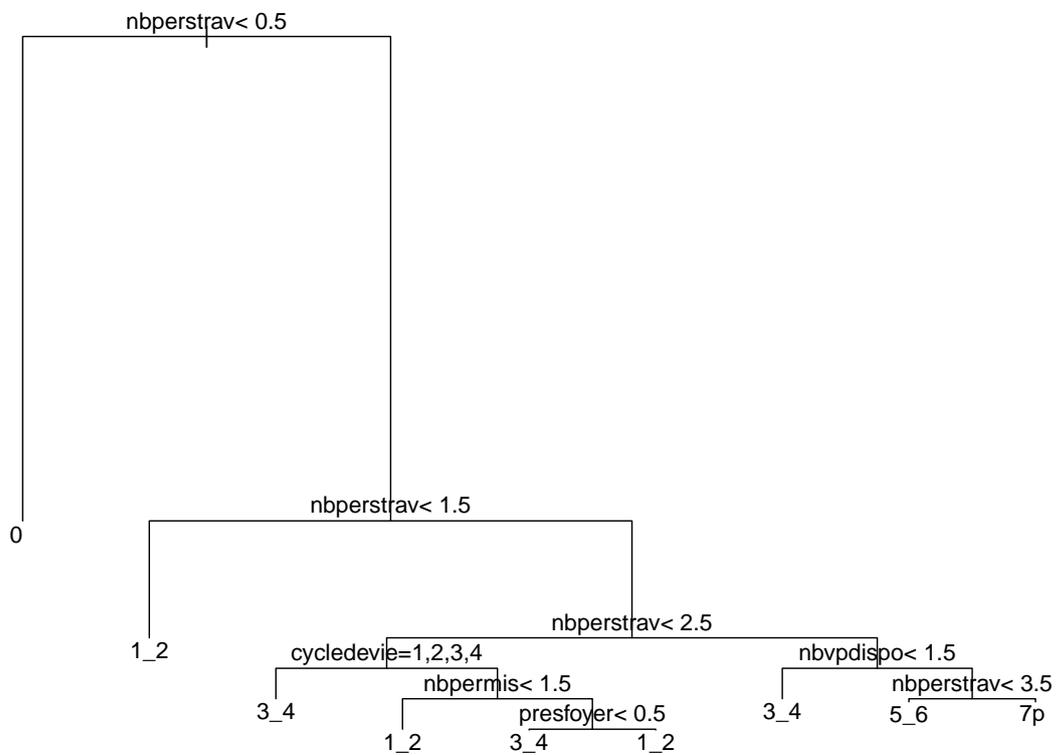
Source : traitement Cabrera Delgado à partir de l'EMD 2006.

FIGURE D.7 – Arbre de classification pour le motif Travail construit à partir des données de l'EMD de 2006 sur le périmètre de 1995 (avec prise en compte des probabilités de tirage).



Source : traitement Cabrera Delgado à partir de l'EMD 2006.

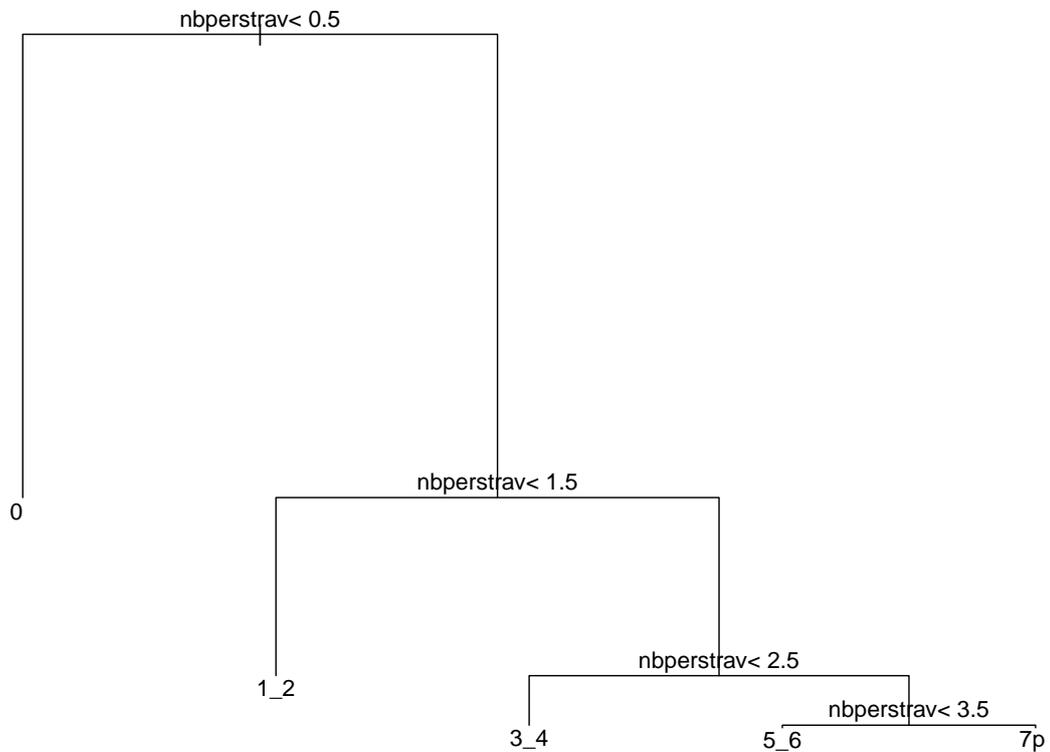
FIGURE D.8 – Arbre de classification pour le motif Travail construit à partir des données de l'EMD de 2006 sur le périmètre de l'aire urbaine lyonnaise (avec prise en compte des probabilités de tirage).



Source : traitement Cabrera Delgado à partir de l'EMD 2006.

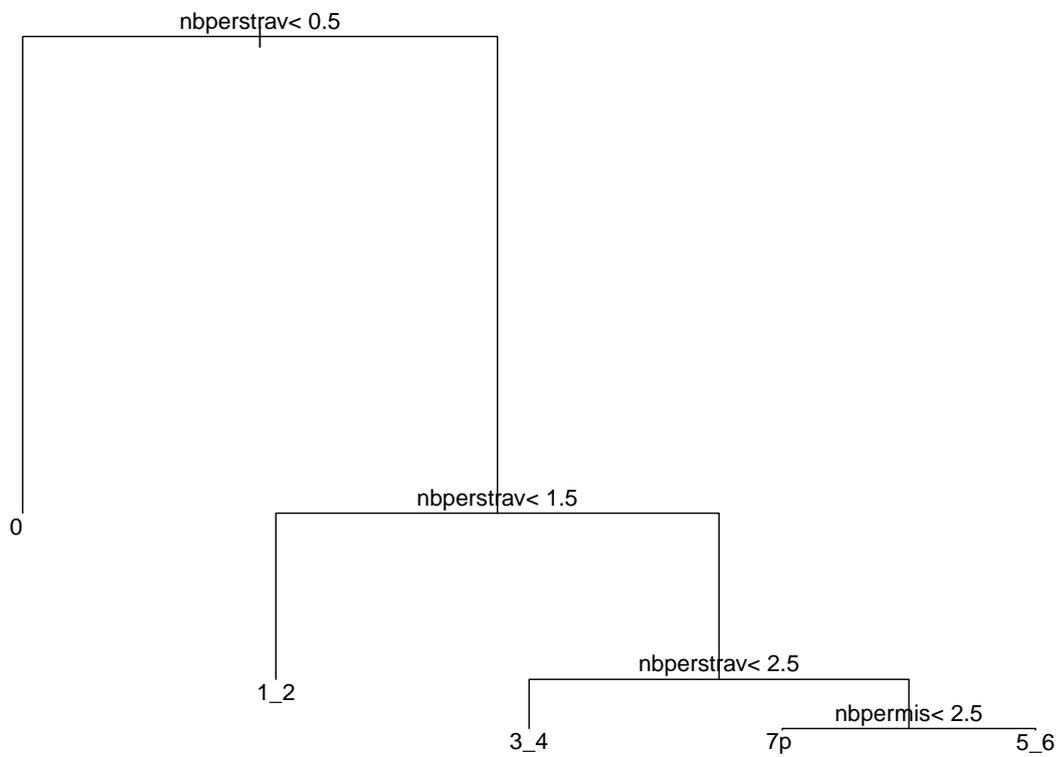
#### D.4.1.2 Sans prise en compte des probabilités de tirage

FIGURE D.9 – Arbre de classification pour le motif Travail construit à partir des données de l'EMD de 1985 (sans prise en compte des probabilités de tirage).



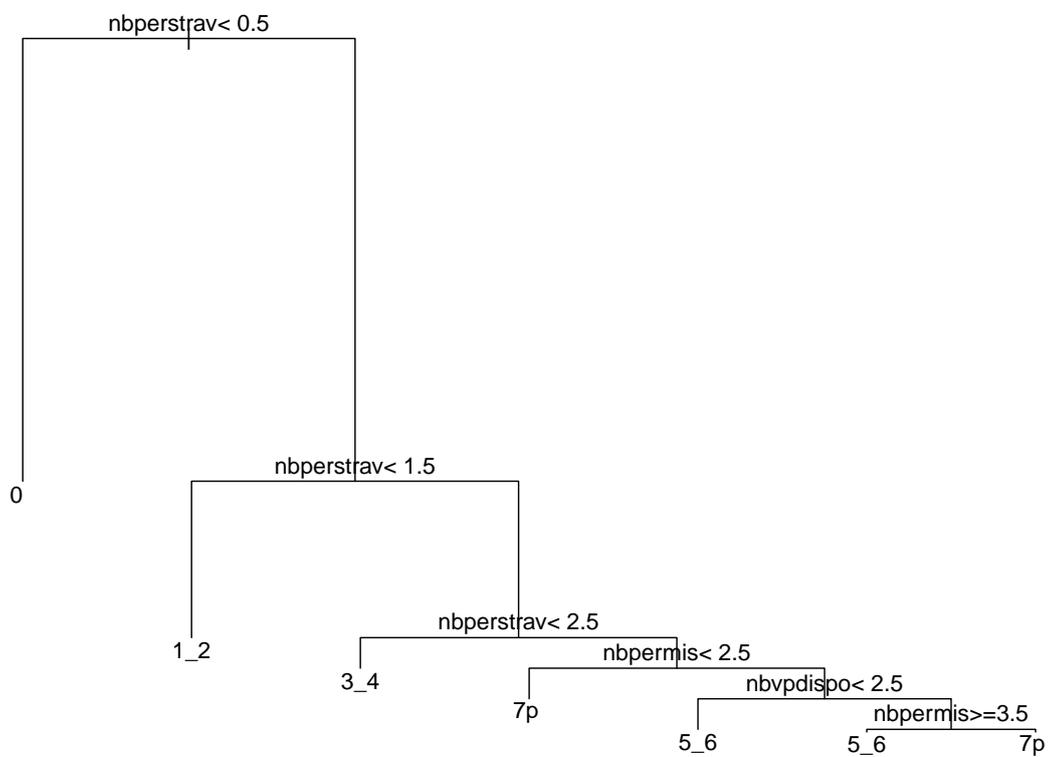
Source : traitement Cabrera Delgado à partir de l'EMD 1985.

FIGURE D.10 – Arbre de classification pour le motif Travail construit à partir des données de l'EMD de 1995 sur le périmètre de 1985 (sans prise en compte des probabilités de tirage).



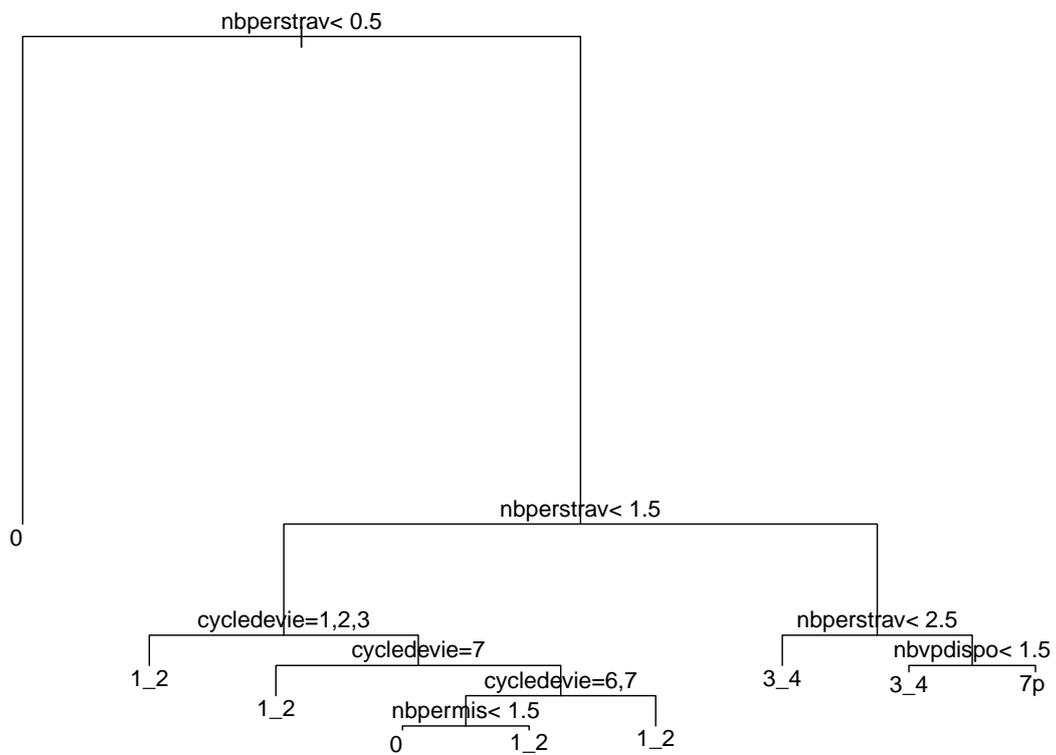
Source : traitement Cabrera Delgado à partir de l'EMD 1995.

FIGURE D.11 – Arbre de classification pour le motif Travail construit à partir des données de l'EMD de 1995 sur le périmètre de 1995 (sans prise en compte des probabilités de tirage).



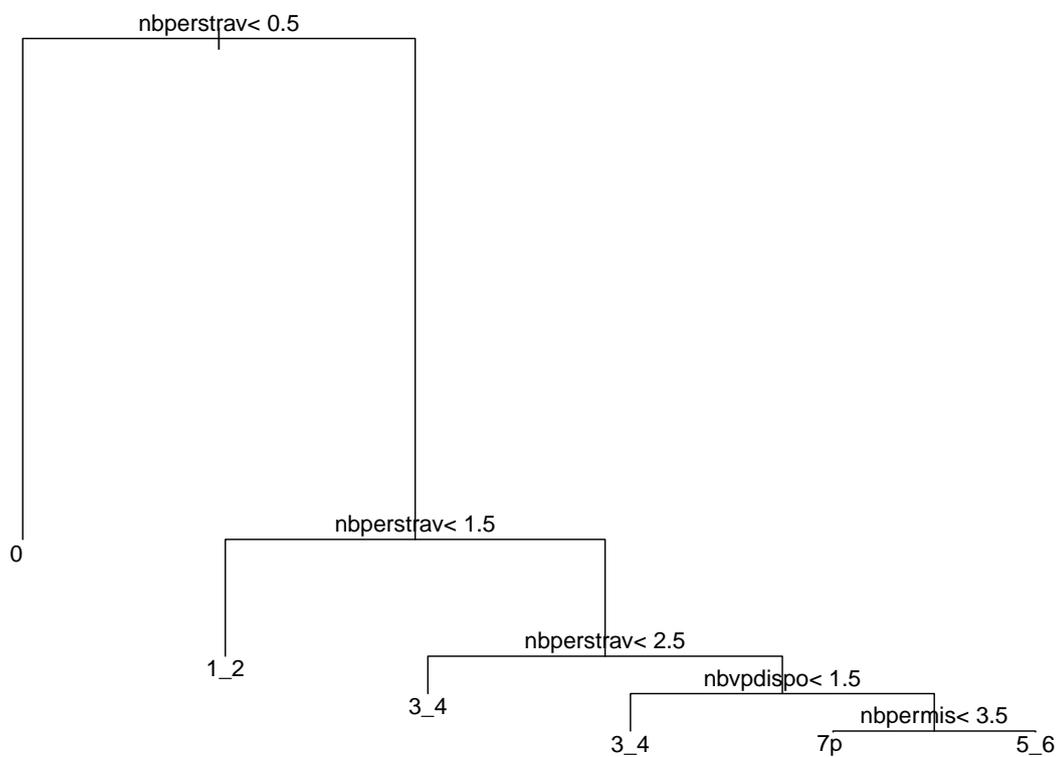
Source : traitement Cabrera Delgado à partir de l'EMD 1995.

FIGURE D.12 – Arbre de classification pour le motif Travail construit à partir des données de l’EMD de 2006 sur le périmètre de 1985 (sans prise en compte des probabilités de tirage).



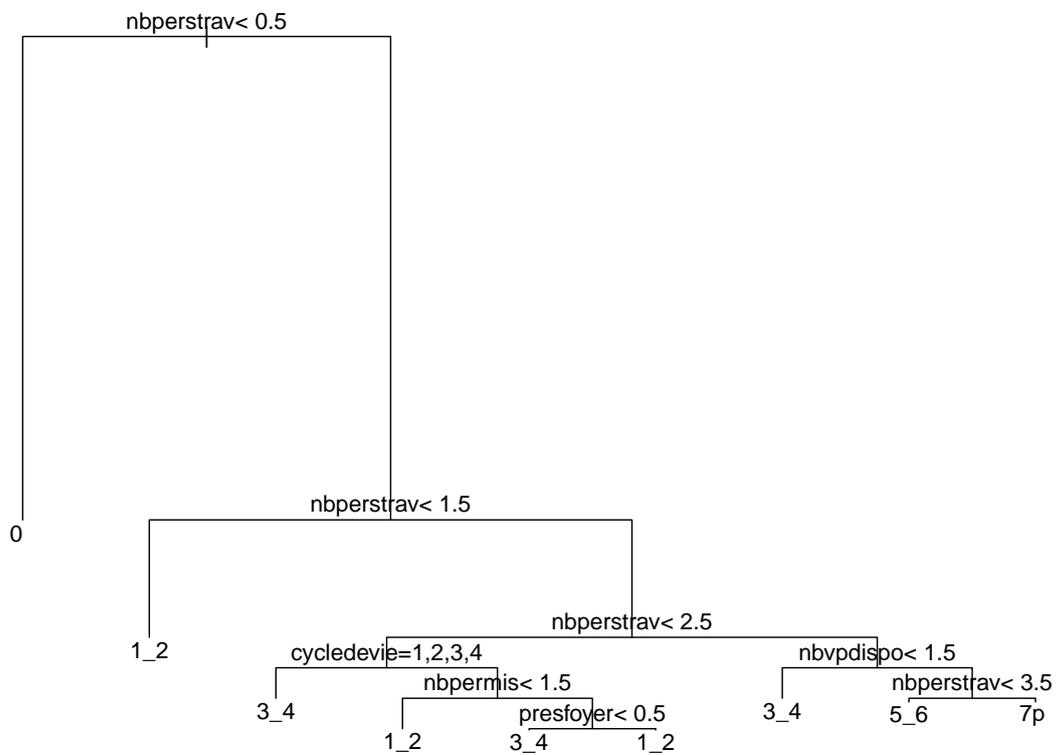
Source : traitement Cabrera Delgado à partir de l’EMD 2006.

FIGURE D.13 – Arbre de classification pour le motif Travail construit à partir des données de l'EMD de 2006 sur le périmètre de 1995 (sans prise en compte des probabilités de tirage).



Source : traitement Cabrera Delgado à partir de l'EMD 2006.

FIGURE D.14 – Arbre de classification pour le motif Travail construit à partir des données de l'EMD de 2006 sur le périmètre de l'aire urbaine lyonnaise (sans prise en compte des probabilités de tirage).

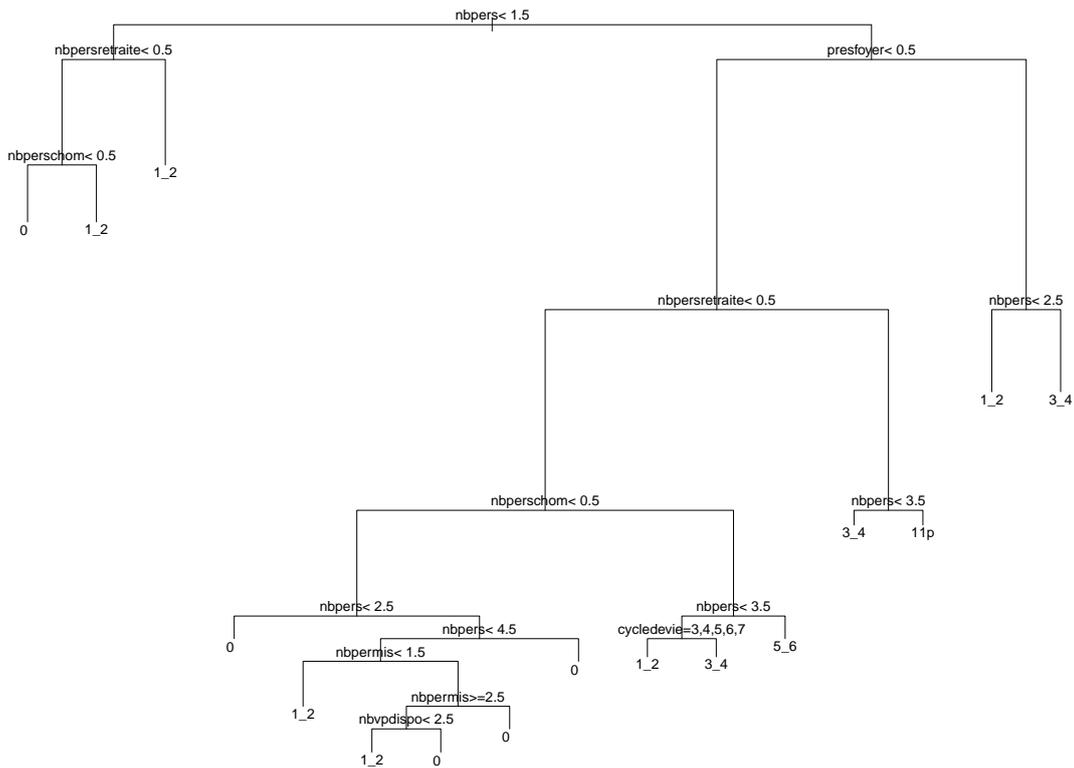


Source : traitement Cabrera Delgado à partir de l'EMD 2006.

## D.4.2 Autres motifs que Travail et Enseignement

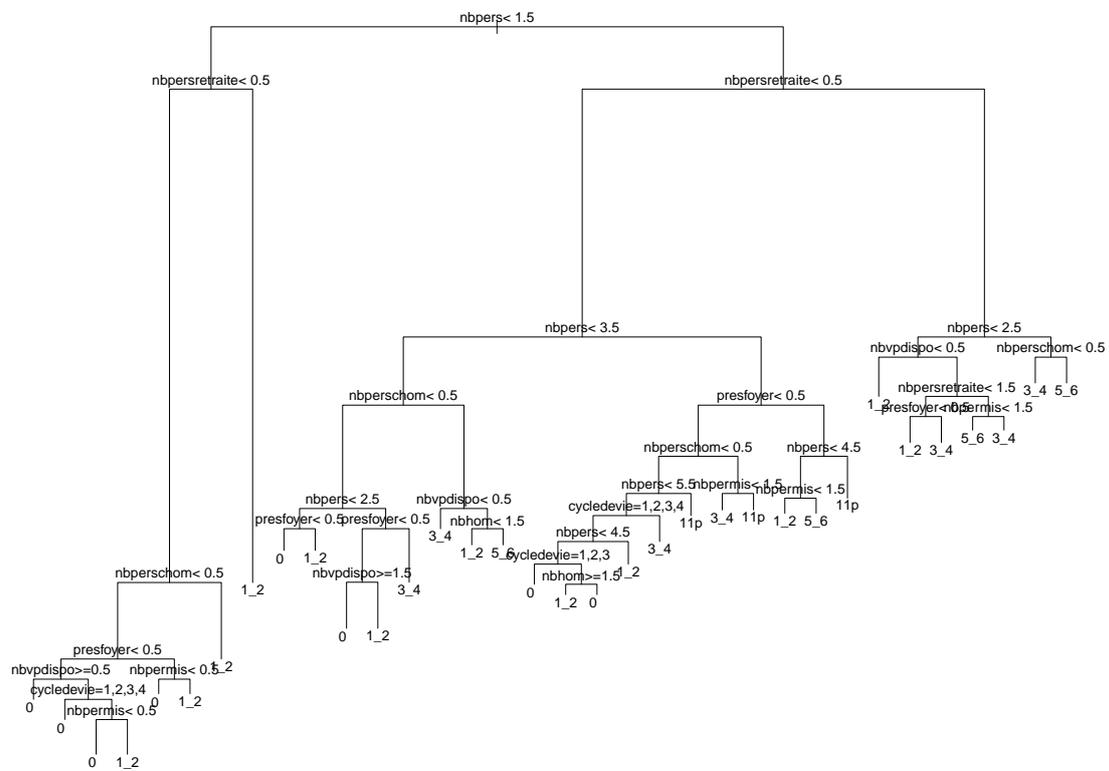
### D.4.2.1 Avec prise en compte des probabilités de tirage

FIGURE D.15 – Arbre de classification pour les motifs autres que Travail et Enseignement construit à partir des données de l'EMD de 1985 (avec prise en compte des probabilités de tirage).



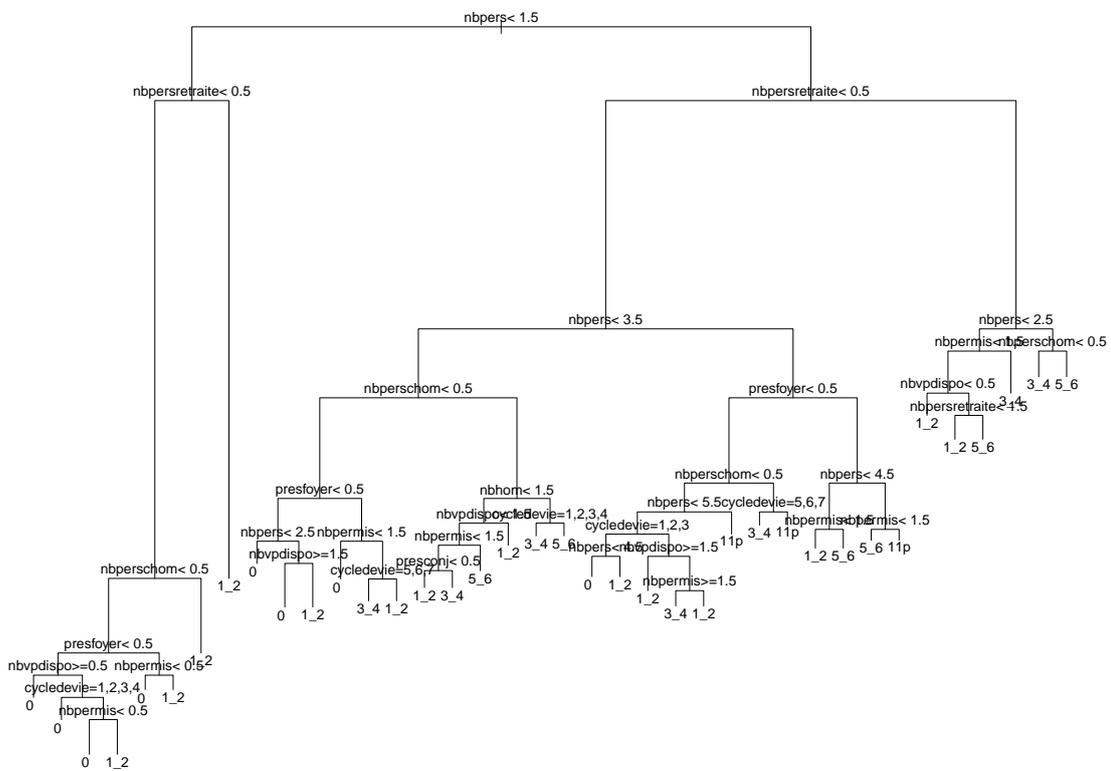
Source : traitement Cabrera Delgado à partir de l'EMD 1985.

FIGURE D.16 – Arbre de classification pour les motifs autres que Travail et Enseignement construit à partir des données de l’EMD de 1995 sur le périmètre de 1985 (avec prise en compte des probabilités de tirage).



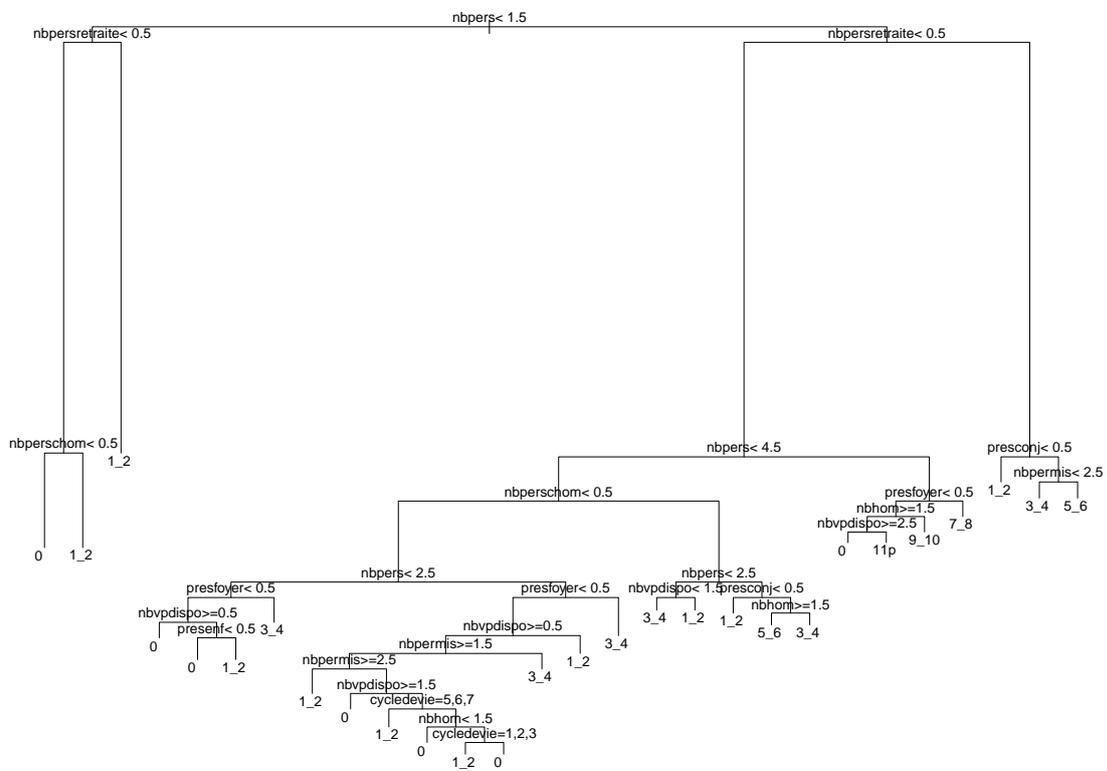
Source : traitement Cabrera Delgado à partir de l’EMD 1995.

FIGURE D.17 – Arbre de classification pour les motifs autres que Travail et Enseignement construit à partir des données de l’EMD de 1995 sur le périmètre de 1995 (avec prise en compte des probabilités de tirage).



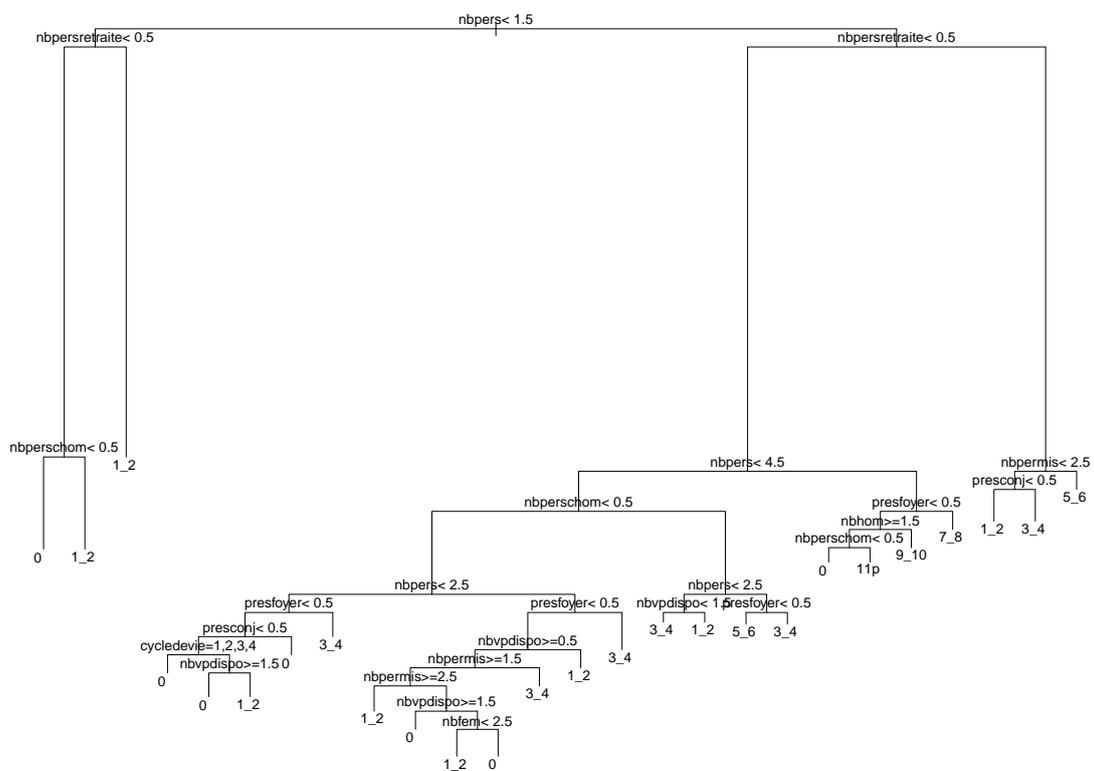
Source : traitement Cabrera Delgado à partir de l’EMD 1995.

FIGURE D.18 – Arbre de classification pour les motifs autres que Travail et Enseignement construit à partir des données de l’EMD de 2006 sur le périmètre de 1985 (avec prise en compte des probabilités de tirage).



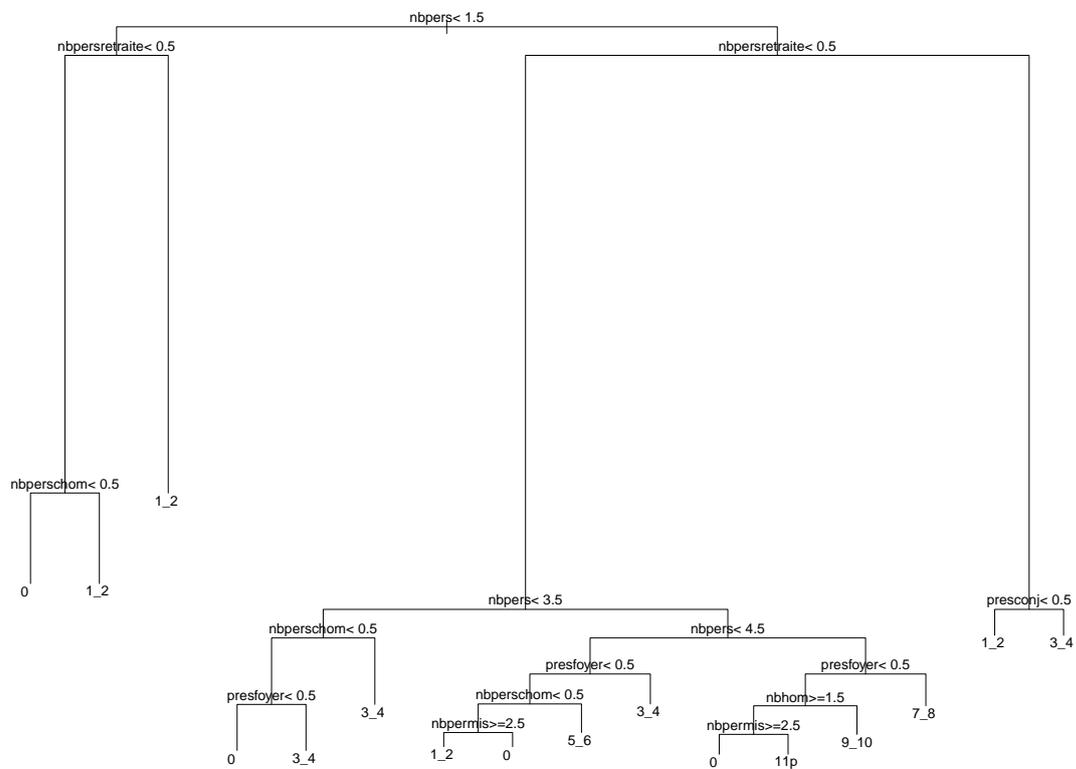
Source : traitement Cabrera Delgado à partir de l’EMD 2006.

FIGURE D.19 – Arbre de classification pour les motifs autres que Travail et Enseignement construit à partir des données de l’EMD de 2006 sur le périmètre de 1995 (avec prise en compte des probabilités de tirage).



Source : traitement Cabrera Delgado à partir de l’EMD 2006.

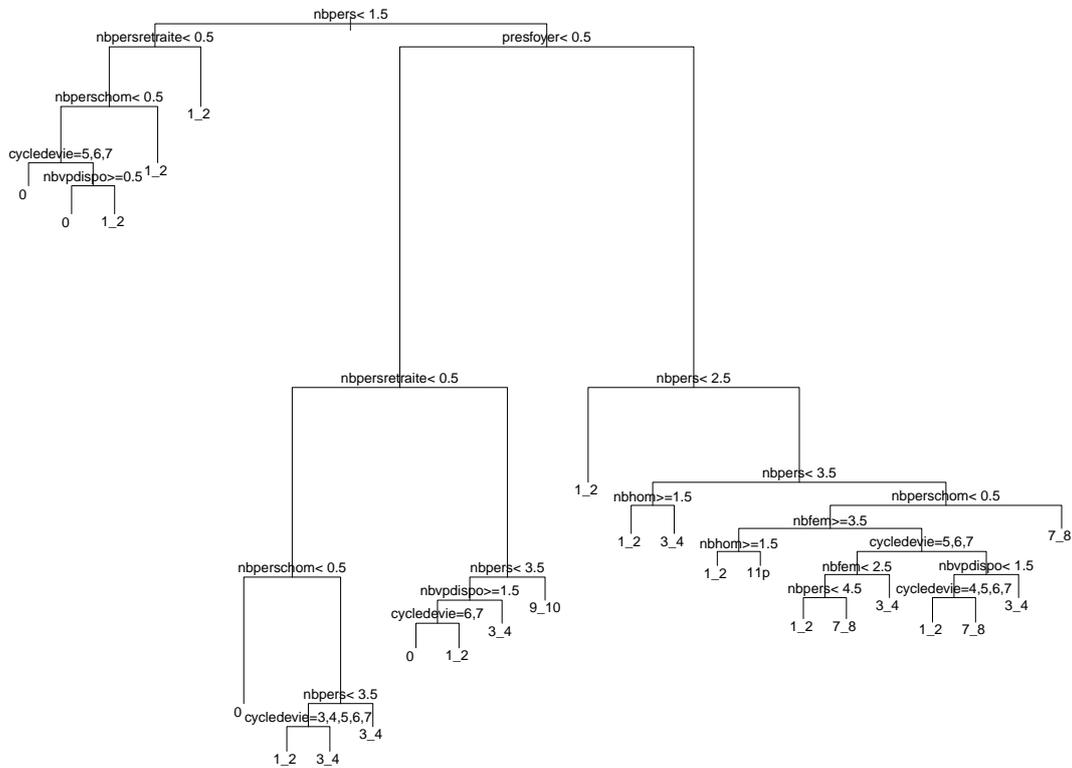
FIGURE D.20 – Arbre de classification pour les motifs autres que Travail et Enseignement construit à partir des données de l’EMD de 2006 sur le périmètre de l’aire urbaine lyonnaise (avec prise en compte des probabilités de tirage).



Source : traitement Cabrera Delgado à partir de l’EMD 2006.

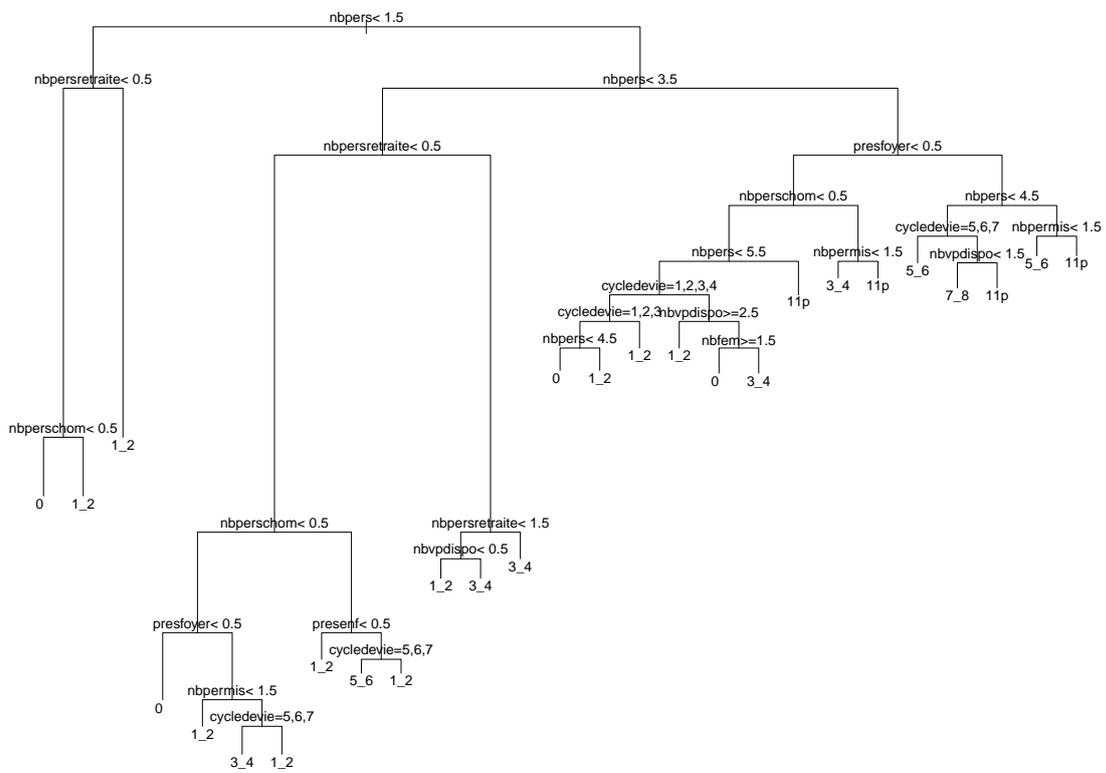
D.4.2.2 Sans prise en compte des probabilités de tirage

FIGURE D.21 – Arbre de classification pour les motifs autres que Travail et Enseignement construit à partir des données de l'EMD de 1985 (sans prise en compte des probabilités de tirage).



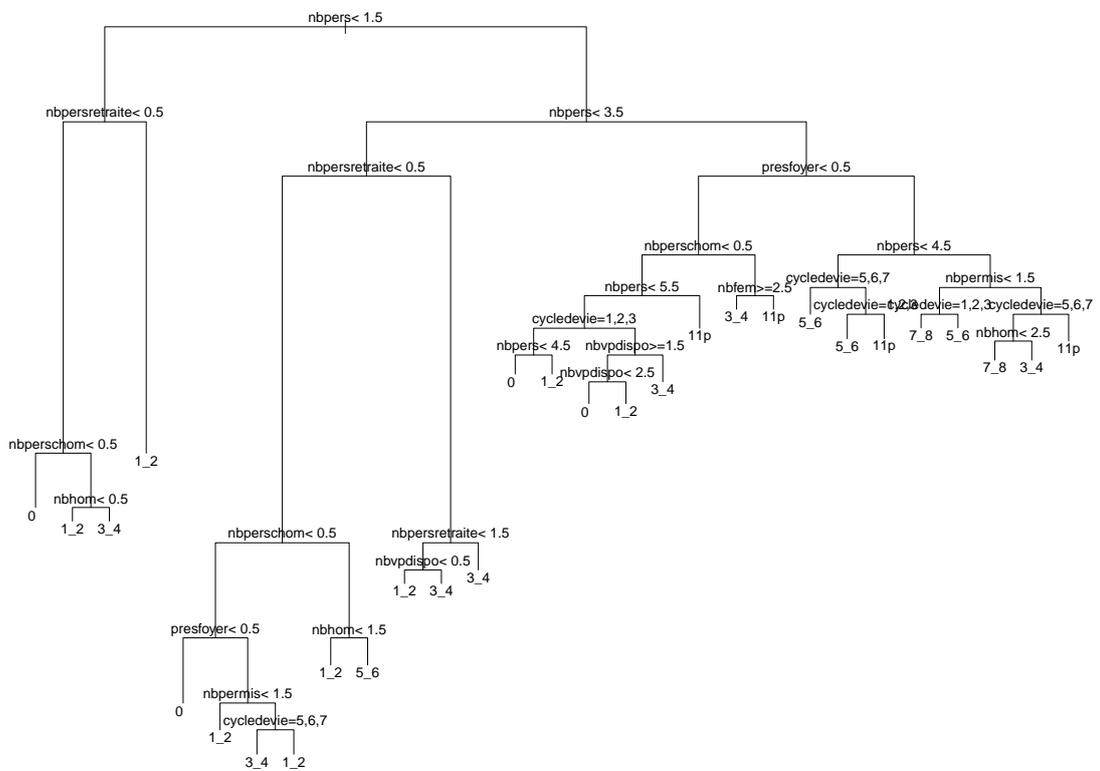
Source : traitement Cabrera Delgado à partir de l'EMD 1985.

FIGURE D.22 – Arbre de classification pour les motifs autres que Travail et Enseignement construit à partir des données de l'EMD de 1995 sur le périmètre de 1985 (sans prise en compte des probabilités de tirage).



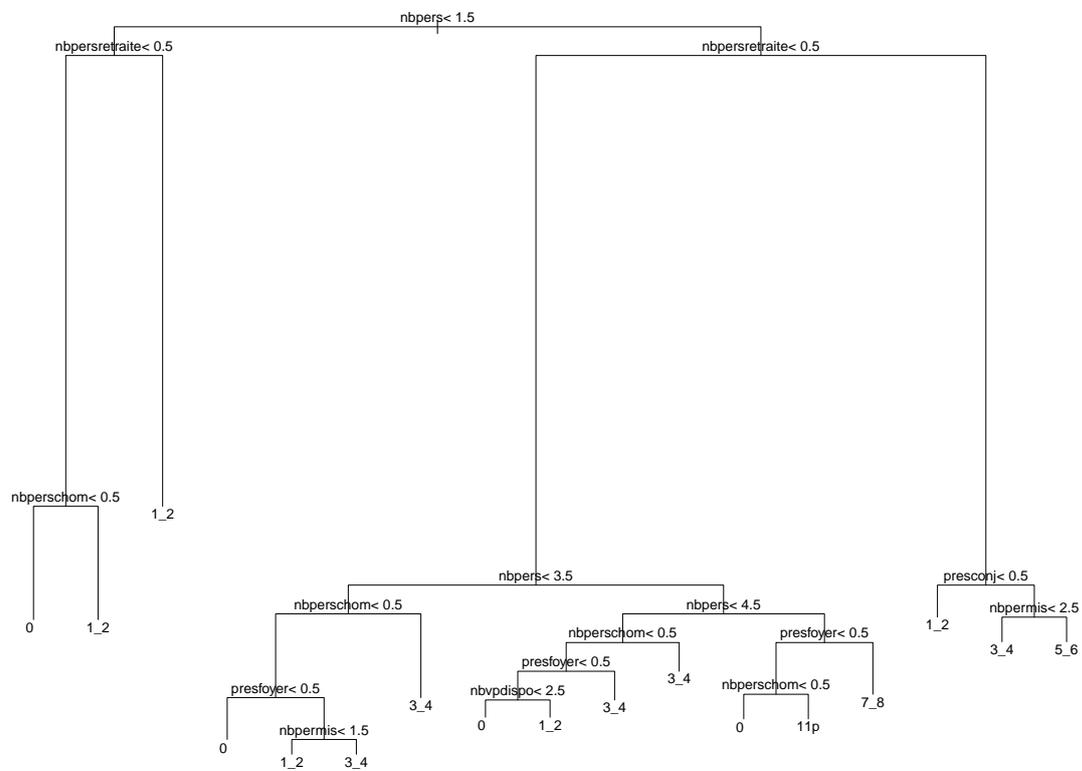
Source : traitement Cabrera Delgado à partir de l'EMD 1995.

FIGURE D.23 – Arbre de classification pour les motifs autres que Travail et Enseignement construit à partir des données de l’EMD de 1995 sur le périmètre de 1995 (sans prise en compte des probabilités de tirage).



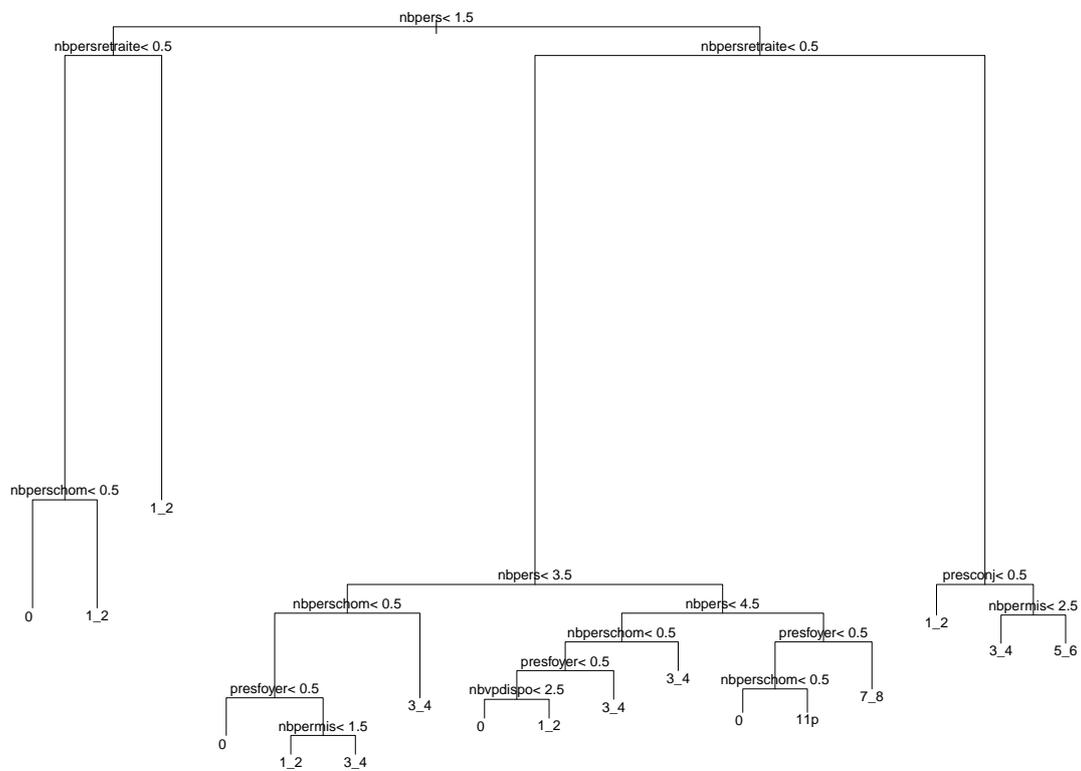
Source : traitement Cabrera Delgado à partir de l’EMD 1995.

FIGURE D.24 – Arbre de classification pour les motifs autres que Travail et Enseignement construit à partir des données de l'EMD de 2006 sur le périmètre de 1985 (sans prise en compte des probabilités de tirage).



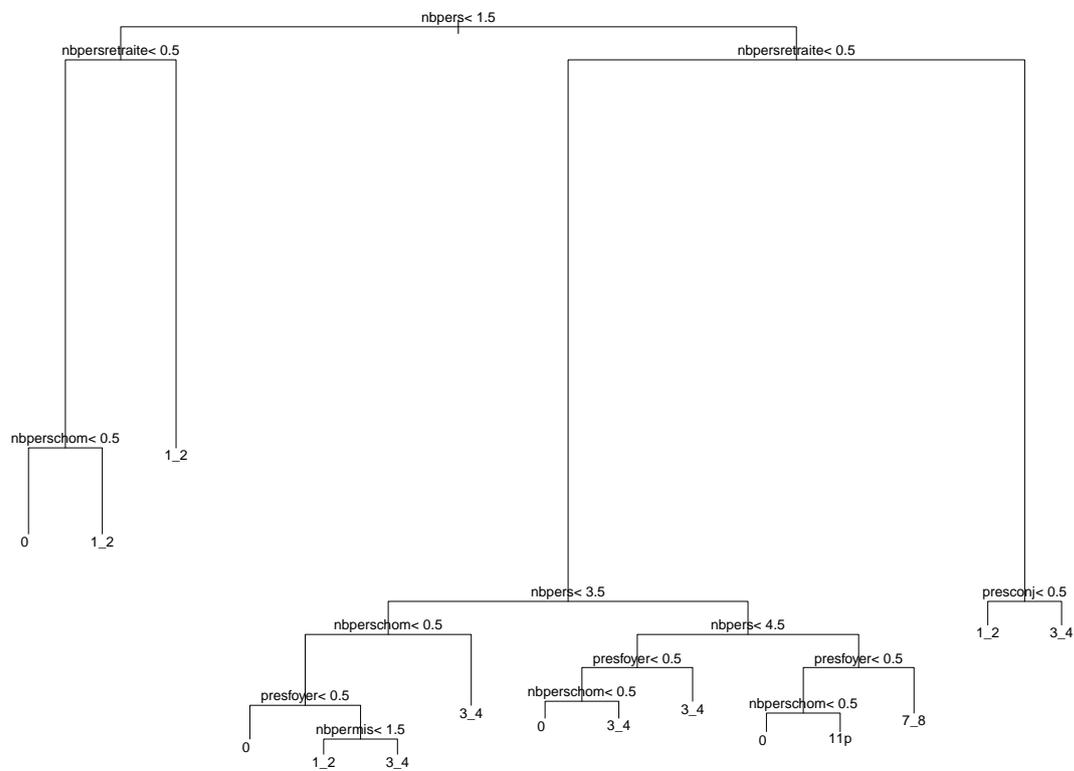
Source : traitement Cabrera Delgado à partir de l'EMD 2006.

FIGURE D.25 – Arbre de classification pour les motifs autres que Travail et Enseignement construit à partir des données de l’EMD de 2006 sur le périmètre de 1995 (sans prise en compte des probabilités de tirage).



Source : traitement Cabrera Delgado à partir de l’EMD 2006.

FIGURE D.26 – Arbre de classification pour les motifs autres que Travail et Enseignement construit à partir des données de l’EMD de 2006 sur le périmètre de l’aire urbaine lyonnaise (sans prise en compte des probabilités de tirage).



Source : traitement Cabrera Delgado à partir de l’EMD 2006.

---

## Liste des figures

1.1	Équilibre offre-demande de transport de Manheim. . . . .	17
1.2	Séquence de modélisation classique à quatre étapes . . . . .	19
3.1	Communes de l'agglomération lyonnaise incluses dans le périmètre des différentes EMD ainsi que de l'aire urbaine lyonnaise (AUL). . . . .	57
3.2	Périmètre et zonage retenus pour le projet SIMBAD : 777 IRIS de l'Aire urbaine lyonnaise en 1999. . . . .	66
3.3	Courbes débit-temps retenues dans SIMBAD en distinguant la partie « avant saturation » et la partie « après saturation ». . . . .	70
3.4	Courbes débit-temps associées aux différents types de liens (capture d'écran DAVISUM). . . . .	71
3.5	Récapitulatif de l'affectation des nœuds aux types de carrefours. . . . .	72
4.1	<i>Forest plot</i> du nombre de demi-sorties par ménage, pour le motif Travail, en fonction du nombre d'actifs employés (pour les différentes dates d'enquête et périmètres). . . . .	108
4.2	<i>Forest plot</i> du nombre de demi-sorties par ménage, pour le motif Travail, en fonction de véhicules à disposition du ménage (pour les différentes dates d'enquête et périmètres). . . . .	110
4.3	<i>Forest plot</i> du nombre de demi-sorties par ménage, pour le motif Travail, en fonction du stade dans le cycle de vie du ménage (pour les différentes dates d'enquête et périmètres). . . . .	111
4.4	<i>Forest plot</i> du nombre de demi-sorties par ménage, pour les motifs autres que le motif Travail ou Enseignement, en fonction du nombre de personnes du ménage (pour les différentes dates d'enquête et périmètres). . . . .	113
4.5	<i>Forest plot</i> du nombre de demi-sorties par ménage, pour les motifs autres que le motif Travail ou Enseignement, en fonction de véhicules à disposition du ménage (pour les différentes dates d'enquête et périmètres). . . . .	114
4.6	<i>Forest plot</i> du nombre de demi-sorties par ménage, pour les motifs autres que le motif Travail ou Enseignement, en fonction du stade dans le cycle de vie du ménage (pour les différentes dates d'enquête et périmètres). . . . .	116

5.1	Évolution de la distribution des distances à vol d’oiseau des demi-sorties, pour les motifs Travail et Loisir. . . . .	149
5.2	Capacité des matrices simulées, à partir du modèle à la Tanner, à reproduire la distribution des distances observées, pour le motif Travail. . . . .	158
5.3	Capacité des matrices simulées, à partir du modèle à la Tanner, à reproduire la distribution des distances observées, pour le motif Loisir. . . . .	159
6.1	Évolution des parts modales des demi-sorties des résidents du Grand Lyon, réalisées à l’intérieur de ce périmètre. Segmentation par motif. . . . .	168
6.2	Évolution des parts modales à l’intérieur du Grand Lyon en fonction de la distance « à vol d’oiseau » des demi-sorties. Segmentation par motif. . . . .	169
6.3	Modélisation et prévision de la part modale des TC à l’intérieur du Grand Lyon en 2006, pour le motif Travail, avec la spécification étendue. Segmentation selon la distance « à vol d’oiseau » des demi-sorties. . . . .	182
6.4	Modélisation et prévision de la part modale des TC à l’intérieur du Grand Lyon en 1985, pour le motif Travail, avec la spécification étendue. Segmentation selon la distance « à vol d’oiseau » des demi-sorties. . . . .	185
A.1	Carte du zonage commun pour l’analyse des modèles de génération et choix modal. . . . .	213
C.1	Capture d’écran du réseau TC de 1985 modélisé sous DAVISUM. . . . .	215
C.2	Capture d’écran du réseau TC de 1995 modélisé sous DAVISUM. . . . .	216
C.3	Capture d’écran du réseau TC de 2001 modélisé sous DAVISUM. . . . .	216
C.4	Capture d’écran du réseau TC de 2006 modélisé sous DAVISUM. . . . .	217
D.1	Relation entre le nombre de personnes ayant un emploi et le nombre de demi-sorties pour le motif Travail. . . . .	219
D.2	Relation entre le nombre de personnes de la zone et le nombre de demi-sorties pour les motifs autres que Travail et Enseignement. . . . .	220
D.3	Arbre de classification pour le motif Travail construit à partir des données de l’EMD de 1985 (avec prise en compte des probabilités de tirage). . . . .	223
D.4	Arbre de classification pour le motif Travail construit à partir des données de l’EMD de 1995 sur le périmètre de 1985 (avec prise en compte des probabilités de tirage). . . . .	224
D.5	Arbre de classification pour le motif Travail construit à partir des données de l’EMD de 1995 sur le périmètre de 1995 (avec prise en compte des probabilités de tirage). . . . .	225
D.6	Arbre de classification pour le motif Travail construit à partir des données de l’EMD de 2006 sur le périmètre de 1985 (avec prise en compte des probabilités de tirage). . . . .	226
D.7	Arbre de classification pour le motif Travail construit à partir des données de l’EMD de 2006 sur le périmètre de 1995 (avec prise en compte des probabilités de tirage). . . . .	227
D.8	Arbre de classification pour le motif Travail construit à partir des données de l’EMD de 2006 sur le périmètre de l’aire urbaine lyonnaise (avec prise en compte des probabilités de tirage). . . . .	228

D.9	Arbre de classification pour le motif Travail construit à partir des données de l'EMD de 1985 (sans prise en compte des probabilités de tirage). . . .	229
D.10	Arbre de classification pour le motif Travail construit à partir des données de l'EMD de 1995 sur le périmètre de 1985 (sans prise en compte des probabilités de tirage). . . . .	230
D.11	Arbre de classification pour le motif Travail construit à partir des données de l'EMD de 1995 sur le périmètre de 1995 (sans prise en compte des probabilités de tirage). . . . .	231
D.12	Arbre de classification pour le motif Travail construit à partir des données de l'EMD de 2006 sur le périmètre de 1985 (sans prise en compte des probabilités de tirage). . . . .	232
D.13	Arbre de classification pour le motif Travail construit à partir des données de l'EMD de 2006 sur le périmètre de 1995 (sans prise en compte des probabilités de tirage). . . . .	233
D.14	Arbre de classification pour le motif Travail construit à partir des données de l'EMD de 2006 sur le périmètre de l'aire urbaine lyonnaise (sans prise en compte des probabilités de tirage). . . . .	234
D.15	Arbre de classification pour les motifs autres que Travail et Enseignement construit à partir des données de l'EMD de 1985 (avec prise en compte des probabilités de tirage). . . . .	235
D.16	Arbre de classification pour les motifs autres que Travail et Enseignement construit à partir des données de l'EMD de 1995 sur le périmètre de 1985 (avec prise en compte des probabilités de tirage). . . . .	236
D.17	Arbre de classification pour les motifs autres que Travail et Enseignement construit à partir des données de l'EMD de 1995 sur le périmètre de 1995 (avec prise en compte des probabilités de tirage). . . . .	237
D.18	Arbre de classification pour les motifs autres que Travail et Enseignement construit à partir des données de l'EMD de 2006 sur le périmètre de 1985 (avec prise en compte des probabilités de tirage). . . . .	238
D.19	Arbre de classification pour les motifs autres que Travail et Enseignement construit à partir des données de l'EMD de 2006 sur le périmètre de 1995 (avec prise en compte des probabilités de tirage). . . . .	239
D.20	Arbre de classification pour les motifs autres que Travail et Enseignement construit à partir des données de l'EMD de 2006 sur le périmètre de l'aire urbaine lyonnaise (avec prise en compte des probabilités de tirage). . . .	240
D.21	Arbre de classification pour les motifs autres que Travail et Enseignement construit à partir des données de l'EMD de 1985 (sans prise en compte des probabilités de tirage). . . . .	241
D.22	Arbre de classification pour les motifs autres que Travail et Enseignement construit à partir des données de l'EMD de 1995 sur le périmètre de 1985 (sans prise en compte des probabilités de tirage). . . . .	242
D.23	Arbre de classification pour les motifs autres que Travail et Enseignement construit à partir des données de l'EMD de 1995 sur le périmètre de 1995 (sans prise en compte des probabilités de tirage). . . . .	243
D.24	Arbre de classification pour les motifs autres que Travail et Enseignement construit à partir des données de l'EMD de 2006 sur le périmètre de 1985 (sans prise en compte des probabilités de tirage). . . . .	244

D.25 Arbre de classification pour les motifs autres que Travail et Enseignement  
construit à partir des données de l'EMD de 2006 sur le périmètre de 1995  
(sans prise en compte des probabilités de tirage). . . . . 245

D.26 Arbre de classification pour les motifs autres que Travail et Enseignement  
construit à partir des données de l'EMD de 2006 sur le périmètre de l'aire  
urbaine lyonnaise (sans prise en compte des probabilités de tirage). . . . 246

---

## Liste des tables

3.1	Évolutions démographiques et socioéconomiques constatées dans l'agglomération lyonnaise à partir des données redressées des EMD, en prenant en compte les différents périmètres d'enquête (figure 3.1).	60
3.2	Typologie des liens de réseau	69
3.3	Pénalités aux mouvements tournants dans SIMBAD (en secondes).	72
3.4	Charges finales en entrée et en sortie de l'AUL	76
3.5	Répartition de la valeur de l'indicateur <i>GEH</i> en fonction du type de réseau (en %).	76
3.6	Matrices de demande VP	80
3.7	Temps spécifiques de correspondances entre systèmes de transports aux arrêts (en minutes).	85
3.8	Fonction de temps généralisé.	86
3.9	Résultats du calage du réseau de transports collectifs de référence	87
3.10	Nombre de lignes codifiées par année	89
4.1	Sorties réalisées pour 1 000 personnes suivant une nomenclature semi-détaillée en 30 classes.	95
4.2	Analyse de la stabilité temporelle du coefficient d'émission de demi-sorties par ménage, pour le motif Travail. Classification réalisée à partir de deux variables : nombre de personnes déclarant avoir un emploi et nombre de véhicules disponibles.	118
4.3	Analyse de la stabilité temporelle du coefficient d'émission de demi-sorties par ménage, pour le motif Travail. Classification réalisée à partir de deux variables : nombre de personnes déclarant avoir un emploi et cycle de vie.	120
4.4	Analyse de la stabilité temporelle du coefficient d'émission de demi-sorties par ménage, pour les motifs autres que Travail et Enseignement. Classification réalisée à partir de deux variables : nombre de personnes et nombre de véhicules disponibles.	122
4.5	Analyse de la stabilité temporelle du coefficient d'émission de demi-sorties par ménage, pour les motifs autres que Travail et Enseignement. Classification réalisée à partir de deux variables : nombre de personnes et cycle de vie.	125

4.6	Prévision de la mobilité de 2006 dans le périmètre de l'EMD 1985 avec les coefficients estimés sur ce même périmètre en 1985 et 1995. . . . .	128
4.7	Prévision de la mobilité de 2006 sur l'ensemble de l'aire urbaine avec les coefficients estimés avec l'ensemble de données de 1985 et 1995. . . . .	128
4.8	Résultats des calages des modèles de régression linéaire à l'échelle zonale pour le motif Travail . . . . .	132
4.9	Résultats des calages des modèles de régression linéaire à l'échelle zonale pour les motifs autres que Travail et Enseignement. . . . .	133
4.10	Capacité des modèles de régression à l'échelle zonale, calibrés en 1985 et 1995, à prévoir la mobilité observée en 2006. . . . .	134
4.11	Résultats de calage des modèles de génération pour le motif Travail à l'échelle des ménages (avec prise en compte explicite du plan de sondage) et test de stabilité temporelle des paramètres. . . . .	136
4.12	Résultats de calage des modèles de génération pour les motifs autres que Travail et Enseignement à l'échelle des ménages (avec prise en compte explicite du plan de sondage) et test de stabilité temporelle des paramètres. . . . .	139
4.13	Prévision de la mobilité de 2006 dans le périmètre de l'EMD 1985 avec les coefficients estimés sur ce même périmètre en 1985 et 1995, avec et sans prise en compte du plan de sondage . . . . .	142
4.14	Prévision de la mobilité de 2006 dans l'AUL avec les coefficients estimés en utilisant toutes les données disponibles en 1985 et 1995, avec et sans prise en compte du plan de sondage . . . . .	142
5.1	Distances moyennes à vol d'oiseau en km des demi-sorties, pour les motifs Travail et Loisirs. . . . .	148
5.2	Résultats des calages du modèle Tanner par année et périmètre pour les motifs Travail et Loisir (zonage IRIS2000). . . . .	155
5.3	Performance prospective du modèle de type Tanner avec les coefficients du calage de 1985/P1985 (zonage IRIS2000). . . . .	157
6.1	Résultats de calage des modèles de choix modal pour le motif Travail et test de stabilité temporelle des paramètres. . . . .	177
6.2	Tests de transférabilité globale des modèles calibrés en 1985 et 1995 pour le contexte d'application (données) de 2006. . . . .	181
6.3	Prévision de la part des TC en 2006 avec les coefficients estimés en 1985 et 1995, avec et sans prise en compte des coefficients de redressement. . . . .	181
6.4	Tests de transférabilité globale de la spécification étendue calibrée en 1985 pour le contexte d'application (données) de 1995. . . . .	183
6.5	Prévision de la part des TC en 1995 avec les coefficients estimés en 1985 et 1995, avec et sans prise en compte des coefficients de redressement. . . . .	184
D.1	Résultats de calage des modèles de génération pour le motif Travail à l'échelle des ménages (sans prise en compte explicite du plan de sondage) et test de stabilité temporelle des paramètres. . . . .	221
D.2	Résultats de calage des modèles de génération pour les motifs autres que Travail et Enseignement à l'échelle des ménages (sans prise en compte explicite du plan de sondage) et test de stabilité temporelle des paramètres. . . . .	222

---

# Table des matières

<b>Introduction</b>	<b>7</b>
<b>1 Objectifs et principes de la modélisation à quatre étapes</b>	<b>13</b>
1.1 Prévoir la demande de transport : l'approche des modèles économiques	13
1.1.1 Objectifs de la modélisation . . . . .	14
1.1.2 Cadrage économique de l'analyse de la demande de transport de personnes . . . . .	15
1.1.2.1 Interdépendance entre offre et demande de transport de personnes : le transport comme service . . . . .	15
1.1.2.2 Espaces et Activités : la demande de transports comme une demande « dérivée » . . . . .	16
1.1.3 Le modèle classique à quatre étapes . . . . .	18
1.1.3.1 Génération . . . . .	20
1.1.3.2 Distribution . . . . .	20
1.1.3.3 Répartition modale . . . . .	21
1.1.3.4 Affectation . . . . .	22
1.1.4 Utilité aléatoire : clivage entre une modélisation agrégée ou désagrégée . . . . .	23
1.2 Familles de modèles pour la spécification des quatre étapes . . . . .	25
1.2.1 Génération . . . . .	25
1.2.1.1 Classification croisée . . . . .	25
1.2.1.2 Régression linéaire . . . . .	26
1.2.1.3 Les arbres de segmentation (classification ou régression)	27
1.2.2 Distribution . . . . .	28
1.2.3 Répartition modale à partir d'une modélisation désagrégée . .	31
1.2.4 Affectation . . . . .	33
1.3 Synthèse et remarques conclusives . . . . .	33
<b>2 Transférabilité temporelle des modèles de demande de transport :   état des savoirs</b>	<b>35</b>
2.1 Validation des modèles prospectifs ou étude de leur transférabilité temporelle . . . . .	35
2.1.1 Vocabulaire de l'approche empirique . . . . .	35

2.1.2	La démarche de prévision : validation temporelle et étude de la transférabilité . . . . .	36
2.2	Évidence sur la transférabilité temporelle des modèles de prévision de la demande de transport . . . . .	38
2.2.1	Génération . . . . .	38
2.2.2	Distribution et modèles gravitaires . . . . .	42
2.2.3	Partage ou choix modal . . . . .	45
2.3	Conclusion . . . . .	48
<b>3</b>	<b>Terrain d'étude et données pour la modélisation</b>	<b>49</b>
3.1	L'agglomération lyonnaise : reflet des mutations dans les politiques d'aménagement des transports en France . . . . .	50
3.2	Les données relatives à la demande de transport . . . . .	52
3.2.1	Les enquêtes ménages déplacements . . . . .	52
3.2.1.1	Présentation générale des enquêtes ménages déplacements . . . . .	52
3.2.1.2	Comparabilité dans le temps des EMD lyonnaises . . . . .	53
3.2.1.2.1	Perturbations difficilement mesurables . . . . .	53
3.2.1.2.2	Homogénéité des données disponibles . . . . .	55
3.2.1.3	Calculs de précision . . . . .	58
3.2.2	Dynamiques urbaines constatées à partir des EMD . . . . .	59
3.3	Données relatives à l'offre de transport . . . . .	62
3.3.1	Codification du réseau routier : vers une codification automatique ? . . . . .	64
3.3.1.1	Méthodologie de la codification des réseaux routiers de référence . . . . .	64
3.3.1.1.1	Niveau d'agrégation . . . . .	64
	Principe . . . . .	64
	Choix du projet SIMBAD . . . . .	65
	Impact pour notre étude . . . . .	65
3.3.1.1.2	Représenter les infrastructures par des arcs et des nœuds . . . . .	66
	Une utilisation des bases de données existantes . . . . .	67
	Typologie des liens . . . . .	68
	Choix des courbes débit-temps et des capacités en heure de pointe . . . . .	68
	Typologie des nœuds et mouvements tournants aux carrefours . . . . .	70
3.3.1.1.3	Faire le lien entre zones et réseau . . . . .	72
3.3.1.1.4	Calage du réseau . . . . .	74
3.3.1.2	Codification de l'historique du réseau routier . . . . .	75
3.3.1.2.1	Principes de codification de l'historique des réseaux routiers . . . . .	75
	Les données disponibles . . . . .	76
	Les principes de codification . . . . .	77
	Matrices de demande . . . . .	79

3.3.2	Codification du réseau de transports collectifs : vers une codification automatique à partir des données publiques ? . . . . .	80
3.3.2.1	Réseau de référence : principes de codification retenus pour le projet SIMBAD . . . . .	80
3.3.2.1.1	Sources de données . . . . .	81
3.3.2.1.2	Codification des lignes de transports collectifs . . . . .	81
3.3.2.1.3	Codification automatique des connecteurs de centroïde . . . . .	82
3.3.2.1.4	Choix de la fonction de temps généralisé . . . . .	84
3.3.2.1.5	Calage du réseau de transports collectifs . . . . .	86
3.3.2.2	Codification de l'historique du réseau TC . . . . .	86
3.3.2.2.1	Principes de codification de l'historique des réseaux TC . . . . .	87
	Les données disponibles . . . . .	87
	Codification dans DAVISUM . . . . .	87
3.3.2.2.2	Détermination des temps généralisés . . . . .	90
3.4	Synthèse et remarques conclusives . . . . .	90
<b>4</b>	<b>Génération</b>	<b>93</b>
4.1	Une méthodologie pour étudier les évolutions temporelles . . . . .	93
4.1.1	Choix de la « sortie » comme unité de mesure de la mobilité . . . . .	94
4.1.1.1	Avantages d'une analyse en termes de nombre de sorties . . . . .	94
4.1.1.2	Évolution de la mobilité en termes de « sorties » . . . . .	94
4.1.1.2.1	La complexité des objets « sorties » . . . . .	94
4.1.1.2.2	Simplification de la nomenclature et demi-sorties . . . . .	97
4.1.2	Modèles de génération . . . . .	98
4.1.2.1	Choix du ménage comme unité de prise de décision . . . . .	98
4.1.2.2	Choix des modèles à étudier . . . . .	99
4.1.2.2.1	Les modèles par classification croisée . . . . .	100
4.1.2.2.2	Les modèles de régression à l'échelle zonale . . . . .	101
4.1.2.2.3	Les modèles de régression à l'échelle des ménages . . . . .	103
4.1.2.2.4	Les modèles par arbres de segmentation à l'échelle des ménages . . . . .	104
4.1.2.3	Évaluer la stabilité temporelle des modèles . . . . .	105
4.1.2.4	Indicateurs de qualité des calages et prévisions . . . . .	106
4.2	Étude de la mobilité moyenne par segment de population : est-ce que les émissions moyennes obtenues dans les modèles de classification croisée sont transférables dans le temps ? . . . . .	107
4.2.1	Stabilité ou instabilité des taux d'émission dans les classifications croisées ? . . . . .	107
4.2.1.1	Une seule variable explicative . . . . .	107
4.2.1.1.1	Le motif Travail . . . . .	108
4.2.1.1.2	Les motifs autres que Travail et Enseignement . . . . .	112
4.2.1.2	Deux variables explicatives . . . . .	117
4.2.1.2.1	Le motif Travail . . . . .	117

4.2.1.2.2	Les motifs autres que Travail et Enseignement	121
4.2.2	Les modèles de classification croisée calibrés en 1985 et 1995 permettent-ils de prédire convenablement les comportements de mobilité de 2006 ?	126
4.3	Le modèle de génération par régression zonale	130
4.3.1	Stabilité des coefficients	130
4.3.1.1	Spécification du modèle	130
4.3.1.1.1	Motif travail	130
4.3.1.1.2	Autres motifs que Travail et Enseignement	131
4.3.1.2	Résultats de calage	131
4.3.1.2.1	Motif travail	131
4.3.1.2.2	Autres motifs que Travail et Enseignement	132
4.3.2	Capacité prédictive des modèles par régression linéaire à l'échelle zonale	132
4.4	Le modèle de génération par régression à l'échelle des ménages	135
4.4.1	Spécification	135
4.4.2	Résultats des calages	135
4.4.2.1	Motif Travail	135
4.4.2.2	Motifs autres que Travail et Enseignement	137
4.4.3	Capacité prédictive des modèles par régression à l'échelle du ménage	140
4.5	Classification automatique : est-ce optimal de conserver les mêmes segments de population en prévision ?	143
4.5.1	Motif Travail	143
4.5.2	Motifs autres que Travail et Enseignement	144
4.6	Remarques conclusives	144
<b>5</b>	<b>Distribution</b>	<b>147</b>
5.1	Méthodologie	147
5.1.1	Évaluation de l'allongement des distances observées	148
5.1.2	Spécifications du modèle gravitaire segmentées par motif de déplacement	150
5.1.3	Calibration et indicateurs	151
5.1.3.1	Estimation des paramètres par la méthode du maximum de vraisemblance	151
5.1.3.2	Évaluation de la qualité des calages	152
5.1.3.3	Évaluation de la transférabilité des modèles	153
5.2	Résultats	154
5.2.1	Résultats de calage	154
5.2.2	Performance en prévision	155
5.3	Conclusion	160
<b>6</b>	<b>Choix modal</b>	<b>163</b>
6.1	Méthodologie pour comprendre l'évolution du choix du mode	164
6.1.1	Définition du mode de déplacement au sein d'une sortie	164
6.1.2	Un regain de parts de marché pour les TC	166
6.1.3	Spécification des modèles	170

6.1.3.1	Principes généraux . . . . .	170
6.1.3.2	Définition d'un ensemble de choix et d'une population de référence . . . . .	171
6.1.3.3	Deux spécifications pour les choix entre TC et VP . . . . .	172
6.1.4	Évaluer la stabilité des paramètres et la qualité des calages . . . . .	174
6.1.4.1	Tester l'égalité des paramètres . . . . .	174
6.1.4.2	Évaluer la capacité prédictive des modèles . . . . .	174
6.2	L'hypothèse de stabilité des paramètres est-elle acceptable dans le cas des modèles de choix modal ? . . . . .	176
6.2.1	Résultats des estimations . . . . .	176
6.2.1.1	Qualité des calages . . . . .	176
6.2.1.2	Évolution temporelle des paramètres . . . . .	178
6.2.2	Aurait-on pu prévoir les parts modales en 2006 avec les modèles calibrés en 1985 et 1995 ? . . . . .	180
6.2.3	Peut-on parler d'un effet de rupture entre 1995 et 2006 ? . . . . .	182
6.3	Conclusion . . . . .	185
<b>Conclusion</b>		<b>187</b>
<b>Bibliographie</b>		<b>191</b>
 <b>Annexes</b>		
<b>A</b>	<b>Compléments méthodologiques</b>	<b>209</b>
A.1	Dérivation du modèle gravitaire selon l'approche classique dite de maximisation de l'entropie . . . . .	209
A.2	Zonage agrégé commun pour la génération et le choix modal . . . . .	213
<b>B</b>	<b>Définition de la variable « cycle de vie »</b>	<b>214</b>
<b>C</b>	<b>Réseau TC modélisé</b>	<b>215</b>
<b>D</b>	<b>Compléments sur les estimations des modèles de génération</b>	<b>218</b>
D.1	Modèle de génération du projet SIMBAD . . . . .	218
D.2	Représentation graphique des données pour la régression zonale . . . . .	219
D.3	Résultats de la régression au niveau des ménages sans la prise en compte du plan de sondage . . . . .	221
D.4	Arbres de classification . . . . .	223
D.4.1	Motif Travail . . . . .	223
D.4.1.1	Avec prise en compte des probabilités de tirage . . . . .	223
D.4.1.2	Sans prise en compte des probabilités de tirage . . . . .	229
D.4.2	Autres motifs que Travail et Enseignement . . . . .	235
D.4.2.1	Avec prise en compte des probabilités de tirage . . . . .	235
D.4.2.2	Sans prise en compte des probabilités de tirage . . . . .	241
<b>Liste des figures</b>		<b>247</b>

<b>Liste des tables</b>	<b>251</b>
<b>Table des matières</b>	<b>253</b>