



HAL
open science

Manifestations spatiales de la congestion et localisation des emplois et des ménages

Vincent Breteau

► **To cite this version:**

Vincent Breteau. Manifestations spatiales de la congestion et localisation des emplois et des ménages. Architecture, aménagement de l'espace. Université Paris-Est, 2011. Français. NNT : 2011PEST1134 . pastel-00690398

HAL Id: pastel-00690398

<https://pastel.hal.science/pastel-00690398v1>

Submitted on 23 Apr 2012

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 PARIS-EST
École Doctorale Ville, Transports et Territoires
Laboratoire Ville Mobilité Transport

THÈSE DE DOCTORAT
Spécialité Économie des transports

Manifestations spatiales de la congestion et localisation des emplois et des ménages

Présentée et soutenue publiquement

le 26/09/2011

par Vincent BRETEAU

Sous la direction de Fabien LEURENT

Composition du jury :

Jean DELONS, examinateur

Florence GOFFETTE-NAGOT, examinatrice

Jean LATERRASSE, examinateur,

Émile QUINET, rapporteur,

Charles RAUX, rapporteur,

Isabelle THOMAS, présidente.

Manifestations spatiales de la congestion et localisation des emplois et des ménages

Thèse de doctorat en économie des transports
présentée par Vincent BRETEAU
sous la direction de Fabien LEURENT

UNIVERSITÉ PARIS-EST
École Doctorale Ville, Transports et Territoires
Laboratoire Ville Mobilité Transport

Remerciements

Bien plus qu'un aboutissement, la thèse est un processus, une construction. Pleine de surprises, jalonnée de moments de doute, de déception, d'envie et de satisfaction, elle s'est nourrie, durant les trois années que j'y ai consacrées, de l'ensemble des rencontres que j'ai pu faire, des soutiens et des encouragements dont j'ai pu bénéficier. Alors que je m'appête à rendre ce travail, je me dois de remercier tous ceux sans qui il n'aurait pas vu le jour.

Mes premiers remerciements vont à Fabien Leurent qui, par ses nombreux conseils, sa disponibilité et sa présence, a su m'orienter tout au long de ce travail et m'a permis de mener à bien cette thèse. Je remercie les membres du jury qui ont consenti à évaluer ce travail : Émile Quinet et Charles Raux tout d'abord, qui ont accepté la lourde tâche d'en être les rapporteurs ; Florence Goffette-Nagot, Isabelle Thomas et Jean Delons ensuite, qui ont bien voulu participer à ce jury.

Je remercie les arbitres et participants des conférences auxquelles j'ai participé et qui m'ont permis, peu à peu, par leurs questions, remarques et conseils d'affiner mes interprétations et d'améliorer la présentation de mes résultats. Je suis en particulier redevable à Masahi Fujita, Jacques-François Thisse et Charles Raux pour leurs conseils et encouragements.

J'adresse également mes remerciements à l'ensemble des membres du LVMT, et tout particulièrement à François Combes, Nicolas Wagner, Frédéric Meunier, Nicolas Coulombel, Olivier Bonin et Thierno Aw. Nos nombreux échanges, toujours riches d'enseignements, m'ont souvent permis de retrouver une confiance en moi parfois en berne. Je remercie également Vincent Aguiléra, dont la relecture attentive de ce mémoire a été précieuse.

Je remercie Jean-Bernard Kovarik, Béatrice Adoléhouné et l'ensemble des membres du groupe de travail *Réactualisation des coefficients Hautreux* pour leur accueil et l'intérêt qu'ils ont porté aux travaux que nous leur avons présentés. Pour m'avoir chaleureusement accueilli au sein du Commissariat Général au Développement Durable depuis ma prise de fonction en septembre 2010, je tiens à remercier Jean-Jacques Becker, Laurence Demeulenaere, Olivier Teissier, David Meunier et l'ensemble de la sous-direction Mobilité et Aménagement. J'ai apprécié leur soutien dans cette délicate période de transition.

Enfin, je tiens à remercier mes amis et les membres de ma troupe de théâtre pour leurs encouragements pendant toute cette période, et tout particulièrement Antoine dont l'aide précieuse, pendant la difficile phase de fin de rédaction, a été essentielle. Je remercie également ma famille et celle de ma conjointe. J'ai une pensée toute particulière pour mon père dont l'absence, toujours douloureuse, se fait encore plus sentir aujourd'hui. Je lui dédie cette thèse.

Mes ultimes remerciements vont à Caroline, qui a partagé mes hauts et mes bas et dont la présence sans faille à mes côtés, les encouragements constants, la patience, la compréhension et la bienveillance m'ont permis de mener à bien ce travail et de m'épanouir autant qu'il est possible.

Table des matières

| | |
|-----------------------------------------------------------------------|------------|
| Introduction générale | 11 |
| | |
| I. Structure urbaine et congestion | 19 |
| | |
| 1. Agglomération et structure urbaine | 21 |
| 1.1. Introduction | 21 |
| 1.2. Une agglomération aux multiples sources | 23 |
| 1.3. La « tyrannie » de l'espace | 39 |
| 1.4. Le modèle urbain monocentrique | 44 |
| 1.5. Conclusion | 56 |
| | |
| 2. Les formes de la congestion | 59 |
| 2.1. Introduction | 59 |
| 2.2. Formes psychologiques de la congestion | 61 |
| 2.3. Formes physiques de la congestion | 69 |
| 2.4. Les coûts économiques de la congestion | 86 |
| 2.5. Conclusion | 97 |
| | |
| II. Les manifestations spatiales de la congestion du transport | 99 |
| | |
| 3. Méthodologie d'analyse de la congestion | 101 |
| 3.1. Introduction | 101 |
| 3.2. La représentation de l'ingénieur | 103 |
| 3.3. Trois perspectives d'évaluation de la congestion | 109 |
| 3.4. Évaluation locale de la congestion | 114 |
| 3.5. Appréhender la variabilité spatiale de la congestion | 117 |
| 3.6. Les indicateurs de congestion : du local au global | 125 |
| 3.7. Conclusion | 127 |
| 3.A. Annexes | 129 |

| | |
|-----------------------------------------------------------------------------|------------|
| 4. Analyse de la congestion routière en Île-de-France | 131 |
| 4.1. Introduction | 131 |
| 4.2. Présentation du contexte francilien | 132 |
| 4.3. La plateforme de simulation employée | 136 |
| 4.4. Les états simulés du trafic | 138 |
| 4.5. Diagnostic statistique de la congestion | 142 |
| 4.6. Diagnostic géographique de la congestion | 153 |
| 4.7. Diagnostic économique agrégé | 163 |
| 4.8. Conclusion | 168 |
| 4.A. Annexes | 170 |
| 5. Coût social marginal et aide à la décision | 175 |
| 5.1. Introduction | 175 |
| 5.2. Concurrence entre modes de transport | 178 |
| 5.3. L'évaluation des gains de décongestion routière | 183 |
| 5.4. Méthodologies d'évaluation des gains de temps | 188 |
| 5.5. Comparaison des méthodes | 196 |
| 5.6. Conclusion | 206 |
| 5.A. Détail des procédures d'évaluation des gains de décongestion | 208 |
| III. Localisation des ménages et des emplois et usage des transports | 211 |
| 6. Modèle d'équilibre urbain | 213 |
| 6.1. Introduction | 213 |
| 6.2. Décentralisation de l'emploi et modélisation | 215 |
| 6.3. Un modèle pour étudier la structure urbaine | 218 |
| 6.4. Équilibre du modèle | 222 |
| 6.5. Propriétés générales de l'équilibre | 231 |
| 6.6. Conclusion | 234 |
| 6.A. Annexes mathématiques | 236 |
| 7. Résolution analytique et sensibilité du modèle | 243 |
| 7.1. Introduction | 243 |
| 7.2. Présentation du cadre et résolution | 244 |
| 7.3. Analyses de sensibilité à l'équilibre urbain | 248 |
| 7.4. Conclusion | 259 |
| 7.A. Hypothèses et valeurs numériques choisies | 261 |
| 7.B. Annexes mathématiques | 262 |

| | |
|---------------------------------------------------------------------------|------------|
| 8. Étalement urbain et transports | 267 |
| 8.1. Introduction | 267 |
| 8.2. Étalement urbain et transport : un sujet à débat | 268 |
| 8.3. L'approche retenue et le modèle | 274 |
| 8.4. L'étalement des emplois, moteur de l'étalement urbain ? | 278 |
| 8.5. Un étalement favorisant un usage raisonné des transports ? | 286 |
| 8.6. Conclusion | 307 |
| 8.A. Annexes | 308 |
| | |
| Conclusion générale | 311 |
| | |
| Bibliographie | 317 |
| | |
| A. Appendices | 339 |
| A.1. Cartes complémentaires de congestion du réseau | 339 |
| A.2. Cartes globales de l'Île-de-France | 344 |
| A.3. Cartes complémentaires des indicateurs zonaux | 350 |

Introduction générale

Cette thèse s'inscrit dans le contexte urbain actuel, caractérisé à la fois par une croissance urbaine continue quoique ralentie dans les pays industrialisés et par l'acuité de certains phénomènes urbains tels que l'étalement urbain et la dépendance automobile. La ville en tant qu'objet d'étude est un système complexe et polymorphe, qui se laisse difficilement appréhender dans sa globalité. Les approches pour l'aborder ainsi que les disciplines qui s'y intéressent sont multiples, allant de la géographie à la sociologie, en passant par l'histoire et l'économie. Nous avons choisi dans cette thèse l'approche économique pour aborder les questions d'adéquation entre l'offre du système de transport et la demande émergeant des agents économiques.

Plus spécifiquement, ce travail de thèse vise à mieux comprendre les liens entre congestion des systèmes de transport et localisation des ménages et des emplois. Les aspects spatiaux de ces liens y sont tout particulièrement soulignés et étudiés. Ainsi, un premier objectif de ce travail est de dégager certaines des logiques spatiales de la congestion des transports. Nous nous appuierons, pour ce faire, sur le cas particulier de l'Île-de-France. Un second objectif est d'apporter une contribution théorique à l'analyse des conséquences de la localisation des emplois sur les choix de localisation des ménages et leur usage des transports.

L'émergence de la congestion urbaine

Le phénomène urbain est aujourd'hui majeur : si le nombre de citadins ne représentait que 29,2 % de la population mondiale en 1950, la population urbaine forme aujourd'hui 50 % d'une population mondiale qui a presque triplé et représentera en 2030 environ 60 % d'une population encore augmentée de plus d'un milliard d'individus (United Nations, 2010). La tendance est encore plus marquée pour les régions moins développées du monde, où le taux d'urbanisation est passé de 18 % en 1950, à près de 45 % aujourd'hui, avec une prévision de 56 % d'ici 2030.

Ce mouvement massif d'urbanisation s'accompagne tout naturellement d'une augmentation spectaculaire du nombre de villes dépassant les 10 millions d'habitants : alors qu'il n'y en avait que deux dans le monde jusqu'en 1970, on en

compte aujourd’hui vingt, et elles seront probablement près de trente en 2030. On commence même à parler de « métapoles » pour les villes de plus de 20 millions d’habitants (Huriot et Bourdeau-Lepage, 2009) : une seule ville dépassait ce seuil en 2007, Tokyo, mais huit nouvelles agglomérations devraient la rejoindre d’ici 2020, la plupart situées dans des pays en développement.

En Europe comme en Amérique du Nord, la croissance rapide de la population urbaine depuis les années 1950, associée à l’augmentation tout aussi rapide du taux de motorisation des ménages et à la périurbanisation, ont conduit à une forte hausse du besoin de transport de passagers et de fret dans les zones urbaines (Mills et Hamilton, 1994). De plus, le partage modal, entre transport public et voiture particulière, déjà favorable à cette dernière, a continué à évoluer dans le même sens. Ainsi, en 2001, la distance totale parcourue en Île-de-France par les usagers en voiture particulière atteignait presque le double de celle parcourue en transport collectif¹. La Table 0.1 présente l’évolution des distances parcourues chaque jour en transport public et en voiture particulière entre 1976 et 2001 (année de la dernière Enquête Générale Transport) en Île-de-France.

| Mode | 1976 | 1983 | 1991 | 1997 | 2001 |
|----------------------|------|------|------|------|------|
| Transport public | 47,0 | 50,9 | 59,3 | 59,4 | 59,3 |
| Voiture particulière | 52,9 | 66,1 | 83,3 | 99,5 | 98,9 |

Source : DREIF (2004a)

TABLE 0.1.: Evolution des distances parcourues par mode entre 1976 et 2001 (millions de kilomètres).

Cette hausse très importante des trafics, sur l’ensemble des réseaux urbains, et tout particulièrement sur les réseaux routiers, est à l’origine de problèmes aigus de congestion. La croissance, en termes de population, explique une partie de l’augmentation des trafics, mais l’extension spatiale des villes et la modification des localisations des ménages et des activités, est également à mettre au rang des causes d’un tel phénomène. Ce sont de ces problèmes de congestion, dans l’ensemble de ses dimensions, dont nous allons traiter dans cette thèse, car si la ville est notre objet d’étude, c’est en réalité surtout la ville congestionnée qui nous intéresse.

1. Il est à noter que la part d’utilisation de la voiture est moins élevée pour les déplacements domicile-travail, mais atteint néanmoins 55 % des déplacements mécanisés, contre 40 % pour les transports collectifs (DREIF, 2004a).

Enjeux scientifiques et méthodologiques

Nous partons, dans ce travail, de l'hypothèse que le développement urbain, ainsi que la forme que prend ce développement, ont une influence sur les coûts urbains subis par les usagers, en particulier les ménages. Ainsi, les coûts liés au transport dépendent de la forme urbaine, car celle-ci modifie l'usage que les agents font des transports, et elle a donc des répercussions en termes de congestion, qui représente une part importante des coûts urbains à l'heure actuelle. C'est pourquoi la compréhension de ses déterminants et de ses manifestations est essentielle pour en limiter l'ampleur et les effets sur le développement urbain. Or, les aspects liés à la distribution spatiale de la congestion sont, aujourd'hui encore, relativement peu explorés et restent, comme nos travaux commandités par le Ministère en charge des Transports l'ont mis en évidence, du domaine de la recherche et peinent à franchir les portes des acteurs opérationnels.

Plusieurs éléments permettent de comprendre les difficultés scientifiques et méthodologiques associées à ces constats. Tout d'abord, les formes diverses de la congestion et les mécanismes variés qui en sont à l'origine aboutissent à une complexité du phénomène. Cela impose une analyse de nature systémique qui doit permettre, entre autres, de maîtriser les liens entre ces formes et ces mécanismes, afin de rendre le réel plus compréhensible. Par ailleurs, l'analyse spatiale classique du phénomène de congestion des transports a ses limites. En effet, l'observation, sur les réseaux, de la congestion ne permet généralement pas de dégager des logiques spatiales claires : les trafics ne sont que la traduction, sur les réseaux de transport, des choix d'itinéraire des usagers, et dépendent donc fortement des origines et destinations des déplacements. Enfin, les interactions entre localisation des emplois et des ménages, et congestion restent délicates à modéliser de manière théorique, pour des raisons à la fois conceptuelles et techniques.

Les enjeux liés à ce travail découlent en partie de ces difficultés. Ainsi, d'un point de vue scientifique, il s'agit d'apporter des éléments de compréhension des interactions à double sens entre congestion et localisation des emplois et des ménages. Plus précisément, la mise en relation d'une forme de développement urbain, à travers ses modalités d'usage du sol, avec les coûts de transport associés, parmi lesquels la congestion représente une part importante, reste une question de recherche ouverte.

Par ailleurs, en termes de politiques de transport et d'aménagement se pose tout d'abord la question de l'adéquation entre l'offre de transport et la demande, émanant en particulier des modalités d'usage du sol. Ensuite, se pose la question de l'orientation du développement urbain vers un usage plus raisonné des transports, et tout particulièrement de la voiture particulière. Enfin, distribution spatiale des activités et manifestations spatiales de la congestion soulèvent la question de l'équité

spatiale de ces phénomènes et des éventuelles politiques correctrices.

Démarche de recherche

Problématique

La compréhension des interactions entre distribution spatiale de la congestion et localisation des activités, dont nous venons de montrer les enjeux, est au cœur de ce travail. Ces interactions complexes proviennent en grande partie du caractère exclusif de l'espace en tant que bien économique : les différents usages permis par l'espace — résidentiels, professionnels, de transport, ou autres — sont exclusifs les uns des autres, et ce à différents niveaux et à différentes échelles, notamment temporelles. Ce caractère exclusif est à l'origine de rivalités entre les usagers ou entre les établissements. Cette rivalité se traduit, dans le cas du transport, par une forme relativement commune de congestion : chaque usager qui emprunte une voie engendre des coûts (en particulier temporels) pour tous les autres usagers de la voie. Dans le cas de l'occupation du sol pour y établir une activité (logement, entreprise), l'échelle temporelle est beaucoup plus longue. La rivalité se traduit alors par une obligation, pour certains usagers, de se reporter ailleurs, lorsque la localisation souhaitée est déjà occupée. Or, localisation et usage des transports sont fortement liés, si bien que les deux phénomènes de congestion que nous venons d'évoquer le sont également.

Notre travail s'inscrit donc dans une problématique large de compréhension de la structure spatiale des contraintes de capacité et de leurs effets, en transport d'une part, et pour l'occupation du sol d'autre part.

Objectifs de la thèse

Parmi l'ensemble des questionnements que nous venons de soulever, seuls certains d'entre eux trouveront des éléments de réponse dans ce travail. Plus précisément, nous nous sommes attaché à analyser en détail les questions de capacité, en transport d'une part, d'occupation du sol d'autre part. Concernant les contraintes de capacité des systèmes de transport, notre objectif est d'élaborer une méthode d'analyse et d'évaluation spatiale de la congestion, applicable à l'ensemble d'un territoire, quel qu'il soit. Cet objectif s'accompagne d'une volonté de tester une telle méthode sur un territoire, celui de l'Île-de-France en l'occurrence. Concernant l'occupation du sol, l'objectif est de mettre en évidence certaines formes moins apparentes de congestion, à travers un modèle analytique simplifié mettant en relation les localisations respectives des emplois et des résidences. De manière à prendre en compte de façon plus raisonnée les questions de rivalités pour le sol, nous avons adopté

le principe selon lequel chaque ménage veut accéder à son lieu d'emploi, et non à l'ensemble des lieux d'emplois².

Approches retenues

Nous avons choisi de nous appuyer sur deux approches qui, selon nous, sont complémentaires. Tout d'abord, nous avons adopté une approche empirique, que nous pouvons qualifier d'*ascendante*. Fondée sur la modélisation des choix d'itinéraire des ménages dans leurs déplacements domicile-travail, elle vise à mettre en évidence la congestion qui en résulte. Cette approche s'inscrit dans un double cadre : celui de la modélisation spatialisée du transport d'une part, et celui de l'économie des transports d'autre part. Plus précisément, en s'appuyant sur la modélisation des itinéraires, nous avons cherché à mettre en relation divers indicateurs économiques de la congestion avec les zones d'origine et de destination des usagers. Autrement dit, nous avons choisi un mode d'analyse en quelque sorte dual du mode classique qui repose directement sur les trafics sur les arcs des réseaux.

Ensuite, nous avons adopté une approche théorique, que nous pouvons qualifier de *descendante*, fondée pour sa part sur la modélisation des choix de localisation résidentielle des ménages. Cette approche s'inscrit dans le cadre de l'économie urbaine théorique, et repose en particulier sur une modélisation de la ville. Elle nous a plus spécifiquement permis de mettre en relation les localisations des emplois et leur évolution avec les localisations des ménages et leur évolution. Nous avons en particulier cherché à analyser les conséquences de ces choix de localisation sur l'usage des transports pour les déplacements domicile-travail qui en résulte.

Plan de la thèse

Cette thèse est organisée en huit chapitres, répartis en trois parties. Celles-ci sont consacrées à une revue bibliographique des thèmes majeurs de la thèse, à une étude des manifestations spatiales de la congestion et à une analyse monocentrique des conséquences de la localisation des activités sur les déplacements domicile-travail.

Première partie – Structure urbaine et congestion

La thèse débute par une analyse bibliographique sur les questions de croissance et de structure urbaine, suivie d'une analyse, de nature systémique, de la congestion des transports.

2. Ce principe est à rapprocher de celui qui sous-entend certains modèles de distribution des déplacements (Leurent, 1999, modèle *Accessibility to Vacant Activities*), ou certaines recherches sur les localisations respectives des ménages et des emplois (Korsu et Massot, 2006).

Le chapitre 1, *Du processus d'agglomération à la structure urbaine : le rôle de l'espace*, vise à remettre en perspective les nombreux travaux, théoriques et empiriques, sur le processus d'agglomération, afin de mettre en évidence comment ces processus font émerger une structure urbaine. En particulier, nous avons cherché à montrer le rôle ambivalent des transports dans ce processus de structuration, en tant que médiateur des économies d'agglomération d'une part, et comme coût du développement urbain d'autre part. Le cadre du modèle monocentrique permet de mettre en scène l'ensemble de ces forces pour en faire émerger une structure. La présentation de ce modèle permet également de mettre en évidence ses limites, dont nous chercherons, pour certaines, à nous affranchir dans la troisième partie de ce mémoire.

Le chapitre 2, *Les formes de la congestion : de la multiplicité à la complexité*, propose une analyse de la congestion à travers trois prismes : psychologique, physique et économique. L'objectif, au-delà d'une meilleure compréhension de l'enchevêtrement des liens existant entre ces formes, est de disposer d'une base permettant la définition d'un cadre méthodologique d'analyse de la dimension spatiale de la congestion en Île-de-France. Ainsi, l'analyse des interdépendances entre les différentes formes de congestion permet de mieux cerner les possibilités de mesure et de modélisation de la congestion, et d'en souligner les limites.

Deuxième partie – Les manifestations spatiales de la congestion du transport

Cette partie est consacrée à l'analyse des manifestations spatiales de la congestion en Île-de-France³. Elle correspond donc à l'approche *ascendante* que nous avons évoquée précédemment.

Le chapitre 3, *Méthodologie d'analyse de la congestion sur un réseau de transport*, rappelle dans un premier temps les principaux éléments de la démarche de modélisation retenue. Dans un second temps, il explicite les différentes dimensions à intégrer dans une analyse d'ingénieur-planificateur. En particulier, une série d'indicateurs, correspondant à trois perspectives d'évaluation, est définie. Par ailleurs, plusieurs méthodes d'analyse, à différentes échelles, sont également proposées. Le chapitre vise à déterminer les indicateurs et les méthodes permettant de saisir, sous différents aspects, les logiques spatiales sous-jacentes à la congestion des transports. L'objectif est bien de disposer d'un outillage le plus complet possible pour aborder l'analyse de la congestion en Île-de-France.

3. Une part importante des travaux présentés dans cette partie ont pour origine un rapport de recherche, rédigé par Fabien Leurent et l'auteur de ces lignes, commandité par le Ministère chargé des Transports visant à réactualiser la méthode d'évaluation des gains de temps de décongestion routière : Leurent *et al.* (2009).

Le chapitre 4, *Analyse de la congestion routière en Île-de-France : l'émergence de logiques spatiales*, commence par une présentation succincte du contexte francilien, avant d'amorcer une exploration de la variabilité spatiale de la congestion en Île-de-France, à partir des indicateurs et des méthodes définis dans le chapitre précédent. Le chapitre fournit ainsi une analyse spatialisée des effets de la congestion en Île-de-France, ainsi qu'une estimation du coût social de la congestion en Île-de-France. Il permet de mettre effectivement en évidence les logiques spatiales de la congestion des transports, et d'analyser, sous cet angle, certaines questions d'équité qu'elles soulèvent.

Le chapitre 5, *Coût social marginal et aide à la décision : la décongestion routière dans l'évaluation d'un projet de transport*, constitue une étude à part entière. Il propose, sur la base des analyses menées dans le chapitre précédent, une comparaison de plusieurs méthodes d'évaluation des gains de temps liés à la décongestion routière. Autrement dit, s'il n'entre pas directement dans le cheminement général de la thèse, il apporte des éléments opérationnels pour améliorer la prise en compte la variabilité spatiale de la congestion dans les évaluations de projets de transport.

Troisième partie – Localisation des activités et usage des transports

Le chapitre 6, *Un modèle d'équilibre urbain à distribution exogène des lieux d'emplois*, décrit un modèle urbain de moyen terme de type monocentrique dans lequel les emplois sont distribués de manière exogène. Le choix de différentes hypothèses est d'abord justifié sur la base d'une analyse bibliographique, puis les conditions nécessaires et suffisantes de l'équilibre du modèle sont déterminées dans le cas général, sous la forme d'un système différentiel, et l'existence d'un tel équilibre est montrée. L'objectif est de disposer d'un modèle permettant d'appréhender la manière dont la distribution spatiale des emplois influe sur la localisation des ménages et l'usage des transports.

Le chapitre 7, *Résolution analytique et sensibilité du modèle à l'équilibre*, propose une analyse, en statique comparative, de l'influence de l'ensemble des paramètres du modèle sur l'équilibre. Cette analyse est menée dans un cadre moins général qu'au chapitre précédent, à partir de formes fonctionnelles simples, donnant accès à une résolution analytique complète du modèle. En particulier, cette analyse de sensibilité permet de rendre compte de la façon dont les choix de localisation des ménages sont influencés par l'intensité de la congestion et sa distribution spatiale.

Le chapitre 8, *Décentralisation des emplois et usage des transports*, s'intéresse plus spécifiquement au rôle joué par l'étalement des emplois sur la localisation des ménages (et l'étalement urbain), les distances et les coûts de transport. En particulier, le chapitre aborde la question de l'étalement urbain et de l'accroissement

du coût de transport et de la congestion qui y sont généralement associés. L'influence des coûts marginaux de transport, lorsqu'ils varient dans l'espace, mise en évidence au chapitre précédent, est également mise en scène dans le cadre d'une décentralisation de l'emploi.

Première partie .

**Structure urbaine et congestion :
une analyse bibliographique**

1

Du processus d'agglomération à la structure urbaine : le rôle de l'espace

1.1. Introduction

Ce mémoire vise une meilleure compréhension des interactions entre les localisations urbaines et les coûts de transport, et du rôle de la congestion dans ces interactions. Pour commencer à appréhender ces liens, le présent chapitre propose de dresser un tableau des processus à l'œuvre dans l'existence et la croissance des villes, afin de mettre en évidence comment l'organisation interne des villes émerge de ces processus.

Ainsi, l'objectif de ce chapitre est de montrer que le phénomène d'agglomération relève de multiples mécanismes imbriqués, se manifestant généralement sous la forme d'externalités positives, et que les transports y jouent un rôle essentiel. Nous chercherons également à comprendre comment la nécessité, pour les agents soumis à ces forces d'agglomération, de se localiser dans l'espace fait émerger une structure urbaine.

La localisation initiale des villes dépend vraisemblablement de facteurs naturels, ainsi que de circonstances historiques (Rosenthal et Strange, 2004 ; Huriot et Bourdeau-Lepage, 2009). L'émergence des villes que nous connaissons est, du moins initialement, le résultat d'un subtil mélange entre hasard et nécessité. Toutefois, le phénomène d'agglomération ne peut se résumer à cela, et de puissantes forces en

sont à l'origine. La ville, agglomération de personnes et d'équipements, est le résultat d'un jeu complexe de forces d'attraction — centripètes — et de répulsion — centrifuges (Fujita et Thisse, 1996). Elle est, de ce fait, un exemple typique d'auto-organisation, c'est-à-dire une organisation globale émergeant spontanément à partir d'un état presque homogène ou aléatoire. Plus précisément, selon Krugman (1998) l'auto-organisation se caractérise par trois aspects. Tout d'abord, la configuration globale d'un système auto-organisé est déterminée par les interactions entre ses éléments. Ce premier trait souligne l'importance des interactions, qui sont en partie déterminées par la structure interne du système que constitue la ville. Autrement dit, étudier la structure interne des villes, c'est également donner des informations sur la manière dont elles se développent. Ensuite, dans un système auto-organisé, la formation d'un ordre global résulte de l'effet cumulatif des interactions individuelles. Enfin, l'auto-organisation engendre elle-même ses propres limites sous la forme d'un processus auto-limitatif qui freine le mouvement vers un ordre extrême.

Les mécanismes à l'origine de ces forces ont été largement étudiées dans le cadre de la *nouvelle économie géographique* (NEG) et de la *nouvelle économie urbaine* (NEU) (Fujita *et al.*, 2001), mais restent, encore aujourd'hui, un des enjeux majeurs de la recherche en économie spatiale. Comprendre ce qui régit la formation, puis l'évolution des villes est en effet crucial si l'on souhaite pouvoir orienter leur développement.

La croissance des villes soulève en effet de nombreuses questions. Comment les villes émergent-elles ? Quels sont les phénomènes qui font croître leur population ? Quels sont ceux qui les font s'étendre ? Pourquoi ne finissent-elles pas par recouvrir l'ensemble de l'espace ? Les deux premières de ces questions font référence à la question de l'agglomération, tandis que les deux suivantes à celle de la dispersion. Les quatre questions sont en lien direct avec la structure interne des villes, leur organisation spatiale, à travers les localisations des activités, entreprises et ménages, et le rôle des transports. Si les deux premières questions sont surtout l'objet d'étude de la NEG, qui s'est focalisée sur le phénomène d'agglomération, les deux autres ont été surtout étudiées dans le cadre de la NEU (Fujita *et al.*, 2001). Ce sont elles qui guideront donc, dans ce chapitre, notre réflexion.

Plus précisément, la deuxième section de ce chapitre est consacrée à la description des mécanismes menant aux économies d'agglomération, et insiste sur le rôle, souvent négligé, des réseaux de transport dans ce processus. La troisième section aborde la question du rôle de l'espace, par l'intermédiaire de la rareté du sol et des coûts de transports, et des arbitrages qui en résultent. Enfin, la quatrième section est consacrée à une exploration du modèle monocentrique, pierre angulaire de l'économie urbaine.

1.2. Une agglomération aux multiples sources

Les villes attirent, comme l'illustre la croissance ininterrompue du taux d'urbanisation dans le monde, passé de 28 % en 1950 à plus de 50 % en 2010 (United Nations, 2010). C'est pourquoi elles sont ce regroupement d'individus et d'activités que l'on connaît. Au-delà de ce simple constat surgit la question des causes d'une telle attirance des individus pour la ville. Si on a pu évoquer une tendance naturelle à se rassembler, un instinct grégaire (Huriot et Bourdeau-Lepage, 2009, chap. 1), il nous semble que cela reviendrait à expliquer l'agglomération par une inclination naturelle à s'agglomérer.

Comme nous l'avons évoqué dans l'introduction générale de ce mémoire, la nouvelle économie géographique s'est attachée à mieux comprendre le phénomène d'agglomération économique. Néanmoins, cette approche est restée essentiellement macroscopique au sens où les échelles en jeu sont de l'ordre de la région plutôt que de la ville. Or, comme le rappellent Fujita et Thisse (2009), l'agglomération économique est un terme vague : il peut désigner aussi bien la concentration des richesses dans les pays du Nord, que le rassemblement de plusieurs restaurants dans un même quartier, voire une même rue. La question de l'échelle spatiale des phénomènes étudiés est donc ici essentielle.

Des travaux ont visé à identifier les fondements microéconomiques des forces d'agglomération, dans un cadre spécifiquement urbain. Ainsi, Duranton et Puga (2004) et Rosenthal et Strange (2004) proposent des revues de ces travaux, pour les aspects théoriques et empiriques respectivement. Nous cherchons, pour notre part dans cette section, à souligner le caractère multi-échelle des mécanismes en jeu, afin de mieux comprendre comment ils peuvent faire émerger les structures urbaines actuelles.

1.2.1. La multiplicité organisée

1.2.1.1. Les dimensions structurantes de la multiplicité des sources d'agglomération

Les rendements d'échelle croissants sont généralement donnés pour responsables de la distribution géographique des activités économiques (Fujita et Thisse, 1996). Ce constat remonte d'ailleurs aux travaux de Marshall (1890, 1920), pionnier de l'économie industrielle et spatiale, qui distinguait quatre causes principales au phénomène d'agglomération :

- la production de masse ;
- la formation d'une main-d'œuvre hautement spécialisée, fondée sur l'accumulation de capital humain et les communications face-à-face ;
- la disponibilité de services spécialisés d'intrants ;

– l’existence d’infrastructures modernes.

Sous son apparente simplicité, cette classification n’apparaît toutefois pas véritablement opératoire. D’une part, elle distingue certains mécanismes en réalité semblables (la disponibilité d’une main d’œuvre spécialisée et de services spécialisés) et, d’autre part, elle regroupe d’autres mécanismes de nature différente (spécialisation de la main d’œuvre et l’accumulation de capital humain).

Depuis, de nombreuses autres typologies des rendements d’échelle croissants ont été proposées. L’une d’elles distingue les mécanismes selon le niveau de l’organisation productive où ils se manifestent (Huriot et Bourdeau-Lepage, 2009) : la firme, le secteur productif ou la ville dans son ensemble. Toutefois, parler de « rendements d’échelle croissants » dans un contexte plus large que la firme est en toute rigueur un abus de langage. Mais il est courant de l’employer pour les deux autres niveaux d’organisation que sont le secteur productif et la ville. De manière à pouvoir distinguer les échelles lorsque cela est nécessaire, nous emploierons les termes *rendements d’échelle croissants internes* pour le niveau de la firme et *rendements d’échelle croissants externes*¹ pour les niveaux supérieurs. Dans cette dernière expression, *externe* fait référence à des externalités au sens de Marshall, pour qui externe signifiait externe à la *firme*. Il ne s’agit donc pas nécessairement d’externalités dans l’acception microéconomie actuelle habituelle, pour laquelle une externalité est un phénomène externe au *marché*. Il est cependant possible de réconcilier ces deux visions et de désigner ces deux types d’interactions sous le même vocable d’*externalité*, en distinguant les externalités technologiques des externalités pécuniaires (Scitovsky, 1954). Une *externalité technologique*, ou effet externe pur, désigne le fait que les actions d’un agent influencent directement les possibilités de choix d’un autre agent, sans contrepartie financière. Formellement, cela revient à ce que les activités d’un agent entrent dans la fonction d’utilité ou de production d’un autre agent. Au contraire, les *externalités pécuniaires* désignent l’effet sur les autres agents d’une modification, par un agent, des prix auxquels ils peuvent engager une transaction (Small, 1999). Ces externalités pécuniaires ne se manifestent donc que dans des situations de concurrence imparfaite, à la différence des effets externes purs (Krugman, 1991b ; Salanié, 1998, chap. 5). De telles interactions de marché peuvent par exemple émerger du fait de l’existence de liaisons verticales entre deux firmes dont l’une est cliente de l’autre, car la proximité permet d’économiser des coûts de transport et plus généralement des coûts de transaction (Venables, 1996).

Dans la suite de ce chapitre, nous emploierons les termes, repris de Marshall (1890, 1920), de *rendements croissants internes* pour désigner les économies d’agglomération se manifestant au niveau de la firme, d’*économies de localisation* pour

1. Certains auteurs utilisent également les termes *rendements croissants agrégés localisés* dans ce contexte.

désigner celles prenant place au niveau d'un secteur productif, et d'*économies d'urbanisation* pour désigner celles qui se manifestent à l'échelle de la ville. Si ces trois types d'économies d'agglomération ne sont pas mutuellement exclusives, elles ont des implications différentes pour la nature de l'activité économique d'une zone urbaine.

De plus, certaines sources d'économies d'agglomération peuvent relever des deux dernières catégories en même temps. Ainsi Eberts et McMillen (1999) évoquent le cas d'une zone urbaine bénéficiant de la proximité d'une source d'électricité bon marché. Ce territoire peut attirer une industrie très consommatrice d'énergie, à travers des économies de localisation. Mais il peut également attirer une multitude de petites entreprises sans relation les unes avec les autres, mais pour lesquelles l'électricité représente une part importante de leurs coûts. Il s'agit alors d'économies d'urbanisation.

Par ailleurs, Duranton et Puga (2004) ont proposé une classification originale des différentes sources d'économies d'agglomération. Elle distingue trois grandes sources microéconomiques : le partage, l'appariement et l'apprentissage. Chacune de ces sources peut être à l'origine d'un phénomène d'agglomération par l'intermédiaire de divers mécanismes détaillés par les auteurs. Cette classification, outre son originalité, met ainsi en évidence la nature profonde des *mécanismes* à l'œuvre dans le phénomène d'agglomération. Néanmoins, elle présente l'inconvénient, selon nous, de ne pas mettre en valeur la nature des *interactions* en jeu — par l'intermédiaire du marché ou hors marché — ni l'échelle à laquelle elles ont lieu.

1.2.1.2. Une classification fondée sur le croisement de deux dimensions

Nous proposons, pour notre part, une classification fondée sur le croisement des deux dimensions que sont la nature et l'échelle des interactions à l'origine du phénomène d'agglomération. Ce choix est guidé par deux éléments. D'une part, les questions d'échelle sont au cœur des problématiques spatiales que nous explorons dans ce mémoire. D'autre part, nous souhaitons bien distinguer les mécanismes mettant en œuvre des externalités pécuniaires de ceux faisant intervenir des externalités technologiques. La Table 1.1 sépare les principaux mécanismes sources d'économies d'agglomération selon notre grille d'analyse échelle - interaction.

La suite de cette section est consacrée à une description ordonnée de ces mécanismes. Elle commence par une présentation des mécanismes faisant intervenir des interactions de marché, pour s'intéresser ensuite aux mécanismes se déroulant hors marché. À chaque fois, nous irons du niveau le plus restreint, celui de la firme, pour aller vers le niveau le plus vaste, celui de la ville ou de la région.

| Échelle | Type d'interaction | |
|-------------------|---------------------------------------------------------------|--------------------------------------------------------------------|
| | Marché | Hors marché |
| Firme | Indivisibilités et technologie Spécialisation individuelle | Échanges d'expérience |
| Secteur productif | Diversification des intrants | Innovation Diffusion de l'information |
| Ville / Région | Appariements Mutualisation des risques | Accumulation de capital humain Équipements publics indivisibles |

TABLE 1.1.: Classification des sources d'économies d'agglomération

1.2.2. Quand le marché crée l'agglomération

Les mécanismes de marché, en particulier les externalités pécuniaires, peuvent être à l'origine de puissantes forces d'agglomération. Ainsi, dans une situation de concurrence spatiale, les stratégies des firmes peuvent les pousser à se regrouper pour se partager plus avantageusement le marché (Hotelling, 1929).

Contrairement aux mécanismes hors marché, ces mécanismes liés au marché nécessitent souvent des modèles relativement complexes, mettant en jeu différentes formes de concurrence imparfaite. En dehors des rendements croissants internes, il s'avère que la diversité (diversification, complémentarité, mutualisation) est le principal moteur de l'agglomération. Que ce soit au niveau du secteur productif ou à une échelle plus large, les avantages qu'elle procure, aux entreprises, aux ménages et aux travailleurs se révèlent être une puissante force centripète.

1.2.2.1. Les rendements d'échelle internes croissants

Au niveau de la firme plusieurs mécanismes expliquent l'émergence de rendements d'échelle croissants (Picard, 2007, chap. 5). En premier lieu, le plus élémentaire d'entre eux est lié à l'indivisibilité des équipements ou des infrastructures, qui entraîne des coûts fixes parfois très élevés, lesquels se diluent au fur et à mesure que le volume de production augmente.

Tout d'abord, certains équipements² utilisés par les entreprises de biens comme de services sont **indivisibles** au sens où le même équipement est nécessaire pour produire une unité de bien ou de service, et à la production d'une grande quantité de ces biens ou services. Ainsi, une entreprise ne peut utiliser une demi chaîne de production automobile ou un demi ordinateur. De même, certains services géné-

2. Le terme *équipement* est ici employé en un sens large pouvant désigner aussi bien une machine qu'un service de support.

raux des entreprises n'ont pas à être développés lorsque le volume de production augmente.

Ces indivisibilités sont naturellement à l'origine de rendements d'échelle croissants, puisque la firme a intérêt à augmenter sa production tant qu'une unité supplémentaire produite fait baisser le coût moyen de production. Lorsque les coûts fixes sont très élevés, comme c'est souvent le cas pour des entreprises industrielles lourdes, elles peuvent suffire à entraîner une agglomération à l'échelle urbaine. Ainsi, des villes comme Le Creusot et Montbéliard doivent en grande partie leur croissance respectivement à l'implantation de Schneider et Peugeot en leur sein (Huriot et Bourdeau-Lepage, 2009).

La croissance de la production finit néanmoins par engendrer des coûts fixes supplémentaires, parfois gigantesques, qui limitent de fait la taille de l'unité de production. Les rendements d'échelle croissants sont généralement valables sur le court ou moyen terme. C'est pourquoi les situations que nous venons d'évoquer constituent plutôt l'exception, et ne concernent que des villes relativement petites.

Par ailleurs, certaines **technologies** de production employées peuvent être à l'origine de rendements d'échelle croissants même en l'absence de coûts fixes. Des rendements d'échelle croissants peuvent également apparaître en l'absence de coûts fixes. Il suffit, pour cela, que la valeur du bien ou du service produit augmente plus vite que le coût de ce bien ou de ce service, lorsque le volume de production augmente. À titre d'exemple, supposons qu'une entreprise ait besoin, pour assurer sa production, d'une cuve de stockage sphérique. D'une part, le coût de production d'une cuve de stockage sphérique est supposé proportionnel à la quantité de métal utilisée pour la produire, quantité elle-même proportionnelle à la surface de la cuve. D'autre part, la valeur de cette cuve, en termes de service rendu, est supposée proportionnelle au volume qu'elle est capable de contenir. L'entreprise a dans ce cas intérêt à massifier sa production, en regroupant éventuellement plusieurs de ses unités de production. En effet, lorsque r augmente, l'utilisation d'une telle cuve présente pour l'entreprise des rendements d'échelle croissants, puisque son volume, donné par $(4/3)\pi r^3$, et donc sa valeur, augmente plus vite que sa surface, $4\pi r^2$, et donc son coût. Autrement dit, le coût de stockage par unité de volume, proportionnel à $3/r$, diminue lorsque le volume augmente.

Ce raisonnement peut être directement appliqué à une ville, assimilant cette dernière à une sorte de *grande* unité de production. Ainsi, la production du mur d'enceinte d'une ville (Huriot et Thisse, 2000) présente des rendements d'échelles croissants : alors que la population protégée par un mur circulaire de rayon r est, à densité constante, proportionnelle à l'aire encerclée πr^2 , son coût est quant à lui proportionnel au périmètre $2\pi r$. Le coût moyen de « protection » par habitant est donc proportionnel à $2/r$, et décroît avec r , donc avec la population de la ville.

Si l'existence de villes primitives ou fortement spécialisées est probablement liée à ce type de grandes indivisibilités, il est peu réaliste d'imaginer que le développement des mégapoles actuelles ne soit dû qu'à elles. Il s'agit plutôt de la somme de nombreuses activités diverses sujettes à de petites non-convexités (de petits rendements d'échelle croissants), qui engendrent globalement des économies d'échelle³.

Enfin, on constate que l'augmentation du volume de la production permet généralement une plus grande **spécialisation** des travailleurs, et par suite une augmentation de leur productivité. L'idée que l'augmentation du volume de la production, et donc de la taille de la firme, permet aux travailleurs de devenir plus efficaces en se spécialisant n'est pas nouvelle. Elle remonte à la manufacture d'épingles de Smith (1776) ou encore à Marshall (1890). Les auteurs parlent d'ailleurs souvent d'approche « *Marshallienne* » pour désigner les économies d'agglomération issues de ce mécanisme, dont la formalisation est récente. Ainsi, Duranton (1998) obtient des rendements d'échelle croissants en combinant une fonction de production globale faisant intervenir plusieurs tâches élémentaires, et un mécanisme individuel de spécialisation grâce auquel les travailleurs peuvent améliorer leur productivité et par conséquent leur salaire.

1.2.2.2. La diversité est une force

Introduites par Marshall (1890, 1920), les économies de localisation de marché découlent des avantages liés à la diversification et à la spécialisation des entreprises au sein d'un secteur productif :

« Lorsqu'une industrie a ainsi choisi une localité, elle a des chances d'y rester longtemps, tant sont grands les avantages que présente pour des gens adonnés à la même industrie qualifiée, le fait d'être près les uns des autres. [...] Bientôt des industries subsidiaires naissent dans le voisinage, fournissant à l'industrie principale les instruments et les matières premières, organisant son trafic, et lui permettant de faire bien des économies diverses. [...] Car des industries subsidiaires se consacrant chacune à une petite branche de l'œuvre de production, et travaillant pour un grand nombre d'entreprises voisines, sont en état d'employer continuellement des machines très spécialisées. » (Marshall, 1890, chap. 10)

Toutefois, certains auteurs ont suggéré que la diversification des activités n'était pas nécessaire pour qu'une ville croisse. Il se peut que seule une diversification intra-

3. On distingue classiquement les économies d'échelle (*scale economies*) des économies d'envergure (*scope economies*). Néanmoins, ces deux notions sont souvent très fortement liées et difficiles à distinguer. Nous emploierons donc le terme économies d'échelle pour désigner indistinctement ces deux types d'économies.

sectorielle associée à une spécialisation sectorielle fasse émerger une force d'agglomération suffisante. Duranton et Puga (2000) montrent d'ailleurs que villes diversifiées et spécialisées coexistent et que les entreprises ont tendance à naître dans des villes diversifiées pour migrer ensuite vers des villes spécialisées. De même, Henderson *et al.* (1995) montrent que la croissance des villes est d'autant plus élevée que la spécialisation de ces villes dans un secteur est forte, en contrôlant pour le nombre d'emplois de ce secteur. Mais Combes (2000b) note qu'un tel contrôle ne permet en réalité que de montrer que les villes petites croissent plus vite que les grandes. Il obtient, quant à lui, à partir de données françaises, que la spécialisation urbaine ne favorise pas la croissance, mais qu'au contraire, la diversité est positive, en particulier pour les services (Combes, 2000a). Autrement dit, si la spécialisation des entreprises et des travailleurs peut constituer une force d'agglomération, il n'est pas acquis que la spécialisation urbaine favorise la croissance. Comme le résumant Rosenthal et Strange (2004), les différents résultats obtenus suggèrent fortement que la diversité est utile.

Les deux phénomènes de diversification et de spécialisation sont intimement liés : en augmentant le nombre de variétés offertes d'un même produit ou service, les entreprises, d'une part, attirent un nombre croissant de clients, car ceux-ci apprécient la variété, et elles peuvent, d'autre part, se spécialiser dans la production d'une seule variété, et donc améliorer leur productivité, par le mécanisme mis en évidence au paragraphe précédent.

La **diversification**, qu'elle ait lieu en amont ou en aval de la production, est donc un mécanisme permettant de faire émerger des rendements d'échelle croissants au niveau d'un secteur productif dans son ensemble, ou plus précisément au niveau de la fonction de production — ou de consommation — agrégée du secteur. Il s'agit d'une approche que Fujita et Thisse (2002) et d'autres auteurs qualifient de « Chamberlinienne ».

Le modèle de Abdel-Rahman et Fujita (1990) formalise ce type de mécanismes⁴. L'économie y est constituée de deux secteurs : le secteur des biens manufacturés, qui ne produit qu'un seul type de biens, et le secteur des services, proposant des services variés. Les auteurs obtiennent, par l'intermédiaire d'une fonction de production de type CES, qu'une telle économie est sujette à des rendements d'échelle croissants avec le nombre de travailleurs. Ce résultat passe par deux mécanismes : d'une part, le secteur manufacturé a une préférence pour la diversité, ce qui stimule l'entrée de nouvelles entreprises de services, et d'autre part, le secteur des services est sujet à des coûts fixes, autrement dit à une forme d'indivisibilité, qui l'encourage à embaucher.

4. Fujita et Thisse (2002, chap. 4) étendent ce modèle à un *continuum* de variétés.

Le même modèle s'applique presque directement aux ménages, en supposant qu'ils ont également une préférence pour la diversité. Autrement dit, ce mécanisme lié à la diversité des produits ou des services permet d'expliquer l'agglomération de l'activité productrice et des ménages, et suggère l'existence d'un cercle « vertueux ». Les ménages se concentrent car en travaillant dans le secteur productif qui tend lui-même à se concentrer ils obtiennent un revenu supérieur et ils bénéficient d'une plus grande variété de produits, et les entreprises se concentrent pour rester proches des ménages. Ce type de processus est évoqué, sous les termes d'agglomérations secondaires et tertiaires par Hochman (2010).

Par ailleurs, Marshall (1890) avançait déjà l'idée selon laquelle, dans une ville, l'augmentation du nombre de travailleurs pouvait faciliter la recherche d'emploi et améliorer l'adéquation entre emploi et travailleur. Là aussi, c'est la diversité des travailleurs et des emplois urbains qui produit l'agglomération :

« [...] une industrie localisée tire un grand avantage du fait qu'elle est constamment un marché pour un genre particulier de travail. Les patrons sont disposés à s'adresser à un endroit où ils ont des chances de trouver un bon choix d'ouvriers possédant les aptitudes spéciales qu'il leur faut ; de leur côté les ouvriers cherchant du travail vont naturellement dans ces endroits où se trouvent beaucoup de patrons ayant besoin d'ouvriers de leur spécialité et où ils ont, par suite, des chances de trouver un marché avantageux. [...] Le propriétaire d'une fabrique isolée est souvent mis dans de grands embarras lorsqu'il a subitement besoin d'ouvriers d'une certaine spécialité, et un ouvrier spécialisé, qui cesse d'être employé par lui, a du mal à se tirer d'affaire. »(Marshall, 1890, chap. 10)

Nous l'avons vu dans les paragraphes précédents, la diversité des produits est à la source de forces d'agglomération. Nous allons maintenant nous intéresser à la diversité des *travailleurs* ou des *activités*, car elles peuvent, elles aussi, être à l'origine de forces d'agglomération. Ce type de forces agit généralement au niveau de la ville dans son ensemble. Il ne s'agit pas, comme au paragraphe précédent, d'une diversification intra-sectorielle, autrement dit d'une augmentation du nombre de variétés d'un bien ou d'un service, mais bien d'une diversification des activités, autrement dit du regroupement, dans la ville, d'activités différentes. Plus précisément, conformément à l'« intuition » de Marshall (1890), le rassemblement d'activités diversifiées entraîne une probabilité plus forte de trouver un partenaire, sur les différents marchés, aux caractéristiques recherchées, et donc de réaliser un meilleur appariement.

L'effet de la diversification des activités sur la qualité des **appariements** peut être mise en évidence en adaptant au marché du travail le modèle de Salop (1979)⁵ (Helsley et Strange, 1990 ; Kim, 1990, 1991). Le modèle repose sur une hypothèse simple : lorsqu'une entreprise embauche un travailleur dont les compétences ne correspondent pas parfaitement à l'emploi, le travailleur subit un coût lié à ce mauvais appariement, ou « *mésappariement* »⁶. Dans ce cadre, chaque entreprise propose un salaire lui permettant de maximiser son profit, et chaque travailleur choisit la firme qui lui permet d'obtenir la rémunération la plus élevée, nette des coûts de mésappariement.

La résolution de ce modèle, telle que proposée par Duranton et Puga (2004), montre l'existence de rendements d'échelle croissants, au niveau de la ville dans son ensemble. Ceux-ci proviennent de deux sources distinctes. Tout d'abord, la concurrence entre les entreprises limite l'entrée de nouvelles firmes lorsque la taille de la main d'œuvre augmente. Plutôt qu'à une multiplication des entreprises, on assiste donc à une croissance du nombre de travailleurs par entreprise, ce qui augmente, du fait de la présence de coûts fixes dans la technologie de production, la production globale. Toutefois, l'intérêt de ce premier résultat est à relativiser, puisque les entreprises modélisées sont elle-mêmes sujettes à des rendements croissants.

Le second résultat donne tout son intérêt à ce modèle du marché du travail. Ainsi, le modèle fait émerger une *externalité d'appariement* qui améliore le revenu par travailleur. On obtient également que le revenu moyen augmente lorsque la main d'œuvre croît, et ce, pour deux raisons : d'une part du fait de l'augmentation de la production agrégée qui se traduit par une hausse du salaire, d'autre part du fait de la réduction des coûts de mésappariement. Il s'agit là d'une puissante incitation, pour les travailleurs, à se regrouper dans la même ville. Toutefois, Duranton et Puga (2004) notent dans le même temps que le caractère *externe* (par rapport à la firme) des gains (ou des coûts) liés à ce processus d'appariement peut être à la source d'inefficacités au niveau de l'économie de la ville. D'un côté, les entreprises, dans le modèle, ne prennent pas en compte l'avantage, au niveau social, qu'elles fournissent en entrant sur le marché, si bien qu'elles peuvent y entrer en un nombre insuffisant. D'un autre côté, en entrant sur le marché, elles débauchent des travailleurs des autres entreprises, ce qui est socialement inefficace, du fait de la présence de coûts fixes. Autrement dit, les mécanismes d'appariement sont une source délicate de l'agglomération.

5. Le modèle de Salop (1979) traite, pour sa part, de la différenciation spatiale des entreprises et du *business stealing* (Salanié, 1998, chap. 9).

6. Cette spécification fait donc entièrement peser le coût du mésappariement sur le travailleur, ce qui est le cas en situation d'information parfaite. L'existence d'un aléa moral pourrait changer la donne, mais ne changerait vraisemblablement pas le résultat.

Nous venons de le voir, la croissance du nombre de travailleurs permet un *meilleur* appariement des compétences avec les exigences des entreprises. Un second phénomène s'y ajoute : une main d'œuvre abondante facilite la rencontre des emplois et des travailleurs. Autrement dit, la *probabilité* d'un appariement serait d'autant plus élevée que le marché du travail serait vaste.

Coles et Smith (1998) proposent ainsi un modèle macroscopique mettant en évidence le rôle de la taille du marché du travail sur la probabilité des appariements. Le modèle de Berliant *et al.* (2006) met, quant à lui, l'accent sur le mécanisme en jeu. Il repose sur l'hypothèse qu'un travailleur possède une certaine compétence et que sa productivité est potentiellement augmentée s'il coopère (s'apparie) avec un travailleur dont la nature de la connaissance n'est ni trop proche ni trop éloignée. Les auteurs obtiennent qu'avec la croissance de la population de la ville, l'écart de compétence moyen se rapproche de l'écart idéal, ce qui signifie que la qualité des appariements augmente, comme dans le modèle que nous avons présenté auparavant. De plus, ces auteurs obtiennent que la probabilité de s'apparier augmente pour un nouvel arrivant. Ce double mécanisme fournit une force d'agglomération pour les travailleurs, qui peuvent augmenter leur revenu espéré de long terme de deux manières : en réduisant le coût de mésappariement et en limitant le temps passé sans emploi.

Marshall (1890) avait déjà pressenti l'intérêt de la diversité dans la **mutualisation des risques** :

« Une industrie localisée offre quelques inconvénients, en tant que marché de travail, si, dans le travail qui s'y fait, une seule espèce prédomine [...] Mais le remède à ce mal est évident, et il est fourni par le développement dans la même région d'industries d'un caractère supplémentaire. [...] Les avantages qu'offre la variété d'occupations se combinent avec ceux de la localisation de l'industrie dans certaines de nos grandes villes manufacturières, et c'est là l'une des principales causes de leur progrès continu. »

« Une région qui vit surtout d'une seule industrie est exposée à une crise très grave, au cas où la demande de ses produits vient à diminuer, comme au cas où la matière première dont elle se sert vient à manquer. Cet inconvénient lui-même est en grande mesure évité par l'existence de ces grandes villes et de ces grandes régions industrielles où plusieurs industries différentes se trouvent développées. Si l'une vient à manquer pendant quelque temps, les autres peuvent lui venir en aide indirectement. » (Marshall, 1890, chap. 10)

La variété est donc un facteur de réduction du risque (Huriot et Bourdeau-Lepage, 2009). Ainsi, une agglomération strictement spécialisée est plus sensible aux aléas

du marché qu'une agglomération diversifiée. Si, là encore, l'intuition remonte à Marshall (1890), Krugman (1991a) en propose une modélisation formelle.

Dans son modèle, les entreprises interviennent chacune sur un marché différent et sont susceptibles de subir un choc de productivité exogène et idiosyncratique. Ainsi, on peut imaginer que ce choc est lié à des aléas du marché sur lequel elle opère. Il obtient que le profit espéré des firmes augmente avec la variabilité du choc idiosyncratique, ainsi qu'avec celle des salaires, mais il diminue avec leur covariance. Ce dernier résultat s'interprète de la manière suivante : si la réalisation d'un choc positif, entraînant une augmentation de la production de la firme, est corrélée avec une hausse des salaires, ou au contraire si un choc négatif, entraînant une contraction de la production, s'accompagne d'une baisse des salaires, alors la marge de manœuvre de l'entreprise est réduite, et le profit de la firme s'en ressent.

Un calcul effectif de la variance du salaire et de la covariance du choc et du salaire permet finalement d'obtenir le résultat recherché : un effet de mutualisation des travailleurs. Chaque firme bénéficie du partage d'un même marché du travail avec les autres firmes, ce qui amortit les chocs. Cet avantage augmente avec le nombre de firmes, mais également avec la variance du choc. Il s'agit bien là d'un phénomène de mutualisation des risques, puisque l'existence d'un large marché du travail, en raison de la présence des autres firmes, permet à une entreprise donnée de pouvoir embaucher en cas d'accroissement de sa production, ou de débaucher dans le cas contraire, en conservant une bonne marge de manœuvre du fait de la constance du salaire.

Ce modèle peut être étendu au cas où les entreprises fixent individuellement les salaires, et où le chômage est possible. L'effet d'agglomération est dans ce cas encore renforcé car les travailleurs cherchent eux aussi à mutualiser les risques du chômage en se regroupant. Par ailleurs, Stahl et Walz (2001) ont modifié ce type de modèles en introduisant des chocs intra-sectoriels spécifiques en plus de chocs touchant les firmes individuellement. Leurs résultats confirment l'importance de la diversité dans la réduction des risques.

1.2.3. « The secrets of the industry are in the air »⁷

Nous venons de le voir, les mécanismes de marché peuvent être à l'origine de forces d'agglomération poussant les entreprises, et donc les ménages, à se regrouper en certains points privilégiés de l'espace. Mais il existe également des interactions hors-marché, c'est-à-dire des externalités pures, qui favorisent l'agglomération économique. La majorité de ces interactions hors marché sont de nature informationnelle. De plus, ce sont généralement des externalités spatiales, ou de proximité, au sens où leur effet diminue avec la distance à l'agent qui les émet. Toutefois,

7. Marshall (1890) : « Les secrets de l'industrie sont dans l'air ».

nous le verrons, certaines infrastructures publiques peuvent indépendamment être à l'origine d'externalités favorisant l'agglomération.

Le fait que l'information, sous toutes ses formes, puissent être à l'origine d'une force d'agglomération n'est pas nouvelle, puisqu'elle remonte aux années 1890 avec Marshall :

« Lorsqu'une industrie a ainsi choisi une localité, [...] Les secrets de l'industrie cessent d'être des secrets ; ils sont pour ainsi dire dans l'air, et les enfants apprennent inconsciemment beaucoup d'entre eux. On sait apprécier le travail bien fait ; on discute aussitôt les mérites des inventions et des améliorations qui sont apportées aux machines, aux procédés, et à l'organisation générale de l'industrie. Si quelqu'un trouve une idée nouvelle, elle est aussitôt reprise par d'autres, et combinée avec des idées de leur crû ; elle devient ainsi la source d'autres idées nouvelles. »
(Marshall, 1890, chap. 10)

Par ailleurs, Mills (1967) et Henderson (1974) ont utilisé, bien avant le développement de la nouvelle économie géographique et de la théorie de l'agglomération, ce type d'économies d'agglomération hors marché en économie urbaine. Les mécanismes à l'œuvre sont variés, comme les modélisations qui en ont été proposées. Par ailleurs, de manière similaire aux mécanismes passant par le marché que nous avons vus plus haut, ils interviennent à plusieurs niveaux, de la firme à la ville dans son ensemble.

À l'échelle de la firme, le fait, pour un travailleur, de côtoyer d'autres travailleurs, plus qualifiés parce que plus **expérimentés**, lui permet d'acquérir plus rapidement un niveau d'expérience à même d'améliorer sa productivité. C'est ce type de mécanismes qu'évoquent par exemple Kogut et Zander (1992). En réalité, il s'agit d'un mécanisme proche de celui de diffusion spatiale de l'information, que nous aborderons plus loin dans ce chapitre, à une échelle plus vaste.

À une échelle spatiale supérieure à la firme, un environnement urbain diversifié favorise vraisemblablement l'expérimentation et l'**innovation**. Duranton et Puga (2001) proposent ainsi un mécanisme d'apprentissage dans lequel les firmes d'un secteur essaient différents *process*, en vue de trouver le meilleur. Les firmes sont par ailleurs d'autant plus efficaces qu'elles sont entourées d'autres firmes du même secteur, utilisant le même type de *process*. Les auteurs obtiennent qu'à l'équilibre coexistent des villes spécialisées et des villes diversifiées. Les travaux empiriques de Duranton et Puga (2000) montrent qu'une telle coexistence est effectivement observable.

L'idée de Marshall selon laquelle la colocalisation permet des échanges informels de connaissances, et donc la **diffusion de l'information**, favorisant l'agglomé-

ration des entreprises a été formalisée dans des modèles visant à représenter les mécanismes internes en jeu de manière à en analyser les conséquences sur la structure urbaine.

Plus spécifiquement, le modèle de Ogawa et Fujita (1980) et Fujita (1982) permet de fixer les idées. Sur une surface donnée S , on suppose l'existence d'un continuum de firmes identiques. En particulier, chacune d'elles « diffuse », sous forme de *spillovers* involontaires, la même quantité d'information utilisable de la même manière par toutes les autres. La ville est ainsi baignée d'un champ informationnel. Si $a(x, y)$ fournit la valeur des avantages tirés par une entreprise située en x de l'information diffusée par une entreprise située en y , et si $f(y)$ détermine la densité de firmes à chaque position $y \in S$, alors le gain agrégé qu'une firme en x peut tirer du champ informationnel de la ville s'écrit :

$$A(x) = \int_S a(x, y)f(y)dy. \quad (1.1)$$

Le profit d'une firme dépend de la production des entreprises autour d'elle et de leur concentration, et ce sans compensation financière. On est donc bien en présence d'une externalité technologique favorisant le rapprochement des firmes dans l'espace.

Ce type de modèles permet d'expliquer l'émergence d'un centre d'emplois (*Central Business District* ou CBD) et de poser ainsi les bases d'une structure urbaine spatiale (Imai, 1982). La diffusion de l'information est, de manière générale et du fait de la simplicité de sa modélisation, un mécanisme souvent employé dans des modèles explorant les questions de structure urbaine. Nous mobiliserons le modèle précédent au chapitre 8. Toutefois, ces modèles ne mettent pas directement en évidence de déséconomies d'agglomération. C'est généralement la congestion des transports qui fournit la force de dispersion et limite le processus d'agglomération.

L'accumulation de connaissances sous la forme d'un **capital humain** peut également être à la source de forces d'agglomération. Pour le mettre en évidence, on peut, comme Palivos et Wang (1996), supposer que les producteurs d'un bien final font individuellement face à des rendements d'échelle constants, mais qu'au niveau agrégé, les rendements d'échelle sont croissants. De plus, le bien final peut être accumulé, sous la forme d'un capital humain, s'il n'est pas consommé. Cette double spécification favorise le rassemblement des travailleurs. Toutefois, ce modèle n'est pas spécifique d'un mécanisme d'accumulation du capital humain et nous semble dès lors relativement abstrait.

Les modèles développés par Lucas (1988) ou Eaton et Eckstein (1997) sont au contraire plus spécifiques dans leurs mécanismes. Ils supposent que les travailleurs, dont la fonction de production dépend de leur capital humain, peuvent consacrer

une partie de leur temps à accumuler ce capital. Or, le mécanisme d'accumulation, ou « fonction d'apprentissage », est d'autant plus efficace que le capital humain de la ville dans son ensemble est élevé. On peut penser ici à des équipements culturels, comme les musées et les bibliothèques. Ce mécanisme introduit une externalité : le fait que les autres travailleurs accumulent du capital accélère ce processus pour chacun d'entre eux. Il s'agit donc d'une externalité d'apprentissage, qui joue le double rôle de moteur de la croissance et de force d'agglomération pour les travailleurs.

1.2.4. Les transports : à la fois sources et conditions nécessaires à la réalisation des économies d'agglomération

Nous l'avons vu, de nombreux mécanismes à l'origine du phénomène d'agglomération sont liés à l'existence d'externalités résultant du partage, par les entreprises en particulier, de « biens » non exclusifs : variété des facteurs de production, nombreux travailleurs permettant d'améliorer les appariements et de réduire le risque, expertise technique et informations. Les **équipements publics indivisibles** peuvent également jouer ce rôle.

Les réseaux de communication et de transport ont dans ce cadre un double rôle, qui nous semble essentiel. Tout d'abord, ils font partie de ces biens publics dont peuvent bénéficier les firmes localisées en ville. Ils jouent donc un rôle de facteur de production pour les entreprises et sont, à ce titre, en eux-mêmes une source d'économies d'agglomération.

Ensuite, ils rendent possible la proximité en temps et en coût des entreprises ou des travailleurs dans l'espace. Plus précisément, comme le notent Huriot et Bourdeau-Lepage (2009, chap. 1), la proximité géographique, permanente ou temporaire, est essentielle dans l'existence et le fonctionnement des villes. Or, ces deux formes de proximité reposent en grande partie sur les réseaux de transport — transports urbains pour la première, interurbains pour la seconde. Les infrastructures de transport affectent directement l'efficacité des activités urbaines, en particulier dans les plus grandes villes, et favorisent donc la réalisation des économies d'agglomération (Eberts et McMillen, 1999). Ainsi, sans un réseau d'autoroutes efficace, les bénéfices obtenus par la proximité élevée des entreprises et des ménages pourraient être entièrement perdus du fait des blocages dans le transport des personnes et des biens. Aussi des villes de taille comparables peuvent-elles bénéficier de niveaux de productivité différents, du fait, notamment, des différences de qualité et de taille de leurs infrastructures de transport. En donnant tout son sens à la *centralité* urbaine, les transports sont donc le vecteur de forces d'agglomération. Dans un article très riche mobilisant des variables instrumentales pour s'affranchir du biais d'endogénéité qui entache de nombreux travaux sur la question, Duranton et Turner (2008a) obtiennent, pour les États-Unis, qu'une augmentation de 10 % de la longueur du

réseau routier d'une ville entraîne une croissance de 2 % de sa population. Ce chiffre peut paraître faible, mais il est significatif et démontre le rôle des infrastructures de transport dans le développement urbain⁸.

Pourtant, comme le notent Eberts et McMillen (1999) et Duranton et Turner (2008a), une partie importante de la littérature portant sur les économies d'agglomération est totalement déconnectée de la question des infrastructures publiques, au sens où la majorité des études s'intéressant aux économies d'agglomération négligent l'influence des infrastructures publiques et en particulier de transport, sur la productivité. Nous nous proposons donc, dans cette section, d'évoquer ces deux modes d'action des transports.

1.2.4.1. Les transports comme facteur de production

Nous l'avons évoqué au 1.2.2.1 au niveau de la firme, certains biens ou équipements peuvent comporter des indivisibilités. Cela peut également se retrouver à l'échelle urbaine, par l'intermédiaire de certains équipements publics de grande ampleur. Ainsi, s'il est envisageable pour une population d'une certaine taille de faire construire et d'entretenir un équipement public comme un musée ou un stade, construire pour chacun un tel équipement d'une taille réduite ne l'est pas (Duranton et Puga, 2004). La raison est semblable à celle citée au niveau de la firme : les coûts fixes qu'impliquent un tel équipement ne sont pas compatibles avec l'utilisation épisodique qu'en fera la majorité de la population. Par conséquent, pour bénéficier des services rendus par un tel équipement, les agents doivent se regrouper. De plus, le fait qu'un tel équipement présente des rendements d'échelle croissants a pour conséquence directe une concentration spatiale de celui-ci. Autrement dit, non seulement la présence d'un musée ou d'un stade est un atout pour la ville dans son ensemble, mais en plus la localisation précise de cet équipement fournit un avantage à ceux qui sont à proximité, et ce, en dehors du marché (Huriot et Bourdeau-Lepage, 2009)⁹.

Ainsi, la présence d'un aéroport international favorise les entreprises de la ville et en attire de nouvelles, alors même que cet aéroport peut être financé au niveau national. L'aéroport procure des avantages à toutes les entreprises de la ville, sans que cette interaction ne passe par les mécanismes du marché. Mais elle en procure, de manière relative, encore plus aux entreprises qui se situent à proximité immédiate

8. De plus, ce résultat concerne les États-Unis, et il est probable qu'un chiffre bien plus élevé aurait été obtenu dans des pays moins industrialisés, en Afrique par exemple, car on peut supposer que le rôle des transports dans le développement urbain est sujet à des rendements décroissants.

9. On pourra noter ici que ce raisonnement n'est en pratique valable que pour les équipements qui ne créent pas de nuisances par eux-mêmes. Dans le cas contraire, on assiste souvent au syndrome NIMBY (*Not In My Back Yard*), voire au syndrome OIOBY (*Only In Others' Back Yard*) (voir, par exemple, Dear, 1992 ; Jobert, 1998).

de l'aéroport¹⁰. Toutefois, ce type d'infrastructure est soumis à des contraintes de capacité à différents niveaux (l'infrastructure elle-même ou l'accès à cette dernière, etc.), qui limitent, de fait, leur rôle attractif.

L'exemple de l'aéroport n'est pas anodin : il met en évidence le rôle particulièrement important des infrastructures de transport dans le processus d'agglomération, notamment en fournissant aux entreprises, comme l'indiquent Eberts et McMillen (1999), un facteur de production non payé. Les infrastructures publiques, en particulier de transport, sont généralement des biens communs et ont donc certaines des caractéristiques des biens privés, et, à ce titre, peuvent entrer dans la fonction de production des firmes. Ainsi, la présence de nombreuses entreprises dans une zone urbaine limite la capacité routière disponible pour chacune d'elles, si bien qu'une firme peut considérer, de son point de vue, qu'elle dispose d'une certaine quantité de cette capacité, qu'elle peut « employer » pour assurer sa production. L'investissement public dans les réseaux de transport permet de diminuer les coûts subis par les entreprises et les ménages, et même, en réduisant les niveaux de congestion, de limiter l'ampleur de cette source de déséconomies (Eberts et McMillen, 1999) que la section suivante présentera plus spécifiquement. De plus, l'existence d'un bon réseau de transport peut également améliorer la productivité des autres facteurs de production, en particulier le travail, par l'intermédiaire d'une baisse des temps de transport des employés. Autrement dit, les réseaux de transport peuvent être vus comme faisant partie intégrante de la technologie des entreprises (Jiwattanakulpaisarn, 2008).

Ainsi, Deno (1988) trouve, pour les États-Unis, que le réseau routier améliore la production des entreprises avec une élasticité de 0,31, ce qui indique un effet très important des infrastructures de transport (à comparer à 0,08 pour l'abduction d'eau dans la même étude). Toutefois, il est souvent difficile de différencier le rôle direct des transports comme facteur de production du rôle plus indirect de « facilitateur » des autres économies d'agglomération. Seitz (1995) fait à ce titre partie des quelques travaux contrôlant explicitement pour l'étendue des économies d'agglomération en utilisant le nombre d'emplois comme mesure. Il obtient, pour l'Allemagne, une élasticité du coût de production de $-0,127$ par rapport au capital d'infrastructures publiques. S'intéressant à la Grèce, Rovolis et Spence (2002) montrent que les infrastructures publiques se substituent partiellement aux autres facteurs de production comme le travail et sont plutôt complémentaires du capital privé.

Dans tous les cas, il semble que l'existence de réseaux de transport soit un gage

10. L'aéroport est un cas typique d'application du syndrome NIMBY : si la présence de cette infrastructure dans une ville peut être source d'avantage pour ses habitants, sa trop grande proximité est au contraire source de nuisances.

d'amélioration de la productivité des entreprises, ce qui constitue une source d'agglomération, puisque ce facteur n'est généralement que très indirectement payé par les firmes.

1.2.4.2. Centralité et accessibilité

Si les transports peuvent donc constituer un facteur de production gratuit pour les entreprises, ils sont aussi des catalyseurs des autres sources d'économies d'agglomération. En effet, l'ensemble des mécanismes que nous avons évoqués dans la section précédente, qu'ils agissent dans le cadre du marché ou sous forme d'externalités, repose sur des *interactions*. Or, ces interactions ne sont rendues possibles que grâce aux infrastructures de transport, qui permettent le face-à-face, et aux infrastructures de communication, qui permettent la proximité virtuelle.

Le fait que les villes constituent des nœuds de transport fait d'elles des centres. Or, agglomération et centralité ne sont *a priori* pas nécessairement associées, mais, comme l'indiquent Huriot et Bourdeau-Lepage (2009, chap. 1), « *l'interdépendance entre agglomération et réseaux de transport et de communication* » implique que « *les lieux d'agglomération maximale sont aussi les lieux d'accessibilité maximale.* »

Autrement dit, en tant qu'elle est un centre, la ville, d'une part, est le résultat des multiples interactions à l'origine des forces d'agglomération, et, d'autre part, favorise les interactions en mettant à la disposition des agents, entreprises et travailleurs, les infrastructures nécessaires. Toutefois, nous le verrons dans la section suivante, la congestion des systèmes de transports vient souvent nuancer ce constat, en cela qu'elle entraîne une diminution parfois très élevée de l'accessibilité potentielle offerte par un réseau de transport urbain.

1.3. La « tyrannie » de l'espace

Nous avons, dans les deux sections précédentes, apporté des réponses partielles aux deux premières questions mises en évidence dans l'introduction de ce chapitre, à savoir celles des causes de l'émergence et de la croissance des villes. Toutefois, la question de l'incarnation spatiale des villes, autrement dit du fait que les agents doivent nécessairement y trouver une place, n'a pas encore été abordée. Cette « tyrannie de l'espace », comme nous l'appellerons en référence à la tyrannie de la distance (Bairoch, 1985) qui en est une des composantes, permet de répondre aux questions suivantes qui font écho l'une à l'autre : pourquoi les villes s'étendent-elles dans l'espace ? Pourquoi ne recouvrent-elles pas tout l'espace ?

Nous montrerons dans cette section comment les forces d'agglomération peuvent engendrer en retour de puissantes forces de dispersion, liées à la rareté du sol (la tyrannie du sol) et aux coûts de transport (la tyrannie de la distance) ; la congestion

des réseaux (la tyrannie d'autrui) venant s'ajouter à ce panorama. Les formes de dispersion engendrées par ces forces se distinguent également par l'échelle à laquelle elles interviennent.

1.3.1. La tyrannie du sol : agglomération et pression foncière

La ville résulte des avantages liés à la proximité. Chaque agent, entreprise ou ménage, cherche à se localiser de manière à bénéficier de la plus grande proximité de ceux avec qui il est en interaction (Huriot et Bourdeau-Lepage, 2009, chap. 4). La raison en est que les économies d'agglomération se manifestent essentiellement à un niveau local, si bien qu'elles engendrent une concentration des agents limitée dans l'espace. Autrement dit, l'existence d'une force d'agglomération localisée entraîne une augmentation de la demande pour le sol, qui passe par une concurrence, d'autant plus intense que l'intérêt à la proximité est fort.

Par ailleurs, cet intérêt à la proximité, qui se traduit en une demande localisée et concentrée pour le sol, entraîne une hausse des prix parce que les agents consomment de l'espace lorsqu'ils se localisent. Par conséquent, lorsque la population d'une ville augmente, la demande de sol augmente également, poussant à la hausse l'ensemble des prix fonciers, quelle que soit la distance au centre.

Si le seul critère de localisation était la proximité du centre, tous les agents s'y localiseraient, avec une densité infiniment élevée. Ce n'est pas le cas, parce que chaque résident consomme de manière exclusive une certaine quantité d'espace. En dépit du fait qu'on assiste le plus souvent à un développement vertical des bâtiments, manière d'augmenter « artificiellement » la surface occupable¹¹, tous les ménages et toutes les entreprises ne peuvent pas se localiser au même point de l'espace, ils doivent se *dispenser*.

La tyrannie du sol, par l'augmentation de la pression foncière qu'elle induit, engendre une dispersion locale autour d'un centre.

1.3.2. La tyrannie de la distance : le coût du transport

Poussés par la tyrannie du sol, les agents acceptent donc de s'éloigner du centre. Ils renoncent ainsi en partie à la proximité pour pouvoir disposer de l'espace dont ils ont besoin au prix auquel ils peuvent l'acquérir. Mais renoncer à la proximité a des coûts, en particulier celui du transport. Ces coûts sont liés aux limites de performance des systèmes de transport et se décomposent en un coût temporel et un coût monétaire. Ces coûts proviennent du fait que les systèmes de transport possèdent un certain nombre de caractéristiques qui définissent le niveau de

11. Ce développement vertical a toutefois ses limites, tant technologiques qu'économiques (coûts).

performance qu'ils sont capables d'atteindre et les conditions auxquelles ce niveau peut être atteint. En particulier, la vitesse de déplacement le long de réseaux est fortement dépendante du système considéré (réseau routier destiné aux véhicules particuliers, réseau de bus, réseau en site propre de type métro, train, etc.). À ce titre, le temps passé en transport constitue donc une part importante du coût du transport. L'usage des systèmes de transport implique également un coût directement monétaire, lié aux conditions de possibilité du niveau de performance atteint : coût d'usage (carburant, entretien) d'un véhicule particulier, titre de transport ou abonnement de transport collectif.

Transport des biens agricoles jusqu'à la révolution industrielle, des biens industriels et des personnes ensuite, dans tous les cas, la croissance des coûts de transport lorsque la population de la ville s'accroît, et donc que la ville s'étend, est une des principales limites à la taille des villes, en particulier dans un cadre monocentrique.

Les mécanismes diffèrent, selon la nature de ce qui est transporté. Pour les biens agricoles, c'est une question d'approvisionnement des villes qu'il faut aller chercher toujours plus loin, ce qui devient vite impossible avec des technologies rudimentaires (Bairoch, 1985). Pour les biens industriels, c'est une question de débouchés : approvisionner l'espace rural peut devenir prohibitif (Huriot et Bourdeau-Lepage, 2009). Enfin, aujourd'hui, c'est le transport des personnes qui, engendrant des coûts liés aux déplacements pour le travail ou pour les achats et le loisir, limite majoritairement la taille des villes. La hausse des coûts de transport entraîne une hausse des prix fonciers et des salaires que doivent verser les entreprises, poussant ces dernières à quitter la ville devenue trop grande. Autrement dit, la tyrannie de la distance se traduit en tyrannie du sol.

Si ces deux forces de dispersion que sont le coût du sol et le coût de la distance sont liées, elles n'agissent pas à la même échelle. La tyrannie du sol pousse les agents à s'éloigner du centre, tout en y restant liés. C'est une force de dispersion intra-urbaine. La tyrannie de la distance, quant à elle, pousse les agents à quitter la ville (ou du moins désincite de nouveaux agents à venir s'y installer). C'est une force de dispersion inter-urbaine.

1.3.3. La tyrannie d'autrui : congestion des transports et pollution

Les coûts du transport subis par un ménage ne sont pas uniquement liés aux distances parcourues par celui-ci. Ils dépendent également du nombre des usagers du système de transport et de leur concentration locale (dans le temps et dans l'espace). La congestion des réseaux de transport, externalité liée à la présence simultanée, sur un même réseau caractérisé par une capacité limitée, d'un grand

nombre d'agents, constitue une puissante force de dispersion, car elle augmente, sans contrepartie complète, les coûts de transport.

Du fait que les usagers ne subissent pas la totalité du coût qu'ils font subir aux autres (cf. chapitre 2), ils font un usage des réseaux plus intense que s'ils devaient supporter entièrement ce coût, ce qui nuit à l'efficacité globale du système urbain. Si l'existence des forces d'agglomération que nous avons évoquées à la section précédente favorise la présence simultanée de nombreux agents, cet avantage peut se transformer en tyrannie, celle d'autrui, lorsque la population devient trop importante.

L'intensité de la congestion est également dépendante de la capacité des infrastructures, si bien qu'une augmentation de celle-ci pourrait apparaître comme une solution au problème. Toutefois, deux éléments conjoints viennent limiter cette possibilité. Tout d'abord, comme le suggère Mills (1967), augmenter la capacité signifie augmenter la surface allouée aux transports, ce qui implique une réduction de la surface dont bénéficie chaque ménage ou entreprise, et aboutit à renforcer la tyrannie du sol. Cela se traduit, pour reprendre l'exemple de Mills, par le fait qu'« *un doublement de la population d'une ville exige plus qu'un doublement de la surface allouée aux transports*¹² ». Par ailleurs, la « loi » de Downs (1962) postule que lorsque la capacité d'une route se voit augmentée, elle est rapidement comblée par de nouveaux usagers. Les résultats de Duranton et Turner (2008b) confirment cette intuition, et ces auteurs l'analysent comme la conséquence de trois phénomènes : lorsque les conditions de circulation s'améliorent, les individus se déplacent davantage, le transport routier de marchandise se développe, et *de nouveaux habitants* s'installent dans la ville.

D'autres effets externes liés à la concentration des agents, et donc à leur proximité spatiale, peuvent être regroupés sous le terme de tyrannie d'autrui. Il en va ainsi de la pollution et du bruit, nuisances généralement associées à la vie urbaine. Par ailleurs, l'intensification de la concurrence liée au rassemblement de nombreuses entreprises peut finir par pousser certaines d'entre elles à quitter un trop grand marché. C'est, il nous semble, une forme particulière de tyrannie d'autrui.

1.3.4. Tyrannie de l'espace et taille des villes

L'existence des trois composantes de ce que nous avons appelé la tyrannie de l'espace engendre une tension entre d'une part les économies externes liées à la concentration géographiques des entreprises et des ménages, et d'autre part les dés-économies associées aux grandes villes — prix du sol, coûts de transport, congestion, pollution, etc. (Fujita *et al.*, 2001).

12. Traduction de l'auteur.

La courbe classique d'Henderson (1974), reprise sur la Figure 1.1, rend compte de cet équilibre. Le point O donne la taille optimale de la ville N^* , autrement dit celle où l'utilité des résidents est maximale. Le point M , quant à lui, fournit la taille maximale de la ville N_m , au-delà de laquelle l'utilité des ménages est inférieure à ce qu'elle serait en l'absence de ville. En se rassemblant, entreprises et ménages commencent par s'approcher du point O , jusqu'à un point où chaque habitant supplémentaire entraîne une diminution de l'utilité de tous les résidents. Mais pour le nouvel arrivant, dont on suppose implicitement qu'il vient des zones rurales, la ville conserve un attrait tant que la ville n'a pas atteint le point M . En conséquence, les villes ont une tendance naturelle à devenir trop peuplée.

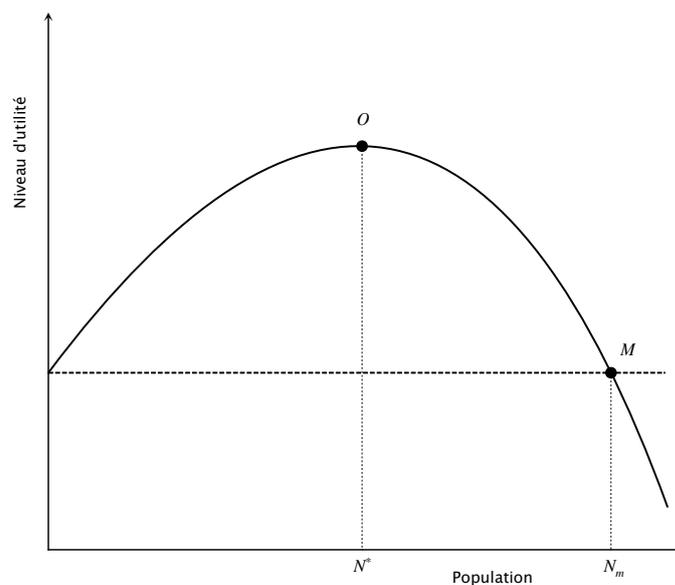


FIGURE 1.1.: Taille de la ville et niveau d'utilité. *Source* : Henderson (1974)

Toutefois, comme le notent Hurriot et Bourdeau-Lepage (2009), cette représentation schématique reste théorique ; déterminer de manière empirique la taille optimale d'une ville n'aurait pas véritablement de sens. En effet, une ville est en perpétuelle évolution, si bien que les points O et M de la figure précédente ne cessent de se déplacer avec les variations des forces en jeu, liées aux caractéristiques économiques, technologiques et institutionnelles.

Henderson (1974) considérait que si les économies d'agglomération se différencient selon le type d'industrie à leur source, les déséconomies ne dépendaient que de la taille de la ville. Or, il est probable que des facteurs tels que l'organisation des systèmes de transport et la localisation relative des emplois et des ménages jouent un rôle dans l'intensité des forces de dispersion.

Les manifestations et les conséquences de la tyrannie de l'espace sont au cœur

de nos préoccupations dans ce mémoire. Tout d'abord, les trois composantes que nous avons évoquées sont, comme nous avons pu le constater, très liées les unes aux autres, et sculptent la structure interne des villes. Les arbitrages qu'elles imposent aux agents, ménages ou entreprises, sont essentiels pour comprendre l'organisation des villes. Ensuite, leur caractère partiellement externe les rend propices à soulever des questions d'équité, en particulier spatiale, thème que nous aborderons à diverses reprises dans ce mémoire.

1.4. L'incarnation des arbitrages de localisation : le modèle urbain monocentrique

Les différentes formes de la tyrannie de l'espace sont liées entre elles. Elles poussent les agents économiques urbains à réaliser des arbitrages. Ainsi, la tyrannie du sol les incite à s'éloigner du centre, tandis que la tyrannie de la distance limite cet éloignement. À l'inverse, c'est la tyrannie de la distance qui engendre des prix élevés du sol et constitue donc la source majeure de la tyrannie du sol.

L'objectif de cette section est de détailler ces arbitrages à partir d'une modélisation urbaine monocentrique. En effet, ce cadre de modélisation nous apparaît idéal pour les mettre en scène, puisqu'il permet de faire émerger de l'ensemble des arbitrages un équilibre à l'échelle urbaine. De plus, il met en évidence certaines conséquences macroscopiques de ces arbitrages, en particulier les questions de ségrégation, tout en offrant la possibilité d'analyser la question du coût de transport et de la congestion.

Les hypothèses du modèle, semblables à celles du modèle initial de Von Thünen (1826), contournent le théorème d'impossibilité spatiale en supposant l'existence, au sein d'une plaine homogène, d'un point central, regroupant l'ensemble des emplois, le CBD. Autrement dit, le modèle monocentrique, dans sa version « basique » (Alonso, 1964 ; Mills, 1967, 1972 ; Muth, 1969), ne concerne que la localisation résidentielle des ménages. Le modèle a été toutefois adapté à la localisation des entreprises (voir par exemple Costes, 2008).

Les ménages se localisent donc dans une couronne résidentielle autour du CBD, où ils doivent se rendre chaque jour pour travailler et faire leurs achats. Ils disposent tous du même revenu Y et sont homogènes en préférences. Ils tirent leur utilité de la consommation en quantité z d'un bien composite pris comme numéraire, et de la surface de leur logement s , par l'intermédiaire d'une fonction d'utilité identique pour tous, $U(z, s)$, supposée strictement croissante et de classe \mathcal{C}^2 en z et s .

Pour se rendre au CBD, ils empruntent un réseau de transport supposé dense, qui leur permet de rejoindre le CBD en suivant une ligne droite à partir de leur

résidence. Le coût de transport, qu'on suppose ne dépendre que de la distance au centre r , est fourni par la fonction $T(r)$, vérifiant $0 \leq T(0) < Y < \lim_{r \rightarrow +\infty} T(r)$.

Le prix du sol est fourni par une courbe de rente foncière $R(r)$, tandis que son coût d'opportunité vaut R_A , la rente « agricole ».

Ces hypothèses suffisent d'une part à définir la localisation d'équilibre d'un ménage, et de mettre ainsi en évidence les arbitrages qu'il réalise, d'autre part à obtenir l'équilibre urbain, autrement dit à déterminer la courbe de rente foncière endogène. Notre présentation, dans la suite de cette section, s'appuie sur celle de Fujita (1989).

1.4.1. La rente foncière : une synthèse de l'arbitrage des agents entre coût du sol et coût du transport

Nous commencerons par détailler le programme d'optimisation auquel font face les ménages et nous montrerons que les courbes d'enchère foncière incarnent leurs arbitrages. Ensuite, nous verrons que la rencontre de l'ensemble de ces courbes d'enchère définit un équilibre urbain, à même de faire émerger une structure urbaine traduisant, de façon macroscopique, les différents arbitrages de l'ensemble des ménages.

1.4.1.1. Les arbitrages du ménage

Le programme du ménage s'écrit :

$$\max_{z,s,r} U(z,s) \quad \text{s.c.} \quad z + R(r)s = Y - T(r). \quad (1.2)$$

Ce programme peut être transformé en un autre, qui met en évidence la capacité des ménages à payer le coût foncier afin de préparer la discussion des arbitrages opérés par les ménages. On définit ainsi la rente d'enchère foncière $\Psi(r, u)$ comme le prix maximal qu'un ménage peut payer pour résider en r tout en jouissant d'un niveau d'utilité u . Plus précisément :

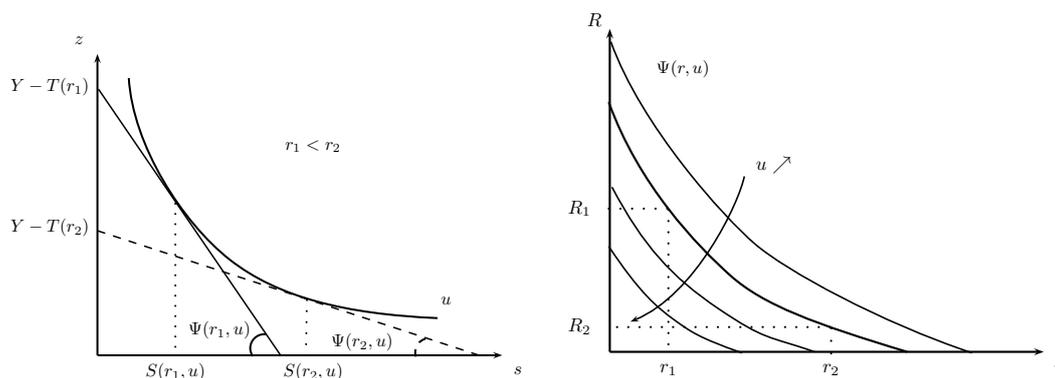
$$\Psi(r, u) = \max_{z,s} \left\{ \frac{Y - T(r) - z}{s} \mid U(z, s) = u \right\}. \quad (1.3)$$

Cette équation peut être réécrite de manière plus compacte, en considérant la solution, en z , de la relation $U(z, s) = u$, à savoir $z = Z(s, u)$:

$$\Psi(r, u) = \max_s \frac{Y - T(r) - Z(s, u)}{s}. \quad (1.4)$$

Deux courbes d'enchère correspondant à deux niveaux d'utilité distincts ne se

coupent pas¹³. Dans un espace (r, R) que l'on peut appeler *espace urbain*, « dual » de l'espace de consommations (z, s) , les courbes d'enchère sont en réalité des courbes d'indifférence. Sur la Figure 1.2(b), un ménage retire la même utilité en résidant en r_1 et en payant une rente foncière R_1 qu'en résidant en r_2 et en payant une rente foncière R_2 . Autrement dit, pour atteindre un niveau d'utilité donné, les ménages ont la possibilité d'arbitrer entre les différentes localisations possibles, en modifiant en conséquence la somme qu'ils sont prêts à payer pour se loger. Plus précisément, les ménages situés plus loin du CBD ont une disposition à payer pour le sol plus faible. De plus, comme l'indique la Figure 1.2(a), la surface des logements augmente avec la distance au centre.



(a) Courbe d'indifférence, contrainte budgétaire et enchère foncière (b) Courbes d'enchère et utilité du ménage

FIGURE 1.2.: Les courbes d'enchère foncière sont des courbes d'indifférence. *Source* : Fujita (1989)

Nous supposons dans un premier temps que la courbe de rente foncière est exogène, si bien qu'un ménage choisit sa localisation en la considérant comme donnée. La résolution du programme du ménage s'effectue de la manière suivante : pour atteindre le niveau d'utilité maximal tout en trouvant un lieu de résidence dans la ville, un ménage doit choisir un couple (r^*, R^*) se situant à la fois sur la courbe $R(r)$ et sur la courbe $\Psi(r, u)$ offrant le meilleur niveau d'utilité, donc la plus « basse » possible sur la Figure 1.2(b). Par conséquent, il s'agit du point de tangence, « par dessous », entre la courbe $R(r)$ et une courbe d'enchère $\Psi(r, u)$. Cette règle se formalise donc :

$$R(r^*) = \Psi(r^*, u^*) \quad \text{et} \quad R(r) \geq \Psi(r, u^*), \quad \forall r. \quad (1.5)$$

13. Cette propriété se démontre par l'absurde. Soient u_1 et u_2 deux niveaux distincts d'utilité. On suppose qu'il existe un point (\tilde{r}, \tilde{R}) tel que $\tilde{R} = \Psi(\tilde{r}, u_1) = \Psi(\tilde{r}, u_2)$. Cela signifie qu'un ménage résidant en \tilde{r} et payant une rente \tilde{R} atteint à la fois les niveaux d'utilité u_1 et u_2 , ce qui est impossible.

où u^* est le niveau d'utilité atteint par le ménage, et r^* sa localisation d'équilibre.

La condition de Muth (1969), conséquence directe de cette règle de tangence entre les courbes de rente et d'enchère, matérialise l'arbitrage des ménages entre prix du sol et coût du transport :

Proposition 1.1 (Condition de Muth). *À la localisation d'équilibre d'un ménage, coût du transport et prix du sol sont reliés par la condition d'arbitrage suivante :*

$$T'(r^*) = -R'(r^*)S(r^*, u^*). \quad (1.6)$$

Démonstration. Le théorème de l'enveloppe, appliqué à (1.4), fournit une relation valable en tout point r , donc en particulier à r^* :

$$\frac{\partial \Psi(r, u)}{\partial r} = -\frac{T'(r)}{S(r, u)}.$$

Par ailleurs, la condition de tangence au point d'équilibre s'exprime par :

$$\frac{\partial \Psi(r^*, u^*)}{\partial r} = R'(r^*).$$

En combinant ces deux dernières relations, on obtient la condition de Muth. \square

Cette condition d'équilibre spatial exprime le fait qu'à la localisation d'équilibre du ménage, le coût marginal de transport qu'il subirait en s'éloignant $T'(r^*)$ est égal au gain marginal sur le coût du sol, $-R'(r^*)S(r^*, u^*)$. Si cela n'était pas le cas, le ménage pourrait en effet atteindre un niveau d'utilité supérieur soit en s'éloignant du CBD si $T'(r) < -R'(r^*)S(r^*, u^*)$, soit en s'en rapprochant si $T'(r) > -R'(r^*)S(r^*, u^*)$.

La condition de Muth peut être étendue à une situation dans laquelle le CBD n'est pas le seul pôle attractif de la ville, et où des aménités apportent également une utilité aux ménages (Brueckner *et al.*, 1999). Plus précisément, en supposant que ces aménités sont distribuées selon une densité $a(r)$, et que la fonction d'utilité dépend de z , s et a , la condition devient :

$$-R'(r^*)S(r^*, u^*) = T'(r^*) - \frac{\partial V}{\partial a}(Y - T(r^*), R(r^*), a(r^*))a'(r^*), \quad (1.7)$$

où V est la fonction d'utilité indirecte. $\partial V/\partial a$ fournit la valeur marginale de l'aménité, après ajustement optimal de la surface de logement par le ménage. L'arbitrage réalisé par le ménage est donc dans ce cas plus complexe, puisqu'il fait intervenir le prix du sol d'un côté, et de l'autre à la fois le coût du transport et la proximité des

aménités. Autrement dit, les ménages sont prêts à supporter un coût de transport plus élevé en échange d'une plus grande densité d'aménités.

Dans les faits, en Île-de-France, Rouchaud et Sauvart (2004) ont mis en évidence, à partir du calibrage d'un modèle monocentrique sur la région parisienne, l'existence d'un arbitrage entre prix du sol et coût du transport, qui se traduit par le fait que « les valeurs du temps révélées par les comportements en transport et celles révélées par les prix des logements sont globalement cohérentes ». Coulombel *et al.* (2007) trouvent que la somme des budgets transport et logement représente une part relativement constante du revenu des ménages, à savoir 44 %, ce qui milite en faveur de l'existence d'un arbitrage. Toutefois, ces mêmes auteurs, ainsi que Polacchini et Orfeuil (1999) nuancent ce résultat en montrant que cette réalité agrégée cache une grande diversité liée aux différences de revenus, et qu'en fait, le budget logement a tendance à rester constant, le budget transport augmentant avec la distance.

1.4.1.2. Concurrence pour le sol, équilibre et ségrégation

Nous venons de le voir, dans un marché du logement existant, un ménage choisit sa localisation d'équilibre en confrontant ses capacités d'enchère, sous la forme de ses courbes d'enchère foncière, avec la réalité du marché, à savoir la courbe de rente foncière. En réalité, la courbe de rente est le résultat de la rencontre de l'ensemble des enchères des ménages. Plus précisément, les ménages sont en concurrence pour l'usage du sol et l'organisation de l'espace urbain monocentrique est le résultat de cette concurrence. L'équilibre urbain, c'est-à-dire une situation dans laquelle aucun ménage n'a intérêt à se délocaliser, émerge de la confrontation entre elles de l'ensemble des fonctions d'enchère foncière des ménages.

Dans le cas, que nous avons considéré jusqu'ici, où les ménages sont tous identiques, ils atteignent nécessairement le même niveau d'utilité u^* à l'équilibre. Si la ville est *fermée*, c'est-à-dire que sa population est fixée à N , l'équilibre est caractérisé par un certain nombre de conditions nécessaires et suffisantes qui permettent de définir la rente foncière d'équilibre $R^*(r)$, la courbe des surfaces des logements $S^*(r)$ et celle des densités radiales $n^*(r)$, ainsi que la limite de la ville r_f et le niveau

d'utilité u^* :

$$\begin{aligned}\Psi(r_f, u^*) &= R_A, \\ R^*(r) &= \begin{cases} \Psi(r, u^*) & \text{pour } r \leq r_f \\ R_A & \text{pour } r \geq r_f \end{cases}, \\ S^*(r) &= S(r, u^*) \quad \text{pour } r \leq r_f, \\ n^*(r) &= \begin{cases} L(r)/S^*(r) & \text{pour } r \leq r_f \\ 0 & \text{pour } r > r_f \end{cases}, \\ \int_0^{r_f} n^*(r) dr &= N,\end{aligned}$$

où $L(r)$ est la distribution de l'espace dans la ville, c'est-à-dire que $L(r)dr$ est la surface totale offerte entre r et $r + dr$. Ces conditions fournissent les principales propriétés de l'équilibre : les prix du sol sont décroissants avec la distance au centre, tandis que la surface des logements augmente. Par conséquent, la densité diminue avec r . Enfin, la relation de Muth est valable *pour toutes les localisations urbaines*. Ce dernier point est à souligner, car il montre qu'à l'équilibre l'ensemble des ménages est soumis à la même condition d'arbitrage.

De plus, en termes agrégés, la dépense totale de transport DTT et la rente différentielle totale RDT sont reliées par (Arnott et Stiglitz, 1981) :

$$DTT \leq \delta RDT \quad \text{si} \quad \forall r, \quad \frac{T(r)L(r)}{T'(r)\mathcal{L}(r)} \leq \delta, \quad (1.8)$$

où $\mathcal{L}(r) = \int_0^r L(x)dx$. Par conséquent, pour une ville linéaire et un coût de transport linéaire, dépense totale de transport et rente différentielle totale sont égales.

D'un point de vue technique, on distingue classiquement quatre hypothèses de modélisation monocentrique. Ainsi, la distinction *ville ouverte / ville fermée* déjà évoquée permet de rendre compte des différences d'intégration des villes dans le contexte économique global d'un pays : les modèles de ville ouverte décrivent mieux les agglomérations s'inscrivant dans un contexte où le niveau d'utilité du pays est fixé par les habitants des zones rurales, tandis que les modèles de ville fermée est plus adaptée au contexte des grandes villes des pays industrialisés. Par ailleurs, on distingue les modèles où les propriétaires des logements sont *absents*, autrement dit où la rente qu'ils collectent quitte la ville et n'entre pas dans le revenu des ménages, et les modèles où s'exerce une propriété publique du sol, impliquant que la rente différentielle totale est redistribuée aux ménages, sous la forme d'une somme forfaitaire. Les propriétés que nous avons présentées ne dépendent pas de ces choix de modélisation, mais les conditions de l'équilibre varient. Nous nous restreindrons, dans la suite de cette présentation, au cas des modèles fermés à propriétaires absents.

Le mécanisme micro-économique d'enchères, à l'origine de l'équilibre urbain, donne par ailleurs lieu à un phénomène de ségrégation spatiale, qui, à la différence de ceux mis en évidence par Schelling (1969), émergent sans qu'une préférence pour l'*entre-soi* n'entre en jeu. La seule concurrence pour le sol, alliée à des dispositions à payer différentes suffit. Fujita (1989) montre ainsi que si les courbes d'enchère d'un ménage i sont plus *pentues*¹⁴ que celles d'un ménage j , alors la localisation d'équilibre du ménage i est plus proche du centre que celle du ménage j .

Cette propriété, très utile, permet d'analyser l'influence des différentes caractéristiques des ménages¹⁵. Ainsi, toutes choses égales par ailleurs, en particulier le coût de transport, le ménage au revenu le plus élevé se localise plus loin du centre que le ménage au revenu plus faible. Toutefois, si le coût unitaire du transport augmente fortement avec le revenu, par l'intermédiaire d'une valeur du temps croissant avec le revenu notamment, le schéma peut s'inverser (Wheaton, 1977)¹⁶. La présence d'aménités, comme nous l'avons évoqué précédemment, peut avoir le même résultat. Dans tous les cas, cependant, la confrontation des courbes d'enchère de ménages différents conduit à un équilibre ségrégué, semblable aux fameux anneaux de Von Thünen (1826). Marshall (1890) évoque également ce phénomène, en l'étendant aux entreprises :

« La valeur que les quartiers centraux d'une grande ville possèdent pour les commerçants permet à ceux-ci d'y payer le sol bien plus cher qu'il ne vaut pour des fabriques [...] et une compétition semblable, au sujet du logement, a lieu entre les employés des maisons de commerce et les ouvriers de fabrique. Le résultat est que, maintenant, les fabriques se groupent dans les faubourgs des grandes villes et dans les régions manufacturières avoisinantes, plutôt que dans les villes elles-mêmes. »
(Marshall, 1890, chap. 10)

Le fait que les ménages aient des caractéristiques différentes induit que les résultats de leurs arbitrages, pourtant réalisés sur les mêmes bases, diffèrent. Dans ce contexte à plusieurs types de ménages, les niveaux d'utilité à l'équilibre diffèrent selon le type de ménages, et la courbe de rente foncière est l'enveloppe supérieure des courbes d'enchère des différents types, comme illustré sur la Figure 1.3.

14. Les courbes d'enchère d'un ménage i sont dites « plus pentues » que celles d'un ménage j si l'on peut trouver deux courbes $\Psi_i(r, u_1)$ et $\Psi_j(r, u_2)$ qui se coupent et telles qu'en leur point d'intersection Ψ_i soit supérieure à Ψ_j à gauche et inférieure à droite. En effet, les courbes d'enchère d'un même ménage ne se coupant pas, si l'une d'entre elles est plus pentue que l'une de celles d'un autre ménage, alors elles le sont toutes.

15. Plus précisément, l'influence d'un paramètre θ sur la pente des courbes d'enchère s'obtient en étudiant le signe de $-(\partial\Psi_r/\partial\theta)|_{d\Psi=0}$.

16. Supposons que $\partial T/\partial Y > 0$, alors $-(\partial\Psi_r/\partial Y)|_{d\Psi=0} = (1/SY)(\partial T/\partial r)(\epsilon_{T,Y} - \epsilon_{S,Y})$, où $\epsilon_{T,Y}$ et $\epsilon_{S,Y}$ sont les élasticités-revenu du coût de transport et de la consommation d'espace, respectivement. Par conséquent, si l'élasticité-revenu du coût de transport est supérieure à l'élasticité-revenu

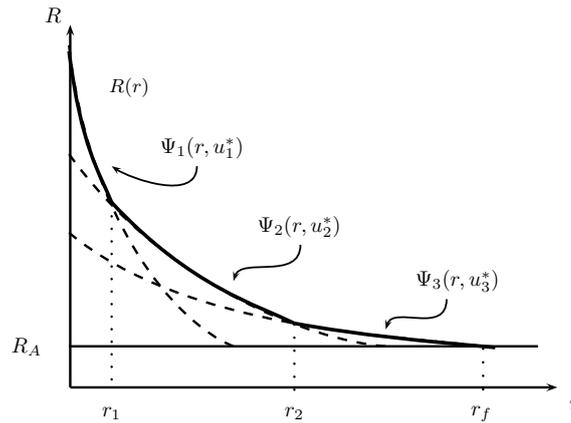


FIGURE 1.3.: La courbe de rente foncière est l'enveloppe supérieure des courbes d'enchère. *Source* : Fujita (1989)

1.4.2. Coût de transport, congestion et étalement

Le transport, et plus particulièrement son coût pour les ménages, joue un rôle fondamental dans l'organisation urbaine interne telle qu'elle est représentée dans le modèle monocentrique, puisqu'il détermine en grande partie les choix de localisation et les prix du logement. Nous souhaitons ici souligner cet aspect, central dans toute la suite de ce mémoire. Plus précisément, une analyse en statique comparative (Wheaton, 1974) nous permettra de mettre en évidence les conséquences d'une modification du coût du transport, dans l'absolu et de manière relative au revenu, que complètera l'introduction d'une forme de congestion.

On compare deux situations correspondant au passage d'une fonction de coût de transport $T_a(r)$ à une fonction $T_b(r)$ telles que :

$$T_a(0) = T_b(0) \quad \text{et} \quad T'_a(r) > T'_b(r) \quad \forall r \quad (1.9)$$

Autrement dit, le coût du transport dans la situation b est plus faible que dans la situation a . Cela se traduit, pour le revenu net I , par :

$$I_a(r) = Y - T_a(r) < Y - T_b(r) = I_b(r) \quad \forall r > 0 \quad (1.10)$$

En conséquence, les courbes d'enchère Ψ_a des ménages dans la situation a sont plus pentues que les courbes Ψ_b ¹⁷. Une baisse du coût de transport engendre donc une

de la consommation d'espace, un haut revenu favorise la localisation centrale.

17. Cette propriété se démontre en choisissant deux représentants $\Psi_a(r, u_a)$ et $\Psi_b(r, u_b)$ qui se coupent en x : $\Psi_a(x, u_a) = \Psi_b(x, u_b) = \bar{R}$. Le sol étant un bien normal, sa consommation par les ménages augmente avec le revenu, si bien que : $S_a(x, u_a) = \hat{s}(\bar{R}, I_a(x)) \leq \hat{s}(\bar{R}, I_b(x)) = S_b(x, u_b)$, \hat{s} étant la demande Marshallienne. Finalement : $-\partial\Psi_a(x, u_a)/\partial r = T'_a(x)/S_a(x, u_a) >$

décroissance plus faible du prix des logements avec la distance au centre. De plus, l'utilité des ménages à l'équilibre s'améliore avec la baisse des coûts de transport, car leur revenu net augmente partout, si bien que $u_a^* < u_b^*$, et la croissance du revenu net augmente la demande d'espace, si bien que la limite de la ville est repoussée plus loin¹⁸ : $r_f^a < r_f^b$. Finalement, le prix du sol au centre de la ville diminue lorsque le coût du transport diminue¹⁹, et la surface de logement augmente à proximité du centre.

En résumé, une diminution du coût de transport rend les terrains en périphérie relativement moins désavantageux par rapport aux terrains proches du centre, si bien que certains ménages s'éloignent du centre, diminuant la pression foncière à proximité du CBD, tout en l'accroissant en périphérie. La courbe de rente foncière s'aplatit et le gradient de densité diminue. On assiste à un étalement urbain semblable à celui constaté par Clark (1968).

Une augmentation du revenu conduit à des conséquences similaires, mais introduit des nuances. Ainsi, en supposant que l'augmentation du revenu ne modifie pas les comportements de déplacement des ménages, l'évolution de la rente foncière dépend de la forme de la fonction de coût de transport $T(r)$ et de la distribution de l'espace offert $L(r)$. Dans la situation la plus probable toutefois, où $L(r)$ augmente plus vite que $T'(r)$, une augmentation du revenu mène aux mêmes résultats qu'une baisse du coût de transport. En revanche, si le coût de transport évolue avec le revenu, en particulier si le coût d'opportunité du temps s'accroît avec le revenu, les conclusions sont moins claires, et on peut assister, en théorie, à une réduction de l'emprise urbaine. Nous reviendrons sur ces questions d'étalement urbain au chapitre 8.

Nous l'avons déjà évoqué, les systèmes de transport sont soumis au phénomène de congestion, dont le coût, en particulier temporel, vient s'ajouter au coût de transport nominal, c'est-à-dire sans congestion. Le fait que la congestion soit une externalité conduit à des choix de localisation sous-optimaux de la part des ménages (Fujita, 1989, chap. 7). Plus précisément, Solow (1973a), Arnott et MacKinnon (1978), Sullivan (1983a,c), parmi d'autres encore, ont montré que la congestion des transports entraîne un trop grand étalement de la ville par rapport à une situation optimale, parce que les usagers ne prennent pas en compte les coûts sociaux qu'ils font peser sur les autres usagers lorsque la distance qu'ils parcourent augmente (Denant-Boèmont et Goffette-Nagot, 2002). Ils ne prennent en considération que le coût marginal *privé*.

$$T'_b(x)/S_b(x, u_b) = -\partial\Psi_b(x, u_b)/\partial r.$$

18. Les preuves de ces deux propriétés figurent dans Fujita (1989, pp. 328 et 79, respectivement).

19. La décroissance de la fonction d'utilité indirecte avec le prix du sol donne : $V(\Psi_a(0, u_a^*), I_a(0)) = u_a^* < u_b^* = V(\Psi_b(0, u_b^*), I_b(0))$, et donc : $R_a(0) = \Psi_a(0, u_a^*) > \Psi_b(0, u_b^*) = R_b(0)$.

Nous avons vu, dans les paragraphes précédents, que la baisse du coût privé de transport était de nature à favoriser un étalement urbain. L'existence de la congestion introduit une inefficacité liée à cet étalement : dans le cadre monocentrique classique, la baisse du coût de transport est de nature à accentuer l'écart entre configuration d'équilibre et configuration optimale, car elle incite les ménages à parcourir des distances plus grandes.

Des politiques de taxation correctrice peuvent être mises en place. Ainsi, une taxe basée sur la localisation des ménages permet d'atteindre un optimum de second rang (Denant-Boëmont et Goffette-Nagot, 2002), de même que l'instauration de cordons de péage (De Palma *et al.*, 2008). Toutefois, Arnott (2007) montre, à partir d'un modèle à deux « îles » reliées par un pont congestible, qu'en présence d'économies d'agglomération liées à la présence simultanée des travailleurs sur leur lieu de travail, taxer la congestion sans connaître l'ampleur des économies d'agglomération peut conduire à une situation sous-optimale pire qu'en l'absence de péage.

1.4.3. Les limites du modèle

Grâce à sa grande simplicité conceptuelle et sa résolution analytique, le modèle monocentrique est la pierre angulaire de l'économie urbaine. Toutefois, il présente un certain nombre de limites, liées en particulier à ses hypothèses simplificatrices. Nous avons choisi d'en évoquer deux, qui nous apparaissent comme les plus fondamentales.

1.4.3.1. Monocentrique ou polycentrique ?

L'une des hypothèses centrales du modèle monocentrique est la concentration de l'ensemble des emplois en un centre ponctuel unique. Or, comme le relèvent Anas *et al.* (1998), l'un des aspects du développement urbain moderne est la tendance de l'activité économique à s'agglomérer en plusieurs centres. Plus précisément, les forces de dispersion que nous avons évoquées, en particulier la tyrannie du sol, poussent les entreprises à rechercher des lieux d'implantation plus éloignés des centres urbains, afin de bénéficier de prix fonciers plus faibles. Dans le même temps, les forces d'agglomération jouent toujours, si bien qu'on assiste à un regroupement de l'activité en de multiples centres. Le modèle monocentrique, dans sa version de base, ne rend pas compte de ce mouvement, pourtant marqué, surtout aux États-Unis. Cependant, des extensions du modèle de base permettent de représenter aussi bien la décentralisation de l'emploi, c'est-à-dire son étalement autour d'un centre, que sa déconcentration, autrement dit son regroupement en plusieurs centres (Goffette-Nagot, 2000). Le chapitre 6 détaillera la question de la décentralisation des emplois qui, nous le verrons, prend très facilement sa place comme

extension du modèle monocentrique. Le modèle monocentrique, dans la version « ménages » telle que nous venons de le présenter, peut être adapté pour rendre compte des choix de localisation des entreprises. Formellement, cela signifie que la fonction de profit des entreprises remplace la fonction d'utilité des ménages. Le modèle de Von Thünen (1826) s'intéressait déjà à cette question de localisation de la production, et beaucoup plus récemment, Cooke (1983) ou Costes (2008), par exemple, proposent un modèle de localisation des entreprises, prenant en compte, pour le second, plusieurs types d'entreprises.

La question de la déconcentration des emplois et des ménages peut également être analysée à l'aide de modèles issus du modèle monocentrique. Ainsi, le modèle Fujita-Imai-Ogawa, ou FIO (Ogawa et Fujita, 1980 ; Fujita et Ogawa, 1982 ; Imai, 1982) modélise les interactions entre firmes sous la forme d'échanges d'information (cf. 1.2.3). Les auteurs obtiennent différentes structures internes, certaines étant monocentriques avec emploi partiellement dispersé, d'autres de nature polycentrique, plusieurs centres d'activité émergeant. Plus précisément, lorsque le profit que les firmes tirent de l'échange d'information décroît linéairement avec la distance, les villes sont monocentriques, avec des degrés divers de mixité entre emplois et résidences. En revanche, lorsque le profit varie de manière exponentielle décroissante avec la distance, les villes sont monocentriques, duocentriques ou tricentriques, selon l'intensité relative de la décroissance du profit et des coûts de transport, comme l'illustre la Figure 1.4.

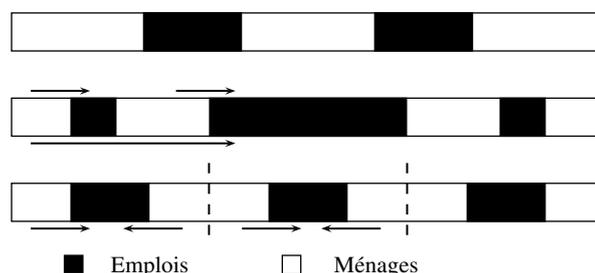


FIGURE 1.4.: Configurations urbaines avec une décroissance exponentielle. *Source* : Huriot et Bourdeau-Lepage (2009)

De plus, certaines de ces configurations montrent un fonctionnement indépendant des différents centres, tandis que d'autres mettent en évidence des centres majeurs, attirant des travailleurs venant de l'ensemble de la zone urbaine. Ota et Fujita (1993) complètent encore l'analyse en supposant que les firmes sont constituées d'une unité de coordination (*front office*), qui a besoin de nombreuses interactions avec les autres unités de coordination, et d'une unité d'exécution (*back office*), qui communique avec son unité de coordination par le biais des technologies de l'information et de la communication (TIC). Ils obtiennent un grand nombre de

configurations et montrent que sous certaines conditions (faible coût des TIC, fort besoin de concentration des unités de coordination qui engendre un niveau élevé des prix fonciers), unités de coordination et d'exécution se séparent, les premières restant au centre, tandis que les secondes s'installent en périphérie.

Ces éléments montrent les limites du modèle monocentrique à représenter l'ensemble du spectre des configurations urbaines. Toutefois, comme le notent Huriot et Bourdeau-Lepage (2009), les villes françaises, à la différence des villes américaines, se caractérisent par une organisation qu'ils nomment *multipolaire-monocentrique*. Autrement dit, si l'existence de multiples pôles est attestée, ils sont généralement organisés selon une structure monocentrique, les pôles les plus importants se situant à proximité immédiate du pôle principal. C'est notamment le cas pour l'Île-de-France, où la croissance des pôles d'emplois est supérieure dans les pôles proches de la capitale que dans ceux qui en sont plus éloignés (Guillain *et al.*, 2006). De même, Riguelle *et al.* (2007) montrent, en s'appuyant sur le cas belge, qu'en dépit d'une tendance générale à la déconcentration, le polycentrisme demeure, en Europe, un phénomène relativement limité. Selon nous, le modèle monocentrique reste apte à représenter, de manière au moins stylisée, la structure urbaine de l'Île-de-France. Les chapitres 6 et 7 viseront, entre autres, à étayer ce point de vue, et à définir ses conditions d'application.

1.4.3.2. Les ménages cherchent-ils à minimiser les coûts de transport ?

Une autre hypothèse centrale du modèle monocentrique est celle de la minimisation, par les ménages, des coûts de déplacement domicile-travail, sous contrainte de satisfaction de leur utilité. Si les contraintes budgétaires, monétaires et temporelles, s'imposent naturellement aux ménages, la question de savoir si le coût du transport pendulaire est un critère central de choix n'est pas tranchée. L'imposante littérature sur ce qui a été appelé par Hamilton (1982) le « *wasteful commuting* » le prouve, la capacité du modèle monocentrique à rendre compte de l'organisation des déplacements des ménages est limitée. Comme le notent Mills et Hamilton (1994), l'écart entre la distance moyenne des déplacements domicile-travail constatée dans une agglomération réelle, d'une part, et obtenue à partir d'un modèle supposant la minimisation des distances domicile-travail calibré sur cette même agglomération, d'autre part, est significatif. Il varie en effet entre 11 % (White, 1988a) et 87 % (Hamilton, 1982). Le même type de résultat est obtenu en prenant en considération la durée du déplacement au lieu de la distance.

Ces résultats peuvent être attribués à deux éléments (Charron, 2007). Tout d'abord, l'auteur évoque un élément morphologique, lié au fait que les réseaux de transport réels ne sont pas denses comme le suppose le modèle monocentrique,

et que résidences et emplois peuvent être distribués de manière complexe dans l'espace. Mais Small et Song (1992) tenant compte de cet aspect, trouvent un écart de distance encore significatif. Toutefois, l'hétérogénéité des ménages et des emplois, l'information imparfaite dont disposent les ménages, ainsi que le fait que de nombreux ménages sont biactifs compliquent l'analyse (Giuliano et Small, 1993). Ensuite, il évoque un élément comportemental, correspondant à la *tolérance* des ménages à la distance (ou à la durée du déplacement) domicile-travail. Si cet élément est prépondérant, l'*excess commuting* a de bonnes chances d'être élevé.

Horner (2002) et Charron (2007) constatent néanmoins que la distance moyenne observée est beaucoup plus proche de celle issue d'une minimisation des distances que de la distance moyenne maximale ainsi que de la distance moyenne issue d'un appariement aléatoire des ménages aux emplois. En d'autres termes, cela signifie que les ménages accordent plus attention à la distance domicile-travail qu'il pourrait sembler, réhabilitant par la même en partie, selon nous, le modèle monocentrique et ses extensions dans leur capacité à analyser les déplacements domicile-travail. C'est pourquoi nous mobiliserons ce modèle, en l'étendant, dans la deuxième partie de ce mémoire, et que nous nous appuierons sur ses propriétés pour étudier cette question des déplacements domicile-travail.

1.5. Conclusion

L'organisation urbaine est le résultat d'une interaction complexe entre forces d'agglomération et forces de dispersion, à plusieurs échelles et à travers de multiples mécanismes. Par leur capacité à *relier*, les transports facilitent la réalisation des économies d'agglomération et favorisent donc la croissance urbaine. Toutefois, le transport a un coût, accentué par le fait que les réseaux de transport sont sujets à la congestion, et génère donc également des forces de dispersion. De même, le besoin de proximité induit une concurrence pour le sol qui pousse à la hausse les prix fonciers.

Pris dans cet écheveau de forces opposées, entre tyrannie du sol et tyrannie de la distance, les ménages sont amenés à faire des arbitrages entre coût du sol et coût du transport. Le modèle monocentrique permet, à partir d'hypothèses simples, de mettre en scène ces arbitrages et d'aboutir à un équilibre urbain à la structure particulière. Nous avons mis en avant les apports de ce modèle, en particulier du fait de sa capacité à expliquer certains phénomènes urbains comme la ségrégation ou l'étalement urbain. Nous avons également abordé certaines de ses limites, en nous efforçant néanmoins de souligner les domaines de validité du modèle, qui constitue, aujourd'hui encore, un outil de choix pour l'économie urbaine.

Il ressort de cette analyse que les arbitrages, réalisés par les ménages mais égale-

ment par les entreprises, entre des forces qui poussent à plus de proximité — économies d'agglomération, minimisation des coûts de transport, etc. — et des forces qui tendent au contraire à la dispersion — prix fonciers, concurrence, préférence pour l'espace, congestion, etc. — sont à la source de la structure urbaine. Ces arbitrages interviennent à diverses échelles, au niveau de l'entreprise comme à celle d'un pôle d'activité, voire d'une région dans son ensemble. Le rôle des transports, comme support des interactions ou comme frein à celles-ci, nous semble essentiel. Si nous avons, dans ce chapitre, évoqué la question de la congestion des transports, nous n'avons pu lui donner, pour l'instant, la place qu'elle mérite dans l'analyse de la structure urbaine. Notre mémoire étant centré sur la question des interactions entre congestion et localisation des activités, nous nous proposons, dans le chapitre suivant, de détailler les formes de la congestion des réseaux de transport.

2

Les formes de la congestion : de la multiplicité à la complexité

2.1. Introduction

L'objectif de ce chapitre est de proposer une analyse des différentes formes que revêt le phénomène de congestion des transports, afin d'en souligner le caractère complexe et systémique et d'offrir un cadre théorique aux analyses menées dans la partie suivante de ce mémoire, en discutant notamment de la question des coûts de la congestion.

La congestion d'une infrastructure désigne un phénomène très courant : elle apparaît lorsque la qualité d'un service rendu par un système, en particulier un système de transport, se dégrade quand le nombre d'utilisateurs de ce service augmente (Small et Verhoef, 2007 ; Arnott et Kraus, 1995, 1998). Mais cette définition, si elle résume bien l'essence du phénomène de congestion, risque, par sa simplicité, d'en masquer la réalité complexe, en particulier en milieu urbain : la congestion est un phénomène complexe s'insérant dans un système encore plus complexe. Comme le souligne le rapport ECMT (2007), cette complexité implique que la connaissance des causes et des mécanismes immédiatement responsables de la congestion ne donne pas directement accès au fonctionnement global du système. Notre propos, dans ce chapitre comme dans les suivants, visera à mettre en évidence cette complexité et à en analyser les composantes.

Nous nous limiterons, dans ce chapitre, à la congestion des systèmes de transport et nous accorderons une attention particulière à la congestion routière. Néanmoins, comme le signalent Lindsey et Verhoef (2000), des parallèles utiles peuvent être faits entre la congestion routière et la congestion d'autres types de services. Un guichet de banque est sujet à la congestion, de même qu'un toboggan dans un square, un réseau informatique, ou une autoroute urbaine. Ainsi nous pourrions être amené à mobiliser ce type d'exemples pour illustrer certaines formes de congestion.

Suivant le contexte et le type de service sujet à la congestion, la notion de qualité de service varie. Le temps est bien souvent un élément important de la qualité de service, mais la congestion peut se manifester sous différentes formes. Ainsi, dans le cas du guichet de banque, c'est presque exclusivement le temps d'attente avant d'atteindre le guichet qui détermine la qualité de service, même si la présence, par exemple, de places assises peut diminuer l'impact de la congestion sur les individus. En revanche, dans le cas d'un réseau informatique, l'intensité d'usage peut, en diminuant la bande passante de chaque utilisateur, augmenter les temps de téléchargement, mais elle peut également entraîner des pertes de données ou un arrêt total du service, occasionnant une baisse de confort de l'utilisateur. Il en va de même pour le transport, pour lequel les éléments déterminant la qualité de service sont le temps de trajet, l'heure d'arrivée prévue, la fiabilité¹ et le confort pendant le déplacement (Small et Verhoef, 2007).

De la même manière, la congestion peut faire intervenir différentes « technologies ». Ainsi, le ralentissement ressenti lors d'une progression au sein d'une foule est issu d'un mécanisme très différent de celui en jeu lors de l'attente à un guichet ou derrière un feu rouge, ou encore lors de la recherche d'une place de stationnement en ville.

On le voit, la congestion se manifeste sous de nombreuses formes, aussi bien en termes de caractéristiques physiques et donc de mécanismes, qu'en termes de conséquences sur les usagers des réseaux de transport et sur le système urbain dans son ensemble. Nous distinguerons les formes psychologiques de la congestion, liées aux perceptions des usagers, et les formes physiques, c'est-à-dire les mécanismes élémentaires et leurs combinaisons. Les aspects économiques, autrement dit les effets de la congestion sur les systèmes économiques urbains, seront également abordés dans ce chapitre.

Les aspects psychologiques de la congestion sont primordiaux, puisque c'est de ces derniers que découlent en grande partie les enjeux de sa maîtrise. Celle-ci passe elle-

1. Dans le cas du transport collectif, la fréquence de desserte entre de manière prépondérante en ligne de compte, car elle influence à la fois le temps de trajet et la fiabilité : lorsque la fréquence est faible, un léger retard du premier service utilisé par un usager peut mener à un retard très important sur l'ensemble du trajet, si les correspondances ne sont pas correctement assurées.

même par la compréhension et la modélisation des mécanismes. Nous explorerons donc successivement dans ce chapitre les formes psychologiques de la congestion dans la deuxième section, puis les formes physiques dans la troisième section. La quatrième section discute la question des coûts de la congestion.

2.2. Formes psychologiques de la congestion

La congestion n'est pas uniquement un phénomène physique, comme il peut en exister en mécanique des fluides. Les volumes de trafic sur les réseaux de transport sont le résultat des choix des usagers, de leurs comportements ponctuels, des réglementations et des habitudes. Ces éléments sont influencés par la perception des conditions de circulation (Lindsey et Verhoef, 2000), dont rend compte le rapport ECMT (2007).

Ce dernier présente ainsi les résultats d'une enquête réalisée par le ministère britannique des Transports auprès des conducteurs, visant à déterminer les principaux problèmes que la congestion engendre pour eux. La Table 2.1, qui en reprend les principaux enseignements, présente les différents items dans un ordre rendant approximativement compte de la fréquence de leur mention par les conducteurs interrogés. La classification en trois catégories est de notre fait.

| Type d'effet | Effet |
|------------------------|----------------------------------------------|
| Temporel | Retards ou rendez-vous manqués |
| | Nécessité de prévoir du temps supplémentaire |
| | Imprévisibilité du temps de parcours |
| | Remise d'un déplacement |
| Comportemental | Stress, énervement |
| | Fatigue |
| | Colère vis-à-vis des autres conducteurs |
| Environnemental | Consommation de carburant supplémentaire |
| | Pollution supplémentaire |

TABLE 2.1.: Principaux problèmes engendrés par la congestion et ressentis par les conducteurs. *Source* : DfT (2001).

Les aspects temporels apparaissent donc comme les plus nombreux pour les usagers. Mais d'autres effets sont également évoqués, en particulier l'inconfort lié à l'attention supplémentaire que requiert la conduite en situation congestionnée. C'est l'ensemble de ces effets que nous appelons formes psychologiques de la congestion. Nous présenterons successivement les formes psychologiques liées au temps, puis celles liées au stress et à l'inconfort.

Les aspects énergétiques et environnementaux apparaissent en dernier. En ce qui concerne la pollution, c'est probablement dû au fait que la pollution supplémentaire subie du fait de la congestion est difficilement perceptible par les usagers, de par son caractère externe. C'est pourquoi nous aborderons ces questions environnementales dans la section 2.4 portant sur les coûts économiques de la congestion.

2.2.1. Temps perdu et fiabilité

De manière plus précise, la congestion, sous ses différentes formes, engendre deux types d'effets temporels. Le premier correspond à une perte de temps par rapport à une situation non congestionnée. Il s'agit de la forme la plus commune et intuitive de congestion : la présence de nombreux usagers allonge le temps de parcours, et ce, d'autant plus que le nombre d'usagers augmente. C'est cette première forme temporelle de la congestion qui nous intéressera dans la suite de ce mémoire et que nous présentons plus en détail dans le paragraphe 2.2.1.1.

Cependant, la variabilité du temps de transport, et donc sa fiabilité, a également une grande importance pour les usagers. Elle fait l'objet du paragraphe 2.2.1.2. Dans la suite de notre mémoire, nous tenterons de l'approcher grâce à certains de nos indicateurs.

2.2.1.1. La valeur du temps

Le temps passé en transport, ou plutôt le temps passé en transport *économisé*, a une valeur pour les usagers. Ainsi, comme l'explique Jara-Diaz (2007) :

« *Les individus préféreraient faire quelque chose d'autre, chez eux, au travail, ou ailleurs, plutôt que de prendre le bus ou conduire leur voiture.* »²

La notion de *valeur du temps*, directement issue de cette constatation, permet de mesurer effectivement la valeur que les usagers donnent au temps perdu en transport, et en particulier celui lié à la congestion. Plus précisément, en supposant que l'utilité U d'un déplacement dépend du temps de parcours T et de son coût C , la valeur du temps s'écrit³ :

$$v_T = \frac{\partial U / \partial T}{\partial U / \partial C} \quad (2.1)$$

DeSerpa (1971) fonde cette formulation sur un modèle mêlant consommation de biens et du temps⁴. Ainsi, on suppose que l'utilité U d'un usager dépend de sa

2. Traduction de l'auteur.

3. Nous reprenons ici les notations de Brownstone et Small (2005).

4. Les premiers travaux sur la valeur du temps sont dus à Becker (1965), mais le modèle de DeSerpa (1971) est plus complet tout en restant suffisamment simple pour être adapté à un exposé

consommation d'un bien composite pris comme numéraire z , du temps passé au travail T_w et des temps T_k passés aux diverses autres activités k .

L'usager cherche à maximiser son niveau d'utilité en tenant compte de contraintes, qui proviennent de l'existence de deux budgets : un budget financier et un budget temporel. Concernant le budget financier, l'agent touche un revenu non salarial Y et un revenu salarial wT_w , w étant le taux de salaire. Concernant le budget temporel, il dispose d'un temps total \bar{T} fixe et chaque activité k nécessite un temps minimal \underline{T}_k pour être réalisée.

Le Lagrangien du problème s'écrit :

$$\begin{aligned} \Lambda = & U(z, T_w, \{T_k\}) + \lambda[Y + wT_w - z] \\ & + \mu \left[\bar{T} - T_w - \sum_k T_k \right] + \sum_k \phi_k [T_k - \underline{T}_k] \end{aligned} \quad (2.2)$$

où λ , μ et $\{\phi_k\}$ sont les multiplicateurs de Lagrange associés. On note que dans ce modèle, les activités n'ont pas de consommation minimale associée, et donc de coût monétaire explicite. Jara-Diaz (2003) améliore ce point que nous avons préféré omettre pour simplifier l'exposé.

Les conditions de premier ordre fournissent :

$$\frac{dU}{dT_k} - \mu - \phi_k = 0 \quad (2.3)$$

$$\frac{dU}{dT_w} + \lambda \left[w + T_w \frac{dw}{dT_w} \right] - \mu = 0 \quad (2.4)$$

en supposant que le taux de salaire est fonction du temps de travail de l'agent.

Par ailleurs, en notant V la fonction d'utilité indirecte, autrement dit le niveau d'utilité à la solution du problème précédent, la valeur de la réduction du temps minimal à passer à une activité k s'écrit donc comme le rapport entre l'utilité marginale de la réduction de ce temps et l'utilité marginale du revenu (non salarial). Ce qui, au point optimal, s'écrit comme le rapport entre les deux multiplicateurs de Lagrange associés :

$$v_T^k = -\frac{\partial V / \partial T_k}{\partial V / \partial Y} = \frac{\phi_k}{\lambda} \quad (2.5)$$

Les activités dont la contrainte de temps minimale n'est pas saturée, autrement dit pour lesquelles $\phi_k = 0$, sont par définition des activités de pur loisir. Les autres, comprenant le transport, sont des activités intermédiaires (par opposition au travail). En combinant les conditions de premier ordre, on obtient la valeur du temps

succinct. Les notations employées ici sont de Small et Verhoef (2007).

d'une activité de transport :

$$v_T^k = \frac{\mu}{\lambda} - \frac{\partial U / \partial T_k}{\lambda} = \left[w + T_w \frac{dw}{dT_w} \right] + \left[\frac{\partial U / \partial T_w}{\lambda} - \frac{\partial U / \partial T_k}{\lambda} \right] \quad (2.6)$$

L'expression est composée de deux termes : le premier est pécuniaire, et correspond au revenu supplémentaire que permettrait d'obtenir une réduction du temps de transport. Si w est supposé fixe, ce terme se réduit au taux de salaire. Le second terme est non pécuniaire et peut être positif ou négatif, selon que le temps passé au travail procure plus ou moins d'utilité que le temps passé en transport. De plus, cette expression met en évidence le fait que la valeur du temps est vraisemblablement croissante avec le temps total de transport, car μ , l'utilité marginale du loisir, augmente lorsque la contrainte temporelle globale devient plus serrée.

Enfin, l'équation (2.5) peut se réécrire en supposant qu'un déplacement implique un coût c_k . Dans ce cas, on a en effet : $\frac{\partial V}{\partial c_k} = -\lambda = -\frac{\partial V}{\partial Y}$. Et donc :

$$v_T^k = \frac{\partial V / \partial T_k}{\partial V / \partial c_k} = \frac{\phi_k}{\lambda} \quad (2.7)$$

Small et Verhoef (2007), s'appuyant sur de nombreuses études menées dans différents pays (Boiteux et Baumstark, 2001, pour la France), avancent, pour cette valeur du temps, un chiffre moyen de 50 % du taux de salaire. De plus, cette valeur évolue avec le revenu avec une élasticité de 0,8 environ, donc moins que proportionnellement (Mackie *et al.*, 2003).

Calfee et Winston (1998), se basant sur une enquête de préférences déclarées ciblant les conducteurs urbains et ne reposant pas sur des choix de mode⁵, ont trouvé une valeur plus faible, autour de 20 % du taux de salaire. Ils expliquent cette différence par le fait que les ménages urbains ont ajusté leur localisation résidentielle, leur lieu d'emploi, leur mode, leur itinéraire et leur horaire de départ en tenant compte de la congestion. Ce résultat suggère une forte intrication des questions de localisations urbaines des activités et des ménages, et des questions de congestion, problématique au cœur de notre mémoire.

2.2.1.2. La valeur de la fiabilité

Si le temps de parcours a son importance, la régularité temporelle de ce temps de parcours joue également, comme le signalent Small *et al.* (1999), un rôle non négligeable dans l'impact de la congestion sur les usagers. En effet, le temps de parcours n'est pas nécessairement vu comme un fardeau par les usagers, car ils

5. La majorité des études repose sur des comparaisons entre modes de transport.

évaluent généralement leur temps de parcours sur la base de leur expérience, en prenant donc en compte la congestion habituellement présente sur leur itinéraire, c'est-à-dire la congestion *récurrente*.

Par contre, comme le rapport ECMT (2007) le souligne, les variations inattendues et imprévisibles du temps de parcours, liées à la congestion *incidente* (accident, véhicule en panne, problème de signalisation, événement exceptionnel, etc.), sont souvent très mal vécues par les usagers. La notion de *valeur de la fiabilité* permet de mesurer cet attachement des usagers à la régularité du temps de transport. En reprenant les mêmes notations que précédemment et en supposant que l'utilité U dépend également de la fiabilité F du temps de parcours, la valeur de la fiabilité s'écrit :

$$v_F = -\frac{\partial U / \partial F}{\partial U / \partial C} \quad (2.8)$$

Toutefois, la valeur de la fiabilité est une notion plus complexe que la valeur du temps. Elle est liée, entre autres, à la désutilité que subit un agent lorsqu'il arrive en retard (au travail ou plus généralement à un rendez-vous), et dans une moindre mesure en avance.

Plus précisément, le cadre de modélisation de Noland et Small (1995) permet de relier les différentes valeurs temporelles que sont la valeur du temps, la valeur du retard ou de l'avance, et la valeur de la fiabilité. Dans ce cadre, le coût généralisé G d'un déplacement donné s'écrit :

$$G = v_T T + \beta A + \gamma R + \theta D \quad (2.9)$$

où T désigne le temps de parcours, qui se décompose en la somme d'un temps prévisible T_p et d'un temps supplémentaire T_s stochastique, A et R désignent respectivement le retard ou l'avance subis par l'agent, et D est une variable binaire valant 0 si l'utilisateur arrive en avance ou à l'heure, et 1 s'il arrive en retard. Le coefficient v_T désigne la valeur du temps, tandis que β et γ sont respectivement les valeurs marginales du temps d'avance et de retard, et θ le coût fixe du retard.

On suppose de plus que l'agent peut ajuster son horaire de départ t_d . Autrement dit, le coût espéré qu'il peut obtenir en choisissant t_d de manière optimale est :

$$G^* = \min_{t_d} \mathbb{E}(G) = \min_{t_d} [v_T \mathbb{E}(T) + \beta \mathbb{E}(A) + \gamma \mathbb{E}(R) + \theta P_R] \quad (2.10)$$

où P_R est la probabilité de retard.

De manière à se focaliser sur la question de la fiabilité, on néglige, comme Small et Verhoef (2007), l'influence de l'horaire de départ sur T_p et donc sur $\mathbb{E}(T)$. Dans ce cas, si $f(T_s)$ désigne la densité de probabilité de T_s , t^* l'horaire désiré d'arrivée

et \tilde{t} l'horaire auquel l'agent doit partir pour arriver à t^* si $T_s = 0$, on a :

$$\begin{aligned}\mathbb{E}(A) &= \mathbb{E}(\tilde{t} - t_d - T_s \mid T_s < \tilde{t} - t_d) \\ &= \int_{-\infty}^{\tilde{t}-t_d} (\tilde{t} - t_d - T_s) f(T_s) dT_s\end{aligned}\quad (2.11)$$

$$\begin{aligned}\mathbb{E}(R) &= \mathbb{E}(t_d + T_s - \tilde{t} \mid T_s > \tilde{t} - t_d) \\ &= \int_{\tilde{t}-t_d}^{+\infty} (t_d + T_s - \tilde{t}) f(T_s) dT_s\end{aligned}\quad (2.12)$$

$$P_R = \int_{\tilde{t}-t_d}^{+\infty} f(T_s) dT_s \quad (2.13)$$

car la condition $T_s \leq \tilde{t} - t_d$ est équivalente à $t_d + T \leq t^*$ et indique donc s'il y a avance ou retard.

Les conditions du premier ordre pour l'horaire de départ optimal s'obtiennent en différentiant (2.11), (2.12) et (2.13) par rapport à t_d , ce qui donne :

$$\beta \frac{d}{dt_d} \mathbb{E}(A) = \beta \int_{-\infty}^{\tilde{t}-t_d} \frac{d}{dt_d} (\tilde{t} - t_d - T_s) f(T_s) dT_s = -\beta(1 - P_R^*) \quad (2.14)$$

$$\gamma \frac{d}{dt_d} \mathbb{E}(R) = \gamma \int_{\tilde{t}-t_d}^{+\infty} \frac{d}{dt_d} (t_d + T_s - \tilde{t}) f(T_s) dT_s = \gamma P_R^* \quad (2.15)$$

$$\theta \frac{d}{dt_d} P_R = \theta \frac{d}{dt_d} \int_{\tilde{t}-t_d}^{+\infty} f(T_s) dT_s = -\theta f(\tilde{t} - t_d^*) \quad (2.16)$$

où P_R^* est la probabilité d'être en retard, étant donné l'horaire optimal de départ. Par conséquent, en sommant les termes, et en égalisant à 0, on trouve la règle de choix de l'horaire de départ :

$$P_R^* = \frac{\beta + \theta f(\tilde{t} - t_d^*)}{\beta + \gamma} \quad (2.17)$$

qui définit implicitement t_d^* car P_R^* et $f(\tilde{t} - t_d^*)$ en dépendent.

Dans le cas particulier d'une distribution uniforme de T_s de densité $1/b$ et si $\theta = 0$, le coût G^* est la somme du coût de transport espéré $v_T \mathbb{E}(T)$ et d'un terme correspondant à la variabilité du temps de transport :

$$-v_F V = \frac{\beta \gamma}{\beta + \gamma} \frac{b}{2} \quad (2.18)$$

La valeur de la fiabilité v_F dépend ensuite de la manière de mesurer la variabilité V (écart-type, intervalle de confiance, etc.).

Brownstone et Small (2005) trouvent que la valeur de la fiabilité est du même ordre que la valeur du temps, résultat que Small et Verhoef (2007) présentent comme

commun à plusieurs autres études : Bates *et al.* (2001) constatent que le rapport généralement trouvé entre valeur de la fiabilité et valeur du temps s'étend entre 0,8 et 1,3. Par ailleurs, contrairement à la valeur du temps, qui augmente moins vite que le revenu, la valeur de la fiabilité augmente plus vite que le revenu (Small *et al.*, 1999).

En ce qui concerne l'Île-de-France, le rapport ECMT (2007) cite une étude de De Palma et Fontan (2000), dans laquelle sont mesurées les valeurs du temps de transport, d'une part, et du temps de retard d'autre part, qui permettent d'approcher la valeur de la fiabilité. Les auteurs trouvent un rapport supérieur à 2 entre les deux, soulignant l'importance de la fiabilité pour les usagers. Enfin, Debrincat *et al.* (2006) obtiennent, pour des liaisons ferrées radiales, une équivalence de 4,6 minutes pour une réduction de 5 % de la probabilité de retard.

2.2.1.3. Engrenages et corrélations

Les deux formes temporelles de la congestion, que nous venons d'évoquer, se combinent pour donner lieu à un effet global de la congestion en termes temporels, comme l'illustre la Figure 2.1.

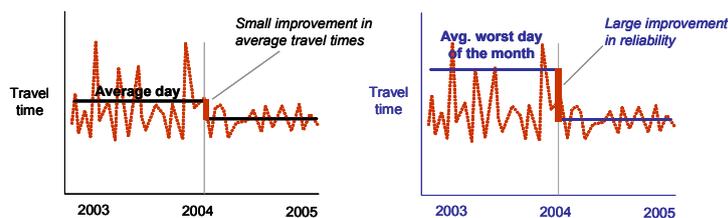


FIGURE 2.1.: Amélioration du temps de parcours moyen *versus* amélioration de la fiabilité. *Source* : ECMT (2007)

Par ailleurs, elles ne sont pas simplement complémentaires, elles sont dépendantes l'une de l'autre, et sont donc corrélées : un itinéraire très congestionné a une plus forte probabilité d'être sujet à une variabilité élevée du temps de parcours qu'un itinéraire sans congestion. Cela tient, selon nous, à deux causes principales :

- Les conséquences, en termes de congestion additionnelle, d'un incident en situation initialement congestionnée sont plus importantes qu'à faible trafic. Ce phénomène s'observe fréquemment en transport public où les répercussions d'un incident en heure de pointe prennent rapidement de l'ampleur, du fait des multiples effets de rebond. Leurent (2010) parle d'engrenages de congestion.
- Le nombre des usagers augmente la probabilité d'un incident. De manière directe d'abord, car plus il y a d'usagers présents simultanément, plus il y a de

risque qu'un incident survienne⁶. De manière indirecte ensuite, par l'intermédiaire de la densité : en transport public, l'inconfort peut entraîner des malaises, qui impliquent l'arrêt momentané de la mission ; sur la route, la conduite en situation de congestion requiert une attention supplémentaire et comporte de nombreux dangers favorisant l'apparition d'incidents. On notera néanmoins que les accidents en situation congestionnée sont généralement moins graves qu'en l'absence de congestion, car les vitesses sont plus faibles.

À ce sujet, Mohring (1999) note que lorsque seule la valeur du temps est prise en compte dans une étude, sans contrôler pour la variabilité, la valeur trouvée est plus élevée et augmente avec le revenu. Autrement dit, les usagers prennent en compte le fait qu'une réduction du temps de parcours liée à une baisse de la congestion entraîne généralement une baisse de la variabilité de ce temps.

Par ailleurs, le fait qu'une forme temporelle, à savoir la variabilité du temps de parcours, soit fortement liée à une forme comportementale, à savoir la difficulté de conduite en situation congestionnée, souligne l'importance de s'intéresser à ces deux dernières formes psychologiques de la congestion, que nous allons détailler maintenant.

2.2.2. Stress et inconfort

La congestion peut entraîner stress et inconfort. Le premier touche plus particulièrement les conducteurs, tandis que le second concerne plutôt les usagers des transports collectifs, pour lesquels la congestion se traduit directement en termes de densité de personnes. Dans les deux cas, ces manifestations de la congestion accentuent potentiellement le phénomène de congestion, puisqu'ils peuvent être à l'origine, entre autres, d'accidents, pour la route, et de malaises, pour les transports collectifs.

L'augmentation du niveau de stress lors de la conduite en situation de congestion est bien établie, comme le note le rapport ECMT (2007). Celle-ci peut entraîner, chez les conducteurs, des comportements comme queues de poisson, conduite agressive, coups de klaxon, freinages et accélérations abrupts, manque de concentration, etc. Les effets de ce stress sur la sécurité routière ont été relativement peu étudiés, principalement en raison des difficultés méthodologiques liées au recueil des données et à leur exploitation.

6. Ce point n'est toutefois pas entièrement clair, car les faibles vitesses liées à la congestion limitent le risque d'accident et leur gravité. En ce qui concerne les incidents liés à des pannes, en revanche, la densité a tendance à augmenter leur fréquence.

Or, l'étude de l'effet d'un accident en situation congestionnée sur la détérioration des conditions de circulation (*via* la réduction de la capacité, la curiosité des autres conducteurs, etc.) de même que celle de l'aggravation potentielle d'un accident du fait de la difficulté de progression des secours en zone congestionnée serait d'un intérêt de premier plan.

En transport collectif, c'est la notion de confort qui prévaut. L'inconfort ressenti par les usagers peut être à l'origine (*via* des malaises, des altercations entre voyageurs, etc.) d'engrenages de congestion (Leurent, 2010), mais peut également modifier leur choix d'itinéraire (Leurent et Liu, 2009), voire leur choix de mode. Outre les conséquences de ces phénomènes pour un opérateur de transport collectif, on perçoit ici une forte intrication entre les conditions de circulation routière et le confort en transport collectif, puisque l'inconfort peut pousser certains usagers à choisir la voiture, d'autant plus que les conditions de circulation routière sont bonnes, et réciproquement.

La valeur du confort s'évalue généralement en lien avec la valeur du temps. Plus précisément, il est possible de déterminer plusieurs valeurs du temps, en fonction du confort — objectif ou subjectif — des usagers. Ainsi, Debrincat *et al.* (2006) trouvent, dans une enquête de préférences déclarées, que la valeur du temps passé debout est 1,6 fois celle du temps passé assis, et même 1,9 fois si le véhicule est bondé. Haywood et Koning (2010) évaluent, quant à eux, la disponibilité à payer des usagers de la ligne 1 du métro parisien en heure de pointe pour voyager dans les conditions des heures creuses, et trouvent une valeur autour de 6,5 minutes, ce qui représenterait une augmentation de près de 35 % de la durée du trajet.

Il apparaît donc que les conditions de confort des déplacements sont très fortement valorisées par les usagers. Pourtant, l'étude et la modélisation des effets du confort sur les choix des voyageurs en sont encore à leurs balbutiements.

Nous avons, dans les paragraphes précédents, présenté les formes psychologiques de la congestion, autrement dit les conséquences de celle-ci sur les usagers des transports, et nous avons insisté sur les interdépendances, parfois complexes, entre ces différentes manifestations. Nous allons, dans la section suivante, nous intéresser aux mécanismes en jeu dans le phénomène de congestion des transports, en insistant là aussi sur les intrications des différentes formes physiques.

2.3. Formes physiques de la congestion

Les mécanismes à l'œuvre dans le phénomène de congestion des transports ont été étudiés dès les années 1950, dans le cadre routier initialement. La première édition du désormais classique *Highway Capacity Manual* date ainsi de 1949. Les travaux

de Beckmann *et al.* (1956) sont également marquants dans l'étude de la congestion des infrastructures routières, par l'effort de modélisation qu'ils ont initié. La suite de ce mémoire s'intéressant plus spécifiquement au mode routier, c'est sur ce dernier que nous porterons une attention particulière. Nous consacrerons toutefois quelques paragraphes à la congestion des transports collectifs.

Les causes à l'origine de la congestion sont multiples selon Small et Verhoef (2007) : ainsi, des files d'attente peuvent se former aux intersections à feux ; les véhicules cherchant à s'insérer dans le flux principal peuvent gêner la circulation ; un véhicule lent peut ralentir l'ensemble du trafic derrière lui si les possibilités de dépassement sont réduites. Ces quelques exemples mettent en évidence une première distinction entre congestion sur les arcs et congestion aux nœuds. En réalité, elle est elle-même fortement liée à une seconde distinction, que nous estimons plus générale, entre congestion de flux et congestion de stock.

Nous commencerons par présenter un cadre d'analyse et de modélisation général, permettant de mettre en évidence ces deux formes de congestion. Nous les présenterons ensuite successivement, puis nous les replacerons dans un contexte d'interactions multiples.

2.3.1. Le diagramme fondamental du trafic et le modèle LWR

Une infrastructure routière se comporte de manière particulière en ce qui concerne la relation entre le nombre d'utilisateurs servis et la qualité de service. Au contraire d'un guichet de banque, par exemple, pour lequel le débit de clients, c'est-à-dire le nombre de clients servi par minute, est globalement constant, le débit qu'offre une infrastructure routière n'est, de manière générale, pas constant.

Ainsi, le débit Q sur une voie est lié à la densité D de véhicules sur cette voie, c'est-à-dire au nombre de véhicules par unité de longueur de voie, et à la vitesse V des véhicules sur la voie, par la relation $D = Q/V$ (cf. par exemple, pour une présentation générale, Quinet, 1998 ; Prud'homme, 1999a). Or, l'expérience montre que, du fait des comportements de conduite, la vitesse sur une voie est inversement proportionnelle à la densité. Par conséquent, débit et vitesse sur une route, autrement dit quantité et qualité de service, sont liées de manière quadratique. Cela se traduit à travers ce qui est couramment appelé, depuis Haight (1963), le « diagramme fondamental du trafic », présenté sur la Figure 2.2, où les valeurs V_0 et D_J y correspondent respectivement à la vitesse à vide et à la densité de blocage (*jam density*).

On notera également que le débit fourni par l'infrastructure est maximal lorsqu'il atteint la capacité κ de l'arc (point K sur le graphique). Autrement dit, c'est lorsque les véhicules circulent à la vitesse V_κ que l'infrastructure délivre le débit maximal

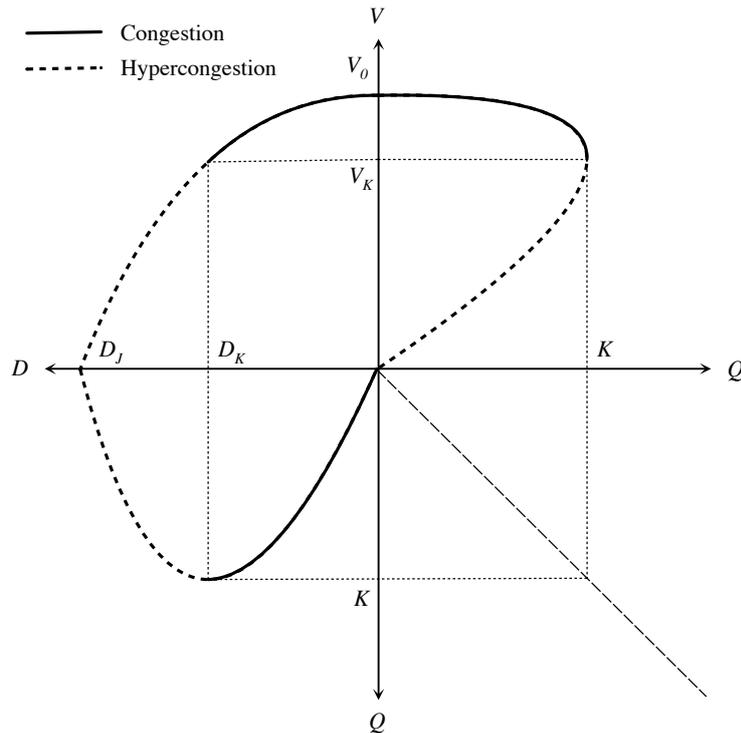


FIGURE 2.2.: Le diagramme fondamental du trafic. *Source* : Verhoef (1999)

dont elle est capable. Par conséquent, et même s'il y a bien entendu d'autres manières de définir la capacité, en particulier celle de la capacité pratique difficilement quantifiable, fournie par le premier *Highway Capacity Manual*⁷, nous retiendrons dans la suite de l'analyse la définition stationnaire de la capacité (Branston, 1976) : il s'agit du débit stationnaire maximal sur cet arc, c'est-à-dire le flux constant maximal que l'arc peut écouler. Cette notion de capacité est centrale dans l'étude de la congestion, en transport comme dans d'autres domaines. Elle constitue à la fois une valeur de référence, et un seuil de fonctionnement du système.

La Figure 2.2 met également en évidence deux régimes de fonctionnement de l'arc : congestion et hypercongestion. Il s'agit là des deux formes physiques principales de la congestion routière. En effet, l'observation des deux quadrants supérieurs permet de noter que la valeur V_K sépare l'ensemble des états possibles du trafic en deux domaines : pour les valeurs supérieures à V_K , il s'agit du régime congestionné, où la congestion prend la forme d'une *congestion de flux*. Pour les valeurs infé-

7. La définition du HCM de 1949 est la suivante : « la capacité d'un arc est le nombre maximal de véhicules qui peuvent passer par un point donné de la route ou d'une voie particulière pendant une heure sans que la densité de trafic devienne tellement élevée qu'elle entraîne un retard ou un danger déraisonnable, ou qu'elle restreigne la liberté de manœuvre des conducteurs par rapport aux conditions normales de circulation » (traduction de l'auteur).

rieures à V_κ , il s'agit du régime hypercongestionné, où la congestion tend à prendre la forme d'une *congestion de stock*, autrement dit une file d'attente. La première forme est bien adaptée à une modélisation statique de la congestion, tandis que la seconde, du fait que l'état du système dépend de l'*histoire* des flux sur l'arc ou sur le réseau considéré, nécessite généralement la mise en place d'une modélisation dynamique. Le diagramme le met en évidence : ces deux formes de congestion sont liées, au sens où un niveau élevé de congestion de flux entraîne une saturation de l'arc considéré et la formation d'une file d'attente. Néanmoins, les mécanismes restent fondamentalement distincts et leurs modélisations généralement très différentes.

Toutefois, le modèle LWR (Lighthill et Whitham, 1955 ; Richards, 1956) exploite directement ce diagramme fondamental du trafic et constitue en cela un cadre conceptuel général particulièrement simple et efficace. Celui-ci repose en effet sur l'hypothèse que la relation débit-densité, valable en régime stationnaire, l'est également en régime non stationnaire. De plus, une équation de conservation, éventuellement discrétisée, qui assure que les véhicules ne sont pas créés ou détruits, est employée :

$$\frac{\partial q(t, x)}{\partial x} + \frac{\partial d(t, x)}{\partial t} = 0$$

Ainsi, le modèle LWR repose sur deux éléments : une hypothèse comportementale — la manière dont les usagers réagissent à l'augmentation du débit sur un arc — et une relation purement comptable — une équation de conservation.

En combinant les deux relations, on obtient, puisqu'on a $q = q(d)$ par la relation débit-densité exogène :

$$q'(d) \frac{\partial d(t, x)}{\partial x} + \frac{\partial d(t, x)}{\partial t} = 0 \quad (2.19)$$

La résolution du modèle consiste donc en la résolution de cette équation aux dérivées partielles du premier ordre.

Le modèle permet d'analyser les conséquences, dans l'espace et le temps, d'une variation brutale des conditions de circulation : débit, capacité, etc. Ainsi, à partir d'un état initial, caractérisé par un débit q_1 et une densité d_1 , donc une vitesse des véhicules $v_1 = q_1/d_1$, on suppose une modification du débit, qui passe à q_2 . La relation débit-densité, supposée exogène, fournit la densité d_2 correspondante, et donc la vitesse v_2 .

Dans un diagramme espace-temps, la ligne de partage entre les deux états, ou *onde de choc*, se propage à une vitesse $w_{12} = (q_2 - q_1)/(d_2 - d_1)$, inférieure à v_1 et v_2 . La direction de propagation dépend du signe de w_{12} et donc des positions relatives des points (d_1, q_1) et (d_2, q_2) . À titre d'exemple, si un flot se déplaçant à haute vitesse (et faible densité) rejoint un flot, situé en aval sur la route, se déplaçant à

plus faible vitesse (et densité plus élevée mais inférieure à d_κ , la densité à capacité), le lieu du début de ralentissement se déplace dans le sens de circulation à une vitesse inférieure aux vitesses des deux flots.

On le voit, le modèle LWR est surtout utile pour étudier les conséquences des variations discrètes des conditions de trafic, en particulier les modifications temporaires de la capacité routière. Il est mal adapté pour traiter des variations continues du flux (Lindsey et Verhoef, 2000).

Par ailleurs, c'est un outil de choix pour l'étude du trafic à l'échelle d'un arc, mais il est relativement mal adapté pour étudier des réseaux de transport dans leur ensemble, du fait de la complexité rapidement croissante des équations à résoudre dans le cas de routes non homogènes, avec présence d'un maillage important, comme c'est le cas en milieu urbain⁸.

Un des intérêts du modèle, pour notre étude des formes physiques de la congestion, est d'offrir un cadre de modélisation commun à la congestion de flux et à certaines formes de congestion de stock. Il constitue ainsi un pont entre ces deux formes de congestion.

Plus précisément, en ajoutant des hypothèses simplificatrices au modèle LWR, il est possible de le « spécialiser » dans l'une ou l'autre des formes de congestion, tout en le simplifiant, facilitant l'obtention d'un équilibre des flux sur un réseau⁹.

Nous présenterons donc successivement les deux formes de congestion que sont la congestion de flux et la congestion de stock. L'ingénieur, comme l'économiste, cherche à rendre la complexité du réel plus simple à manipuler à des fins d'évaluation, de conception ou de prévision. La modélisation constitue, pour ce faire, l'approche privilégiée. C'est pourquoi nous avons choisi d'aborder cette question par le biais des modèles permettant de représenter les différentes formes physiques.

2.3.2. Congestion de flux

Tant que la vitesse des véhicules est supérieure à V_κ , le flux est ininterrompu. La congestion se manifeste par une baisse de vitesse, due à la diminution de la distance de sécurité entre les véhicules, elle-même liée à l'augmentation de la densité locale, comme en rend compte le diagramme fondamental du trafic. Il s'agit donc d'un phénomène agissant au sein d'un *écoulement* de flux, relativement uniforme.

En termes de modélisation, cette forme de congestion se prête bien à un traitement statique, que nous détaillons au 2.3.2.2.

8. Le *Cell Transmission Model* de Daganzo (1994, 1995) procède à une forme de discrétisation du modèle LWR. Il est employé à l'échelle de réseaux entiers, mais reste complexe à manipuler.

9. La question de l'équilibre sur un réseau sera abordée au 2.3.4.3.

2.3.2.1. Relations débit - vitesse

Avant d'aborder le traitement statique de la congestion de flux, il nous semble utile de préciser que la modélisation statique des flux repose essentiellement sur des relations débit-vitesse. Plus précisément, Small et Verhoef (2007) distinguent quatre types de relations : les relations instantanées, les relations moyennées dans l'espace, les relations stationnaires et les relations moyennées dans le temps. Parmi ces quatre types de relations, ce sont principalement les deux dernières — les relations stationnaires et les relations moyennées dans le temps — qui sont employées en modélisation statique.

Les relations *instantanées* sont très utiles pour l'ingénieur du trafic, car elles permettent d'analyser la réponse instantanée d'un arc routier à la sollicitation instantanée à laquelle il est soumis. En revanche, pour l'économiste, dont l'objectif est d'analyser la qualité de service dans son ensemble, en fonction de la demande totale des usagers cherchant à se déplacer, les autres types revêtent un intérêt beaucoup plus grand.

Les relations *moyennées dans l'espace* s'obtiennent en mettant en relation, sur une infrastructure donnée, typiquement une portion d'autoroute, et pendant une période donnée, la vitesse moyenne et le taux de charge moyen. Elles permettent ensuite d'étudier, de façon simple, la réponse de l'infrastructure à une demande en entrée pendant une période de temps. Elles peuvent donc être vues comme des courbes d'offre puisqu'elles relient un coût, la vitesse ou plutôt le temps de franchissement, à une quantité, le débit. Néanmoins, ce type de relation présente l'inconvénient d'être spécifique d'une infrastructure donnée car dépendant de ses caractéristiques précises. Il n'est donc pas facilement transposable à d'autres infrastructures.

Les relations *stationnaires* font pour leur part référence à une situation dans laquelle le débit est constant dans le temps, et sur une durée infinie. Outre le fait que la variation temporelle de la demande, et donc du débit, n'est pas prise en compte dans les modèles stationnaires, ces derniers présentent également l'inconvénient de ne pas pouvoir admettre de débit supérieur à la capacité physique de l'infrastructure, car cela engendrerait des temps de parcours infinis.

Les relations *moyennées dans le temps*, de leur côté, s'apparentent à des relations stationnaires, mais présentent sur ces dernières l'avantage de renseigner sur le comportement d'une infrastructure lorsque la demande excède la capacité, au moins pour une période de temps limitée. La raison en est simple : si un état stationnaire dans lequel la demande dépasse la capacité aboutit par définition à des temps de franchissement infinis, dans le cas d'une relation moyennée sur une période de temps

finie, un dépassement de la capacité entraînera un temps de franchissement élevé, correspondant à la durée de résorption de la file d'attente, mais néanmoins fini. Les relations temporellement moyennées tendent vers les relations stationnaires lorsque la période de temps considérée tend vers l'infini. La variabilité temporelle peut être analysée pour sa part en considérant plusieurs périodes temporelles successives.

2.3.2.2. Modélisation statique des flux

Les modèles statiques mettent en œuvre une relation fixe, *statique*, entre le débit et la vitesse des véhicules, fournissant, pour chaque niveau de débit, la ou les vitesses des usagers sur l'arc, autrement dit le temps de franchissement de l'arc. En d'autres termes, la dimension temporelle n'est pas explicitement décrite. Pour fixer les idées, la modélisation statique s'obtient, à partir du modèle LWR, en supposant qu'aucune variation de la capacité et du débit entrant n'intervient pendant la période considérée (Lindsey et Verhoef, 2000). Par conséquent, aucune onde de choc n'apparaît, et le débit comme la vitesse sont uniformes sur l'ensemble de l'arc, et dans le temps.

Nous l'avons vu, les relations débit-vitesse, à la base de la modélisation statique, peuvent prendre différentes formes, selon les hypothèses à l'origine de leur obtention (stationnaires, moyennées dans le temps ou l'espace) d'une part, et selon les formes fonctionnelles particulières retenues pour la modélisation de l'affectation routière, d'autre part. Ainsi, nous comparerons, au chapitre 3, quelques formes fonctionnelles couramment employées en modélisation statique opérationnelle.

À titre d'illustration, une forme fonctionnelle couramment employée est celle du BPR (1964), qui fournit le temps de parcours d'un arc en fonction du débit Q qui le parcourt pendant une période donnée :

$$T_p = T_0 \left[1 + \alpha \left(\frac{Q}{\kappa} \right)^\beta \right], \quad (2.20)$$

où α et β sont deux paramètres rendant compte de la sensibilité du temps de parcours au volume de trafic, κ la capacité de l'arc et T_0 le temps de parcours à vide.

Cette formulation fait apparaître l'augmentation du temps de parcours (et donc la baisse de la vitesse) avec l'augmentation du débit. Par ailleurs, on peut vérifier que cette fonction ne rend *a priori* pas compte de la branche hypercongestionnée de la relation débit-vitesse. En réalité, les relations débit-vitesse moyennées dans le temps permettent de rendre compte du temps de parcours moyen des véhicules pendant une période donnée en fonction du débit moyen, incorporant donc le temps d'attente en file. En choisissant bien la période considérée de manière à couvrir le temps de résorption de la file d'attente, l'absence de la branche hypercongestionnée

peut être palliée par des temps de parcours très longs lorsque le débit devient très élevé.

Les principaux atouts de la modélisation statique des flux sont d'une part sa simplicité à la fois conceptuelle et d'utilisation, liée à la relation fixe entre le débit (ou la densité) et la vitesse, et d'autre part son intégration relativement aisée dans des modèles d'affectation des flux sur les réseaux, car le caractère statique permet de s'affranchir de nombreuses difficultés liées à la chronologie de l'écoulement¹⁰.

C'est ce qui explique le succès de ce type de modélisation, et donc la prépondérance, dans les utilisations opérationnelles, de la prise en compte de la congestion de flux par rapport à la congestion de stock.

Par ailleurs, si le domaine d'application par excellence de ce type de modélisation est la modélisation routière, la congestion de flux se manifeste également dans d'autres modes de transport : il suffit de penser à un flux de piétons dans un couloir du métro, ou au flux de trains le long d'une ligne de chemin de fer. Les différents types de relations débit-vitesse peuvent ainsi être employés en modélisation de la congestion des transports collectifs, en particulier pour rendre compte de la circulation des véhicules (Leurent, 2010).

Si la modélisation statique présente, nous l'avons vu, *a priori* l'inconvénient de ne pas prendre en compte le régime d'hypercongestion, l'utilisation de relations moyennées dans le temps permet en partie de pallier le problème, à condition de choisir une période suffisamment large pour englober entièrement le temps de résorption de la file d'attente.

Toutefois, la prise en compte explicite de telles files d'attente — durée de dépassement de capacité, temps de parcours réel et non moyen — nécessite une modélisation dynamique. Les files d'attente sont en effet caractéristiques de la congestion de stock.

2.3.3. Congestion de stock

La congestion de stock traite des situations où les véhicules ne s'écoulent pas en un flux relativement uniforme, mais où ils s'accumulent au contraire en files d'attente. Autrement dit, les situations de congestion de stock sont caractérisées par des temps *d'attente* plutôt que des temps de parcours. En transport routier, ce type de congestion se rencontre dans différentes situations : les carrefours à feu et les cas d'hypercongestion en section courante. Ces derniers cas correspondent à une demande excédant la capacité de l'arc, et plus particulièrement au cas du goulot,

10. Nous évoquerons plus en détail cette question au 2.3.4.3.

défini comme une section de route dont la capacité est inférieure à la capacité des autres sections de route, en amont et en aval (Cohen, 2005).

En transport collectif, en revanche, plusieurs situations peuvent être à l'origine d'une file d'attente, et donc d'une congestion de stock. Ainsi, Leurent (2010) cite, outre l'analogie du goulot pour la circulation des trains, les stocks de voyageurs en attente d'un train (ou d'un bus) alors même que ceux-ci sont bondés.

Le phénomène de stock peut être scindé en deux formes particulières : la première est déterministe, au sens où elle ne tient pas compte de l'arrivée stochastique des véhicules en entrée d'un arc ou d'une intersection ; la seconde est aléatoire, et prend en compte ce phénomène, d'ampleur généralement moins importante en dehors des intersections proches de la saturation (Newell, 1965).

2.3.3.1. Une forme déterministe : le goulot

Le modèle de goulot est le résultat d'une simplification du modèle LWR, de manière à le rendre analytiquement soluble : la relation débit-densité est supposée horizontale jusqu'à la capacité, puis verticale. En d'autres mots, on néglige la congestion de flux, et lorsque le flux entrant dans le goulot dépasse la capacité, il le quitte au taux constant égal à la capacité, si bien que la baisse de débit observée en situation d'hypercongestion est également négligée. L'objectif du modèle est de représenter plus fidèlement le phénomène de file d'attente, notamment dans sa dimension temporelle. Ainsi, le modèle de goulot rend explicitement compte de l'évolution de la longueur de la file d'attente et du temps de franchissement de l'arc qui en résulte. Nous proposons ici une présentation succincte du modèle, adaptée de Lindsey et Verhoef (2000) et Small et Verhoef (2007).

Soit $Q_e(t)$ le volume de trafic se présentant à l'instant t à l'entrée d'un goulot ponctuel de capacité κ . $Q_s(t)$ est le volume de trafic quittant le goulot à t . Le nombre de véhicules présents dans la queue à t est $N(t)$, la somme nette des entrées et des sorties du goulot. Autrement dit, la variation temporelle de N s'exprime simplement en fonction de Q_e et Q_s :

$$\dot{N}(t) = Q_e(t) - Q_s(t). \quad (2.21)$$

Par ailleurs, le débit de sortie du goulot est limité par la capacité. Autrement dit :

$$Q_s(t) = \begin{cases} Q_e(t) & \text{si } Q_e(t) \leq \kappa \\ \kappa & \text{sinon} \end{cases}. \quad (2.22)$$

Ces deux équations impliquent immédiatement qu'une file d'attente ne se crée que si le débit d'entrée dépasse la capacité à un moment donné.

Supposons, pour fixer les idées, que le débit en entrée croît à partir d'une valeur

faible, jusqu'à une valeur finie supérieure à la capacité, puis décroît. On note t_κ^i l'instant où Q_e dépasse κ pour la première fois. Par conséquent, $N(t) = 0$ pour $t \leq t_\kappa^i$. La file d'attente augmente jusqu'à l'instant où Q_e atteint son maximum, puis décroît jusqu'à un instant $t_\kappa^f > t_\kappa^i$ défini par :

$$\int_{t_\kappa^i}^{t_\kappa^f} (Q_e(t) - \kappa) dt = 0. \quad (2.23)$$

Les usagers arrivant au goulot entre t_κ^i et t_κ^f doivent donc attendre, en file, que les véhicules devant eux passent le goulot. Autrement dit, ils subissent un retard donné par :

$$T_d = \frac{N(t)}{\kappa} = \int_{t_\kappa^i}^t \left(\frac{Q_e(z)}{\kappa} - 1 \right) dz. \quad (2.24)$$

À titre d'illustration, la Figure 2.3, qui présente les nombres cumulés de véhicules entrant et sortant du goulot en fonction du temps, permet de trouver, graphiquement, la longueur de la file d'attente et le temps d'attente.

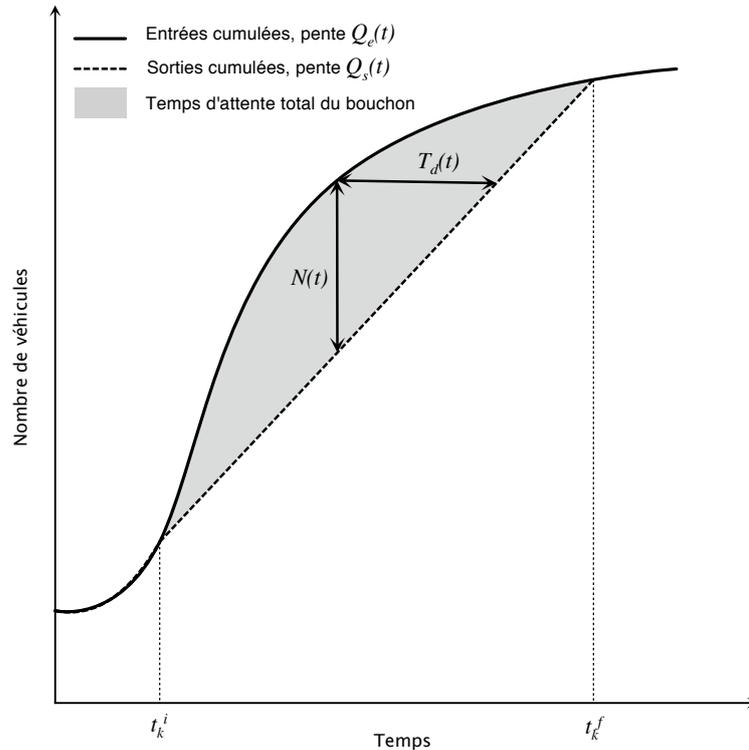


FIGURE 2.3.: Longueur de file et temps d'attente dans le modèle de goulot. *Source* : Small et Verhoef (2007)

Les atouts de ce modèle sont multiples. Citons-en deux : il permet, nous l'avons dit, de rendre compte des niveaux élevés de congestion, de manière dynamique,

sous la forme d'un temps d'attente en file explicitement représenté. De plus, il est soluble de manière analytique et constitue donc, pour les économistes, le modèle privilégié pour analyser les questions de choix d'horaire (Vickrey, 1969).

Le formalisme permet par ailleurs de représenter la capacité en sortie comme un profil temporel, ce qui convient aussi quand la capacité varie de manière contrôlée (par exemple, un débouché sur une intersection à feux, comme nous allons le voir au 2.3.4.1).

Du côté des limites, la version de base du goulot suppose implicitement que le stock est vertical en sortie (puisque'un véhicule arrive en stock *après* un temps de parcours t_0). Cette hypothèse présente deux inconvénients (Lindsey et Verhoef, 2000) : tout d'abord, elle entraîne une surestimation du retard, puisqu'elle néglige le fait qu'en réalité, une partie du temps d'attente en stock sert à se déplacer du début de la file d'attente jusqu'au goulot proprement dit. De plus, si la file d'attente remonte suffisamment en amont pour bloquer ou gêner d'autres jonctions (voies d'insertion ou de sortie, par exemple), les capacités des arcs qui aboutissent en ces jonctions sont modifiées par l'état de saturation de l'arc en aval. Il s'agit du phénomène de *spillback*.

2.3.3.2. Une forme aléatoire : la file d'attente

L'analyse des phénomènes de file d'attente peut également être abordée à partir de modèles stochastiques, rendant compte de l'influence du caractère aléatoire des arrivées sur la formation des files (Newell, 1965).

En effet, en supposant que le temps de passage (à un goulot, une intersection, un guichet, etc.) est constant, le fait que les agents (véhicules par exemple) se présentent de manière aléatoire peut engendrer une file d'attente temporaire, même si l'intervalle moyen entre deux arrivées est supérieur au temps de passage, autrement dit même si la capacité est supérieure au débit d'arrivée. En d'autres termes, du fait du caractère aléatoire des arrivées, une file d'attente peut émerger là où le modèle de goulot n'en prédit aucune.

En utilisant la notation de Kendall¹¹, les situations typiques rencontrées en transport sont de type (M/D/1) ou (M/M/1). Le premier cas correspond à un goulot fixe, à une intersection à feux ou à l'attente en station d'un service de transport collectif, tandis que le second correspond à une intersection sans feux (avec priorités).

11. Cette notation permet de représenter un modèle de file d'attente sous la forme (Processus d'arrivée / Processus de service / Nombre de serveurs), auquel on ajoute éventuellement (Discipline de service / Nombre d'utilisateurs). Chaque élément de cette notation est représenté par une lettre fixée par l'usage et les conventions. Ainsi, la lettre M (*Markovian*) désigne un processus de Poisson et donc une distribution exponentielle, D un processus déterministe et donc une distribution constante, G un processus général quelconque ; par défaut la discipline de service est de type FIFO (*First In First Out*), et le nombre d'utilisateurs infini.

Ces modèles permettent de calculer le temps moyen d'attente en file des usagers. Plus précisément, dans le cas (M/D/1), en supposant que les intervalles inter-arrivées suivent une loi exponentielle de paramètre λ et que la durée de service est $1/\mu$, ce temps s'exprime de la manière suivante :

$$T_a = \frac{\lambda}{2\mu(\mu - \lambda)}. \quad (2.25)$$

La condition $\mu > \lambda$ assure que le serveur n'est pas saturé, sans quoi la file s'allonge à l'infini.

De même, dans le cas d'un modèle (M/M/1), en supposant que la loi des durées de service est exponentielle de paramètre μ (la durée moyenne de service est donc $1/\mu$), le temps d'attente est :

$$T_a = \frac{\lambda}{\mu(\mu - \lambda)}. \quad (2.26)$$

Le caractère aléatoire de la durée de service introduit donc un doublement du temps d'attente moyen.

2.3.4. De multiples interactions : de l'arc au réseau

Les paragraphes précédents ont mis en évidence la multiplicité des mécanismes de congestion en transport et des modèles permettant de les représenter. Pour le moment, nous n'avons présenté ces modèles et mécanismes que dans le cadre d'un arc unique, comme isolé du reste du monde. Or, les déplacements réalisés par les usagers empruntent généralement plusieurs arcs reliés en un ou plusieurs réseaux. Nous allons donc ici montrer comment ces éléments se combinent et interagissent entre eux.

Nous avons précédemment évoqué deux cas d'interaction : les « ondes de choc » de transition et le phénomène de *spillback*. Le premier cas correspond au passage d'une forme de congestion à une autre. Il est lié à une évolution du régime de trafic : lorsque, sur un arc, le flux devient très important, on peut passer d'une situation de congestion de flux à une situation de congestion de stock, où les véhicules forment une file d'attente. Cette transition prend la forme d'une « onde de choc » comme le montre le modèle LWR. Le deuxième cas est un engrenage de congestion : une file d'attente de véhicules sur un arc peut, par l'intermédiaire du phénomène de *spillback*, engendrer une réduction de la capacité d'un autre arc, et accentuer le phénomène de congestion (qu'il s'agisse d'une congestion de flux ou de stock) qui y a lieu. La représentation de ce type de phénomènes est très complexe.

Nous allons à présent évoquer deux autres phénomènes mettant en évidence l'importance des interactions : le cas des intersections, qui aborde le rôle des interactions dans un cadre simple, et la question du stationnement. Enfin, nous soulignerons la difficulté de prendre en compte ces interactions dans le cadre d'une modélisation opérationnelle en abordant les questions d'équilibre des flux sur un réseau.

2.3.4.1. Les intersections

D'une manière générale, la modélisation d'une intersection combine un modèle de goulot et un modèle de file d'attente aléatoire. C'est donc une bonne illustration d'une interaction entre plusieurs formes de congestion, ici les congestions de stock déterministes et stochastiques. En effet, le retard moyen subi par un véhicule se présentant à une intersection à feux est la somme d'un retard déterministe, lié à l'alternance des périodes de « rouge » et de « vert », et d'un retard aléatoire, conséquence de l'arrivée stochastique des véhicules.

Plus précisément, le retard déterministe peut être obtenu par un modèle de goulot particulier, où la capacité passe alternativement de 0 pendant le temps de rouge à S pendant le temps de vert. On suppose par ailleurs que le cycle de durée c se décompose en un temps de vert g et un temps de rouge r . La capacité moyenne à l'entrée est donc $\kappa = gS/c = pS$, en notant $p = g/c$ la proportion de temps de vert dans le cycle.

Le retard déterministe moyen se calcule à partir du retard total, obtenu en comparant, au cours d'un cycle, le cumul des entrées et le cumul des sorties du carrefour :

$$T_d = \frac{r^2}{2c(1 - pR)}, \quad (2.27)$$

où $R = qc/gS = q/\kappa$ est le taux de charge net en entrée, c'est-à-dire le rapport entre le débit en entrée et la capacité moyenne.

En ce qui concerne le retard aléatoire, on modélise, comme nous l'avons déjà évoqué, une intersection à feux par un modèle (M/D/1) : les arrivées se font selon un processus de Poisson¹² et le temps de service est constant (Cohen, 2005). Par ailleurs, la discipline de service est de type FIFO. Plus précisément, les paramètres du modèle sont les suivants :

- les arrivées suivent une loi exponentielle de paramètre $\lambda = q$, le débit des arrivées ;
- le taux de service est constant et correspond à la capacité nette de l'entrée $\kappa = gS/c$, avec $\kappa > q$;

12. Newell (1965) montre que la loi choisie pour représenter le processus d'arrivée ne revêt pas une grande importance.

Le temps moyen d'attente en file s'écrit donc :

$$T_a = \frac{\lambda}{2q(q - \lambda)} = \frac{R^2}{2q(1 - R)}. \quad (2.28)$$

Le temps moyen de traversée de l'intersection¹³ est donc la somme de (2.27) et (2.28) :

$$T_c = T_d + T_a = \frac{r^2}{2c(1 - pR)} + \frac{R^2}{2q(1 - R)}. \quad (2.29)$$

Le calcul de la dérivée de T_c par rapport à q étant positive, T_c est, comme attendu, une fonction croissante du flux d'entrée, caractéristique d'un phénomène de congestion.

Ce résultat s'applique, moyennant quelques modifications, à d'autres situations similaires à l'origine de files d'attente. En particulier, en transport collectif, l'attente avant la montée en véhicule s'y apparente : le temps de présence à quai du véhicule correspond au temps de vert, la durée du cycle étant l'intervalle entre deux véhicules. En revanche, la discipline de service peut éventuellement ne pas être une FIFO stricte.

Par ailleurs, on peut obtenir un résultat similaire pour des intersections sans feux (avec règles de priorités). Le modèle de file d'attente généralement utilisé dans ce cas est le modèle (M/M/1) : la loi de passage d'un véhicule venant de l'axe non prioritaire dépend de la loi des arrivées sur l'axe prioritaire (Troutbeck, 2000).

Le cas des intersections illustre donc une interaction relativement aisée à modéliser, puisque les deux phénomènes en jeu sont purement additifs. Toutefois, il ne faut pas perdre de vue que nous ne nous sommes ici intéressé qu'à une intersection isolée, sans tenir compte de ce qui se passe sur les arcs adjacents. La mise en réseau de plusieurs intersections peut se révéler complexe à modéliser, malgré le caractère additif des deux formes de congestion en jeu ici : la chronologie de l'écoulement peut rapidement compliquer la modélisation, tant conceptuellement qu'opérationnellement.

2.3.4.2. Le stationnement

Les usagers de la voiture doivent garer leur véhicule avant de prendre part à leurs activités ou de poursuivre leur déplacement avec un autre mode. Le stationnement constitue à ce titre un élément important du système de transport (Young, 2000).

De plus, la recherche d'une place de stationnement est elle-même sujette à un phénomène proche de la congestion : un bien commun, ici l'ensemble des places de

13. Il s'agit, à un coefficient correcteur multiplicatif près, de la formule simplifiée de Webster, bien connue des ingénieurs chargés de la régulation des feux.

stationnement, subit la « tragédie des biens communs » (Hardin, 1968). Celle-ci est accentuée par le caractère spatial du stationnement : non seulement cette ressource est, dans son ensemble, trop utilisée (Arnott et Rowse, 1999), mais son utilisation est distribuée inefficacement dans l'espace, les lieux les plus attractifs subissant une surexploitation bien plus intense que les lieux moins désirables (Anderson et De Palma, 2004).

Si l'analyse et la modélisation du stationnement sont complexes et font intervenir des mécanismes assimilables à une forme particulière de congestion, nous souhaitons plutôt insister ici sur les interactions entre le stationnement et les différentes formes de congestion évoquées précédemment.

Ainsi, les véhicules cherchant une place de stationnement ou effectuant une manœuvre pour s'y ranger gênent la circulation des autres véhicules et accentuent le phénomène de congestion. Ce type d'interaction peut avoir des répercussions importantes sur le système de transport dans son ensemble, et ce, par plusieurs mécanismes.

Tout d'abord, la recherche d'une place de stationnement augmente les flux sur les arcs. Plusieurs auteurs (Arnott et Rowse, 1999 ; Arnott et Inci, 2006) indiquent que plus de la moitié des véhicules circulant dans les centres-villes pendant les heures de pointe cherchent une place de stationnement. D'autres avancent des chiffres plus modestes, et Shoup (2006), compilant plusieurs études, trouve un intervalle s'étalant entre 8 % et 74 %. Autrement dit, une partie importante des flux en centre-ville est liée à la recherche de stationnement. La contrainte de capacité sur les places de stationnement se répercute sur le réseau de transport, en augmentant les flux qu'ils doivent supporter.

Ensuite, les manœuvres effectuées par les conducteurs pour entrer ou sortir de leur place de stationnement réduisent temporairement la capacité de l'arc sur lequel ils se trouvent. Dans le cas de rues étroites de centre-ville, il peut en résulter, surtout si ces manœuvres sont fréquentes, un goulot sur l'arc considéré, menant à une situation d'hypercongestion (Lindsey et Verhoef, 2000). Anderson et De Palma (2004) modélisent ce phénomène de manière simplifiée, en supposant que la présence de véhicules se garant à une distance x du CBD diminue la vitesse des véhicules circulant entre le CBD et x . D'autres modélisent de manière statique cette interaction par l'intermédiaire, d'une part, de *facteurs d'absorption*, certains véhicules étant « absorbés » par la rue, et d'autre part par une modification de la fonction de temps de parcours de l'arc, pour rendre compte des processus d'entrée et de sortie des places de stationnement (Young, 2000).

La prise en compte de manière réaliste et détaillée, sur un réseau de transport complet, de l'interaction entre stationnement et circulation des véhicules n'est aujourd'hui possible que dans le cadre de modèles de microsimulation, en l'absence

d'un traitement déterministe. Toutefois, il est difficile de tirer des conclusions fermes à partir de tels modèles, en particulier parce qu'ils n'assurent en rien l'existence d'un équilibre des flux.

Nous allons maintenant nous pencher sur cette question de l'équilibre sur un réseau. L'obtention d'un tel équilibre peut s'avérer, comme nous allons le voir, rapidement complexe, alors même que les interactions entre plusieurs formes physiques de congestion sont généralement négligées.

2.3.4.3. L'équilibre des flux sur un réseau

Sur un même arc peuvent circuler des usagers qui suivent des itinéraires différents. Ils proviennent d'arcs différents en amont et emprunteront des arcs différents en aval. Autrement dit, dans le cadre d'un réseau de transport, les niveaux ainsi que les formes de congestion présents sur les arcs sont interdépendants. Ils sont liés aux demandes de déplacements des usagers. C'est pourquoi la question de l'affectation des flux sur un réseau de transport a été l'objet de très nombreuses attentions depuis les travaux phares de Wardrop (1952) et Beckmann *et al.* (1956).

La méthode d'obtention des flux d'équilibre sur un réseau a d'abord été formulée dans le cas statique. Ainsi, des formes fonctionnelles directement dérivées de courbes débit-vitesse empiriques, employées au sein de modèles d'affectation routière, fournissent, pour chaque arc du réseau, le temps de franchissement auquel font face les usagers qui l'empruntent. Ces temps influencent les choix des usagers, qui peuvent modifier leurs itinéraires, et donc influent, en retour, sur les temps de parcours. La situation d'équilibre émerge de ces ajustements progressifs.

Plusieurs définitions de l'équilibre existent, même si l'une des plus couramment employées est celle fournie par le premier principe de Wardrop (1952) :

« Les temps de parcours sur tous les itinéraires effectivement empruntés sont égaux, et inférieurs à ceux que subirait un véhicule unique sur n'importe quel itinéraire non emprunté¹⁴. »

Ce principe peut être, de façon quasi équivalente, exprimé de la façon suivante, qui souligne l'aspect comportemental du choix de l'équilibre d'affectation (Holden, 1989) :

« Le trafic tend vers une situation d'équilibre, dans laquelle aucun conducteur ne peut réduire son temps de parcours en modifiant son choix d'itinéraire¹⁵. »

Autrement dit, l'équilibre de Wardrop est un équilibre de Nash. Si aucune taxe n'est prélevée, il s'agit de l'équilibre à l'*optimum de l'utilisateur*, qu'il convient de bien

14. Traduction de l'auteur.

15. *Idem.*

distinguer d'un équilibre à l'*optimum social*, correspondant à une situation où c'est la somme des temps passés par l'ensemble des usagers qui est minimisée, et non le temps passé par chaque usager. Pigou (1920) (cité par Mohring, 1999) avait déjà identifié ces deux types d'équilibre :

« Supposons qu'il y ait deux routes ABD et ACD menant toutes deux de A à D. Laissez à lui-même, le trafic se distribuerait de manière à ce que la gêne subie en conduisant une charrette "représentative" le long des deux routes soit identique. Mais, dans certaines circonstances, il serait possible, en transférant quelques charrettes de la route B à la route C, de diminuer grandement la gêne subie par les usagers restant sur B, tout en n'augmentant que très légèrement la gêne subie sur C¹⁶. »

Dans tous les cas, le problème de la recherche de l'équilibre est équivalent à un problème d'optimisation (Beckmann *et al.*, 1956). Cette distinction entre optimum de l'utilisateur et optimum social est la traduction du coût économique de la congestion : le coût social diffère du coût privé.

Dans le cas de flux non stationnaires, autrement dit dans le cas dynamique, la recherche de l'équilibre est plus complexe, conceptuellement et mathématiquement (Lindsey et Verhoef, 2000). Les difficultés résident justement dans l'interaction dynamique entre les arcs, notamment en termes de chronologie : le chargement des arcs doit se faire de manière chronologique, en suivant les itinéraires des véhicules, en s'assurant qu'une discipline de type FIFO est maintenue (Leurent, 2003 ; Leurent et Aguiléra, 2009). De plus, au choix d'itinéraire présent également en affectation statique, s'ajoute le choix de l'horaire de départ, ce qui augmente le nombre de degrés de liberté, et rend la convergence à la fois plus lente et plus difficile à caractériser.

Par ailleurs, la notion d'équilibre repose sur l'hypothèse forte que les usagers ont une information parfaite, autrement dit qu'ils connaissent, en temps réel, les temps de parcours sur l'ensemble des itinéraires du réseau. C'est néanmoins une hypothèse couramment admise, même si des travaux introduisant un déficit d'information existent, et mettent en évidence une dégradation de la situation dans ce cas (Daganzo et Sheffi, 1977 ; De Palma et Picard, 2006 ; Leurent et Nguyen, 2010).

La multiplicité des formes physiques de la congestion complexifie, nous venons de le voir, la modélisation de ce phénomène à l'échelle d'un réseau. Cette difficulté impose des choix de modélisation qui, par essence, réduisent la complexité conceptuelle et mathématique, mais ce, au prix d'une perte de réalisme préjudiciable. En ce qui concerne nos propres travaux, le chapitre suivant nous permettra de présenter, en les justifiant autant que possible, les choix que nous avons faits.

16. *Idem.*

2.4. Les coûts économiques de la congestion

Nous avons, dans les deux sections précédentes, constaté la pluralité des formes que prend la congestion, aussi bien dans la manière dont les usagers des transports la vivent que dans les mécanismes qui la font émerger. Mais si l'étude de la congestion revêt une telle importance, qui justifie les nombreuses études qui lui sont consacrées, c'est d'une part parce qu'elle constitue un coût économique pour les usagers, et d'autre part parce que ce coût privé est inférieur au coût social. La congestion est donc une externalité, qui entraîne surconsommation et problèmes d'équité.

L'objectif de cette section est de préciser la définition du coût économique de la congestion. Plus particulièrement, nous distinguerons l'analyse de ce phénomène en termes de coûts directs, privés et sociaux, à court et à long termes d'une part, et d'autre part en termes de coûts indirects, en soulignant la difficulté de prendre en compte ces derniers. Nous insisterons, tout au long de la section, sur l'importance d'une mesure adéquate de ces coûts.

Rappelons tout d'abord que la congestion est une externalité. Plus précisément, le fait que la qualité de service — la vitesse de circulation ou le temps de parcours — dont bénéficie *l'ensemble des utilisateurs* dépende du nombre d'utilisateurs du service est le premier signe d'une externalité d'usage, ou plus précisément, puisque la qualité de service se dégrade lorsque le nombre d'usagers augmente, d'une dés-économie externe de consommation (Picard, 2007). Le caractère *involontaire* de la dégradation de service que l'utilisateur supplémentaire impose aux autres est le deuxième point qui en fait une externalité (Verhoef, 1994).

La présence d'une telle externalité est liée au fait que la capacité d'un système est un bien public particulier : c'est un bien commun, c'est-à-dire non excluable mais rival, qui est donc sujet, sans tarification, à une utilisation trop intensive (Hardin, 1968).

Supposons que lorsqu'un usager supplémentaire utilise le système, il diminue la qualité du service délivré par le système d'une valeur δq . Chaque usager du système, y compris l'utilisateur supplémentaire, subit cette baisse de qualité, si bien que si le système comprend $N + 1$ usagers, la présence de l'utilisateur supplémentaire entraîne, de manière agrégée, une baisse globale de qualité du service de $(N + 1)\delta q$. En ce qui le concerne directement cependant, l'utilisateur supplémentaire ne subit qu'une perte δq par rapport à la qualité de service à laquelle il s'attendait. Autrement dit, son choix d'utiliser le service entraîne pour la « société » un coût $N\delta q$ qu'il ne paie pas (il ne paie que δq). Il s'agit donc bien d'une externalité au sens où une partie des interactions entre les agents se déroulent *hors marché*, certains effets restant sans compensation. Il en résulte des inefficacités, au sens où l'équilibre de Nash, qui émerge des interactions entre les agents, diffère de la situation optimale, où le

bien-être collectif est maximisé.

Il convient de noter néanmoins que la congestion, en termes de perte de temps, c'est-à-dire de manière directe, est une externalité *de club*, c'est-à-dire qu'elle n'impacte que les usagers qui sont eux-mêmes sources de congestion (Buchanan, 1965 ; Verhoef, 1994). En cela, elle est très différente d'une externalité *environnementale*, comme la pollution ou le bruit. Mais nous verrons que ces deux dernières externalités peuvent être des effets secondaires indirects de la congestion.

2.4.1. Les coûts directs : du court terme au long terme

Les infrastructures de transport requièrent généralement des investissements lourds, qui représentent des coûts fixes importants. Dans le cas routier, en particulier, c'est le cas pour la construction d'une nouvelle route, ou pour l'augmentation de la capacité d'une route existante. On distingue donc, comme il est classique de faire en économie industrielle, les coûts de court terme et de long terme. Les premiers correspondent à un raisonnement à *capacité fixe*, tandis que les seconds prennent en compte les investissements nécessaires à l'évolution de cette capacité.

2.4.1.1. Coût de court terme et optimum de l'usager

De manière générale, le coût subi par un usager se déplaçant sur une route de longueur donnée peut se décomposer en un coût fixe c_f et un coût lié à la congestion c_g (Small et Verhoef, 2007). Le coût fixe se décompose lui-même en c_{f0} qui rend compte des frais de carburant et de maintenance, et en un coût temporel fixe $v_T T_f$, dépendant de la valeur du temps v_T et du temps de parcours à vide T_f . Le coût lié à la congestion se décompose¹⁷ quant à lui en un coût lié aux pertes de temps, $v_T(T - T_f)$ et un éventuel coût de retard (ou d'avance) c_s , lié à la variabilité du temps de parcours. Autrement dit, le coût c , appelé coût généralisé, s'écrit :

$$c = c_f + c_g = (c_{f0} + v_T T_f) + (v_T(T - T_f) + c_s). \quad (2.30)$$

Cette formulation met en évidence le coût de la congestion pour un usager de la route. En termes économiques, il est le résultat de la rencontre d'une demande, le nombre d'usagers souhaitant emprunter la route, et d'une offre, le temps de parcours de la portion de route. La fonction de demande est classique, décroissante avec le coût. La fonction d'offre, quant à elle, est particulière. Walters (1961) est le premier à suggérer d'employer la courbe débit - temps de parcours, tirée du diagramme fondamental du trafic (Figure 2.2) comme courbe d'offre¹⁸ de court terme

17. Le coût de la congestion est supposé dans ce cadre uniquement temporel. Le coût monétaire dû à une usure prématurée liée à l'utilisation prolongée du véhicule n'est pas pris en compte.

18. Comme Gaudry (2007) le note, une courbe débit - temps de parcours n'est pas une véritable

(à capacité fixe), en régime stationnaire. Autrement dit, cela revient à considérer que l'équation (2.30) s'écrit :

$$c(Q) = c_{f0} + v_T T(Q). \quad (2.31)$$

Toutefois, une telle courbe présente l'inconvénient de ne pas correspondre à la définition classique d'une courbe de coût, à savoir une courbe qui fournit le coût minimal de production d'une quantité donnée du bien produit. Une des conséquences est qu'elle peut engendrer plusieurs équilibres possibles, comme l'illustre la Figure 2.4, sur laquelle l'équilibre x est un équilibre classique, tandis que y et z correspondent à des équilibres en situation d'hypercongestion.

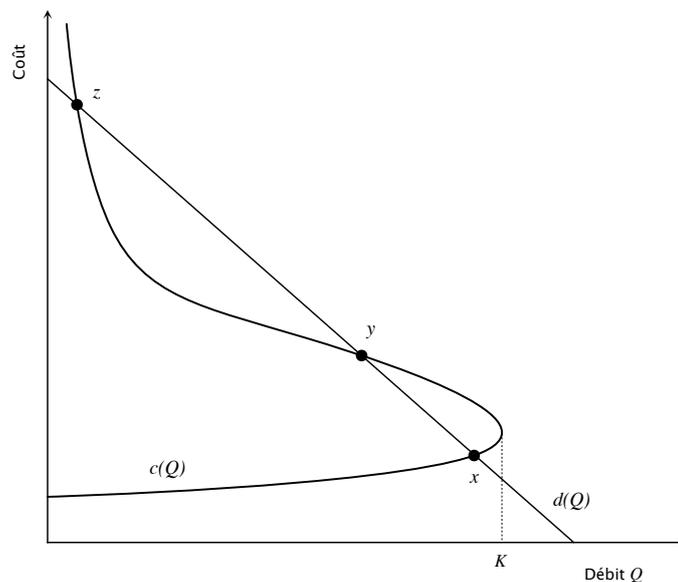


FIGURE 2.4.: Fonction d'offre et équilibres multiples. *Source* : Verhoef (1999)

Cette figure amène plusieurs remarques essentielles à notre propos. Tout d'abord, dans le cadre statique dans lequel nous nous sommes placé, la courbe de coût moyen (c'est-à-dire la moyenne individuelle des coûts subis par les usagers empruntant l'infrastructure) est également la courbe de coût marginal privé. Elle diffère de la courbe de coût marginal social. C'est la conséquence du fait que la congestion est

courbe d'offre, car la qualité du bien fourni varie avec la quantité, la durée d'un déplacement n'étant pas directement assimilable à un prix, même temporel. De même, nous attirons l'attention du lecteur sur le fait que parler de courbe d'offre peut être trompeur, car si le temps de parcours d'une infrastructure peut être assimilée à une offre, ce n'est pas l'offreur (l'opérateur) qui en assume les coûts, mais l'utilisateur. Autrement dit, les usagers sont à la fois les demandeurs et les offreurs. Malgré ces remarques, nous assimilerons néanmoins, comme la majorité des auteurs, les courbes débit - temps de parcours à des courbes d'offre de manière à pouvoir utiliser les outils classiques de la théorie économique.

une externalité technologique : le coût moyen subi par un usager dépend directement des choix des autres usagers de se déplacer ou non.

Ensuite, parce que les usagers ne prennent en compte que le coût moyen (ou coût marginal privé), sans avoir accès au coût marginal social¹⁹, les points d'équilibre se situent à l'intersection entre cette courbe de coût moyen et la courbe de demande (Mohring, 1999).

Enfin, la question de la stabilité des deux points d'équilibre sur la branche hypercongestionnée n'est pas encore tranchée, mais Verhoef (2001) a montré, en utilisant un modèle dynamique de poursuite, que seul l'équilibre x était dynamiquement stable. Autrement dit, les situations d'équilibre y et z ne représenteraient que des situations transitoires, valables en un point de la route à un instant donné. L'hypothèse, évoquée par Mohring (1999), selon laquelle certaines agglomérations pourraient être dans un tel état d'équilibre la plupart du temps semble donc erronée. En conséquence, la courbe de la Figure 2.4 doit être remplacée, en régime stationnaire, par une courbe calquée sur la précédente jusqu'au volume κ , et se terminant en une demi-droite verticale en ce point. En se restreignant à une période temporelle donnée, on peut approcher cette courbe par une relation débit - vitesse moyennée dans le temps, telle que la relation BPR (Small et Verhoef, 2007), qui autorise les dépassements temporaires de capacité.

Dans ce cadre, l'équilibre entre la demande et l'offre sur une portion de route peut être représenté comme sur la Figure 2.5. Nous utiliserons, dans la suite, ce formalisme et cette représentation de la fonction de coût de court terme.

La situation d'équilibre diffère de la situation optimale du fait du caractère externe de la congestion. Cette situation optimale est atteinte lorsque le prix payé par chaque usager est égal au coût marginal social. Or, en présence d'une externalité technologique comme la congestion, ce coût marginal social est supérieur au coût moyen, puisqu'il est la somme de ce coût moyen et d'un coût externe correspondant au coût que l'utilisateur marginal fait peser sur l'ensemble des autres usagers de la route. La Figure 2.6 compare les situations d'équilibre et d'optimum. Elle met en particulier en évidence le fait qu'à l'équilibre, en l'absence de tarification de l'externalité que constitue la congestion, le nombre d'utilisateurs et le temps de parcours sur l'arc sont trop élevés par rapport à la situation optimale.

Par ailleurs, si équilibre et optimum sont sensibles à la forme de la fonction de coût, ils le sont également à celle de la fonction de demande. Or, la forme de la fonction de demande, en particulier son niveau et son élasticité, dépendent, en transport, du motif du déplacement, et donc de son horaire : à l'heure de pointe du

19. Comme l'indique Mohring (1999), le coût marginal social correspond au coût marginal auquel ferait face une entreprise en situation de concurrence parfaite dont le bien produit serait le débit.

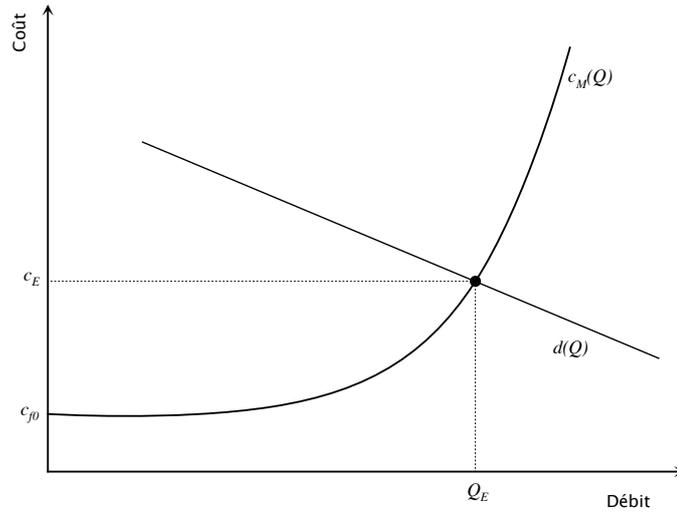


FIGURE 2.5.: Fonctions d'offre et de demande sur une route. *Source* : Mills et Hamilton (1994)

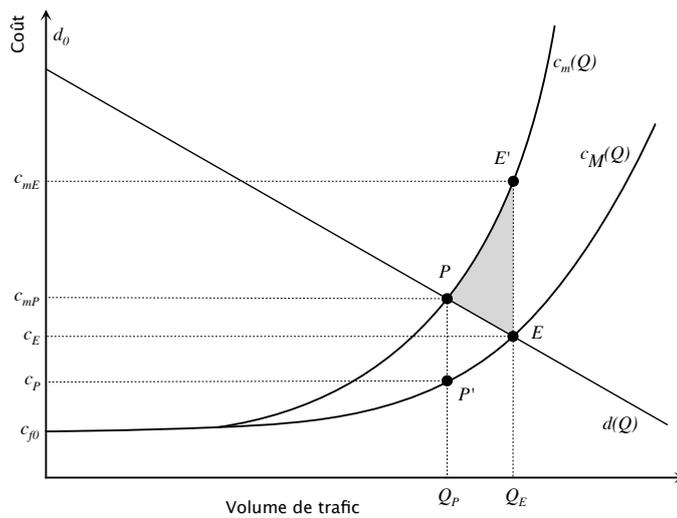


FIGURE 2.6.: Comparaison de l'équilibre et de l'optimum. *Source* : Small et Verhoef (2007)

matin, où les déplacements se font majoritairement pour motif domicile-travail, la demande sera vraisemblablement plus forte et moins élastique qu'en heure creuse le week-end, où les déplacements sont moins nombreux et majoritairement pour des motifs loisirs. Ainsi, la Figure 2.7 illustre, pour deux courbes de demande d_1 et d_2 , respectivement forte et inélastique, et faible et élastique, les situations d'équilibre et d'optimum. Elle souligne le fait que d'une part lorsque la demande est forte

un niveau plus élevé de congestion est acceptable à l'optimum et que d'autre part lorsqu'elle est peu sensible au coût, l'écart entre le volume optimal et le volume d'équilibre se réduit.

Par conséquent, une analyse empirique du coût de la congestion doit distinguer les périodes temporelles considérées, de manière à pouvoir prendre en compte la variabilité temporelle de la demande.

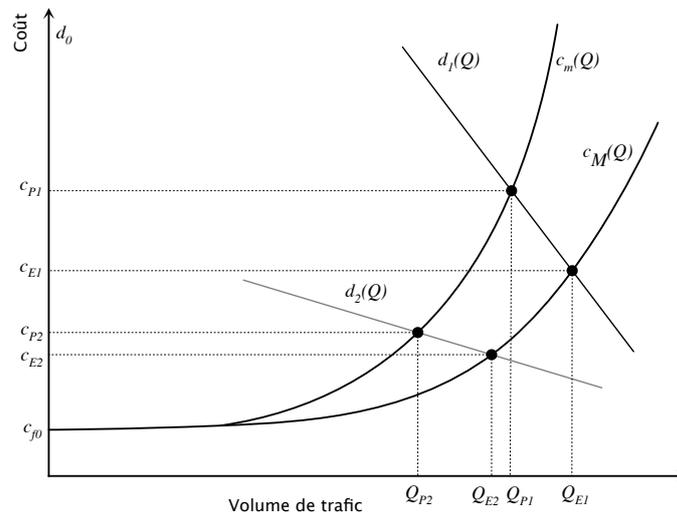


FIGURE 2.7.: Équilibre et optimum pour deux motifs distincts. *Source* : Auteur

Nous pouvons, à partir de la Figure 2.6, donner plusieurs définitions du coût de la congestion (Newbery, 1990 ; Prud'homme, 1999a), et préciser leur validité comme mesure économique de la congestion.

Coût naïf de la congestion Le coût total de congestion est parfois exprimé de manière naïve de la façon suivante :

$$C = (c_E - c_{f0})Q_E. \quad (2.32)$$

Cette expression correspond au temps total « perdu » par les usagers à l'équilibre, par rapport à une situation *sans congestion*. Elle se réfère donc à une situation où la route n'est employée par aucun usager. Or, cette situation n'est en rien optimale, car elle ne permet en fait à aucun usager de se rendre à sa destination. Les routes sont faites pour être utilisées, car le transport est un bien particulier qui correspond à une consommation *secondaire*, au sens où il n'est consommé qu'*en vue d'effectuer une autre activité*. Si les routes sont vides, cela signifie que les usagers ne se

déplacent pas et qu'ils n'effectuent pas les activités qu'ils souhaiteraient²⁰. Cette situation n'est donc pas économiquement désirable. Il nous semble, à l'instar d'auteurs comme Prud'homme, qu'évaluer le coût de la congestion par cette grandeur n'a pas d'intérêt économique, au sens où cela revient à oublier l'objet initial du transport. De plus, si la congestion est un coût social, c'est parce qu'elle est une externalité. Or l'expression précédente n'en tient pas compte. Enfin, comme nous le montrerons dans le chapitre 4, cette méthode ne fournit pas un ordre de grandeur correct du coût de la congestion (les chiffres qu'elle fournit sont généralement dix fois trop élevés).

Pourtant, comme le note le rapport ECMT (2007), malgré son absence de sens économique, cette méthode de calcul reste d'usage dans un nombre significatif de pays, même si la diffusion des méthodes économiques d'évaluation de la congestion tendent à faire disparaître cette approche. À titre d'illustration, c'est la méthode longtemps retenue par la Commission Européenne pour estimer le coût de la congestion à l'échelle de l'Europe (CEC, 1995).

Coût fiscal de la congestion Une autre manière, souvent employée, d'aborder le coût de la congestion est de considérer la taxe qu'il faudrait imposer aux usagers pour passer de la situation d'équilibre à la situation optimale. Cette taxe d'internalisation correspond, en termes temporels, à la longueur du segment PP' sur la Figure 2.6²¹. En effet, à ce niveau de taxe, les usagers dont la disposition à payer est inférieure à c_{mP} renoncent à prendre la route, si bien que le volume effectivement présent sur l'arc est Q_P , le volume à l'optimum. Le produit de la taxe qui serait prélevée est parfois considéré comme le coût de la congestion. Autrement dit, celui-ci s'exprimerait par :

$$C = (c_{mP} - c_P)Q_P. \quad (2.33)$$

Toutefois, ce coût fiscal (virtuel, lorsqu'aucune taxe n'est effectivement appliquée) surestime le coût de la congestion, car il néglige le surplus des usagers qui prennent la route à l'équilibre, et y renoncent en raison de la taxe.

Coût économique de la congestion En réalité, une partie de la congestion subie par les usagers est optimale, au sens où elle est le résultat d'un arbitrage *social* entre le temps perdu par les usagers et la disposition à payer des usagers pour se rendre à leur destination. Le coût de la congestion purement *inefficace*, autrement dit le coût économique de la congestion, peut donc être évalué par la différence entre le

20. Nous considérons ici implicitement que les usagers ne peuvent se déplacer qu'en voiture. L'existence d'autres modes de transport complique le schéma sans l'invalider : considérer que l'optimum consiste en une non-utilisation des réseaux semble injustifiable.

21. La valeur de la taxe d'internalisation correspond à l'écart entre le coût marginal social et le coût privé à l'optimum.

surplus dégagé en situation optimale et le surplus dégagé en situation d'équilibre. Le premier correspond à l'aire de $c_P d_0 P P'$, qui se décompose elle-même en un surplus des usagers en situation optimale et en un revenu de la taxe d'internalisation, tandis que le second correspond à l'aire de $c_E d_0 E$. Le coût économique de la congestion à l'équilibre vaut donc :

$$\mathbf{C} = (c_E - c_P)Q_P - \frac{1}{2}(c_{mP} - c_E)(Q_E - Q_0). \quad (2.34)$$

En effet, à l'équilibre, parmi les Q_E usagers présents, Q_P seraient également présents à l'optimum, mais subiraient un coût moindre, de c_P au lieu de c_E . De plus, $Q_E - Q_P$ usagers peuvent utiliser la route, alors qu'ils ne le pourraient pas à l'optimum, dégageant un surplus de $1/2(c_{mP} - c_E)(Q_E - Q_P)$.

Ce coût économique est également mesuré par le classique triangle de Harberger, grisé sur la Figure 2.6, qui représente la perte sèche liée à la congestion. Ce triangle met en évidence deux caractéristiques importantes. Tout d'abord, plus l'écart entre le coût marginal social et le coût moyen à l'équilibre est élevé, plus la perte sèche est importante. Autrement dit, la mesure, en situation d'équilibre, du coût marginal social, et sa comparaison avec le coût privé sont essentiels pour évaluer le coût de la congestion. Ensuite, toutes choses égales par ailleurs, la perte sèche augmente avec l'élasticité de la demande. En effet, si la demande est inélastique, les usagers sont prêts à subir des niveaux élevés de congestion, car ils tirent de leur déplacement une forte utilité (à la destination). En conséquence, la perte sociale entraînée par la congestion est faible dans ce cas. Ainsi, pour un même niveau de congestion, la perte sociale en termes temporels sera plus importante s'il s'agit de déplacements de loisir plutôt que de déplacements professionnels. Néanmoins, cela est en bonne partie compensé par les volumes en jeu, beaucoup plus élevés à l'heure de pointe du matin qu'en heures creuses.

La mesure du coût économique de la congestion nous a donc montré que la situation optimale pour les usagers n'est pas l'absence de congestion, mais un niveau inférieur au niveau d'équilibre. Par ailleurs, nous avons mis en évidence que le niveau optimal de congestion varie *dans le temps* au cours d'une même journée.

2.4.1.2. Coût de long terme et optimum du gestionnaire

Les paragraphes précédents ont analysé le coût de la congestion dans une perspective de court terme, où l'infrastructure existe et sa capacité est fixe. La prise en compte des coûts de construction et de maintenance de la capacité routière permet d'obtenir la fonction de coût moyen total de court terme, comme l'illustre la Figure 2.8.

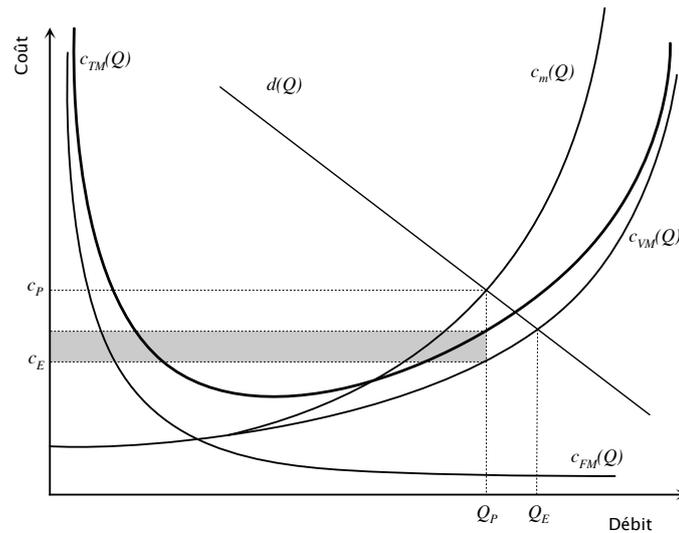


FIGURE 2.8.: Coût moyen total de court terme. *Source* : Mills et Hamilton (1994)

La courbe de coût moyen mise en évidence précédemment y est notée $c_{VM}(Q)$ pour indiquer qu'il s'agit de la partie variable du coût total, tandis que le coût moyen lié au financement de l'infrastructure est noté $c_{FM}(Q)$ et est décroissant. La courbe de coût moyen total $c_{TM}(Q)$ est la somme de ces deux courbes. L'aire grisée représente le coût d'infrastructure à l'optimum, tandis que le produit de la taxe d'internalisation est donné par $(c_P - c_E)Q_P$. Autrement dit, dans ce cas, le produit de la taxe est supérieur au coût de la route.

Mohring et Harwitz (1962) ont montré que si la technologie de construction des routes se fait à rendements d'échelle constants, la capacité optimale de l'infrastructure est telle que le revenu de la taxe d'internalisation couvre exactement le coût de la route. Graphiquement, cela se traduit par le fait que les courbes de demande, de coût marginal social et de coût moyen total se rejoignent en un point.

Dans ce cadre, l'optimum de long terme, une fois la capacité ajustée, varie, pour une même demande, selon les coûts de construction et de maintenance de l'infrastructure. Plus précisément, comme l'illustre la Figure 2.9, un coût de construction élevé (coût du sol, difficulté de construction liée à la densité d'occupation, etc.), par exemple en centre-ville, conduit à choisir une capacité plus faible que si la construction est peu onéreuse, comme en périphérie. L'optimum O_1 , correspondant au centre-ville, conduit à un niveau de congestion plus élevé et un trafic plus faible que l'optimum O_2 correspondant à la périphérie.

Des travaux ont cherché à estimer la vitesse optimale de circulation en heure de pointe, compte tenu des coûts en capital des infrastructures. Ainsi, Starrs et Starkie (1986, cité par Small et Verhoef, 2007), s'appuyant sur le modèle de Keeler et Small (1977), trouvent, pour les voies artérielles du centre-ville d'Adélaïde, une vitesse de

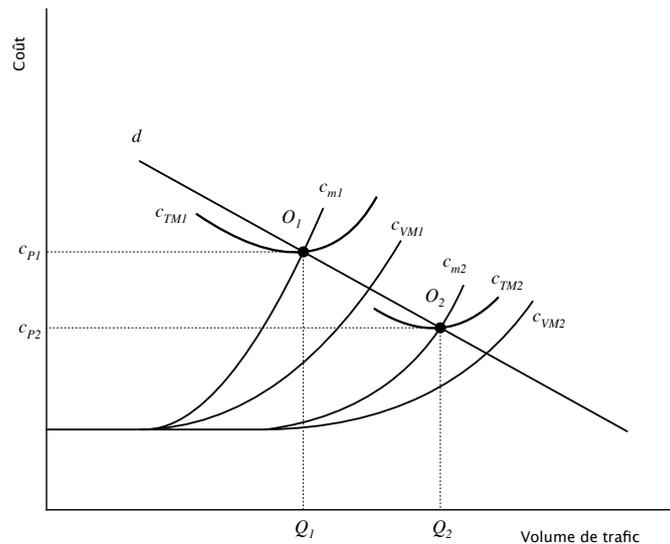


FIGURE 2.9.: Variation spatiale du niveau optimal de congestion. *Source* : Mills et Hamilton (1994)

38 km/h. Ce résultat montre deux choses : d'une part, la situation optimale, du point de vue du gestionnaire, n'est pas l'absence de congestion ; d'autre part, en centre-ville, le coût de construction justifie des niveaux élevés de congestion.

L'analyse du coût de long terme de la congestion nous a donc montré que, parallèlement au coût de court terme, l'optimum n'est pas exempt de congestion. De plus, nous avons mis en évidence que le niveau optimal de congestion varie *dans l'espace*.

2.4.2. Les coûts indirects : une difficile prise en compte

Si les coûts directs de la congestion, c'est-à-dire les coût temporels, peuvent être pris en compte et mesurés, notamment à partir du cadre d'analyse que nous avons présenté dans les paragraphes précédents, les impacts et donc les coûts indirects, en revanche, sont délicats à mesurer et à évaluer. La raison principale de cette difficulté provient, comme le souligne le rapport ECMT (2007), du fait que la majorité des impacts indirects de la congestion sont liés à l'usage de la route en général, si bien que déterminer la part de ces impacts réellement imputables à la congestion est une tâche difficile, qui nécessite la comparaison entre des situations avec et sans congestion. De plus, contrairement aux coûts directs, qui ne touchent que les usagers de la route, les coûts indirects sont souvent environnementaux et touchent donc d'autres usagers.

Dans la suite de ce mémoire, nous ne nous intéresserons pas directement à ces

coûts, si bien qu'une présentation exhaustive nous semble inutile. En revanche, nous souhaitons pointer quelques effets particuliers de la congestion sur certains impacts environnementaux, afin d'en montrer le caractère parfois paradoxal.

Sécurité routière Nous avons déjà évoqué, au 2.2.1.3, le fait que la probabilité d'un incident augmente avec la densité des usagers, mais que dans le même temps, la faible vitesse des véhicules en situation congestionnée diminue cette probabilité, ainsi que le gravité des accidents. L'accidentologie de la congestion apparaît donc délicate, d'autant que le recueil des données l'est encore plus.

Consommation de carburant et pollution La consommation de carburant au kilomètre augmente avec la congestion. En effet, elle varie de manière non monotone avec la vitesse (ECMT, 2007). Elle est minimale pour des vitesses comprises entre 60 et 80 km/h et les situations congestionnées entraînent des vitesses généralement inférieures. L'émission de CO₂ étant directement reliée à la consommation de carburant, elle augmente également en situation congestionnée.

C'est également le cas de la majorité des polluants, mais les NO_x font exception, car leur émission dépend directement de la température du moteur, et diminue donc à faible vitesse.

Toutefois, la dépendance des émissions à la vitesse est généralement évaluée à partir de vitesses moyennes. Or, les situations congestionnées se caractérisent souvent par des régimes de circulation alternant arrêts et accélérations. De même, le régime de fonctionnement du moteur, à chaud ou à froid, influence également les niveaux d'émission, si bien que les émissions diffèrent en début et en fin de déplacement. Autrement dit, estimer ces coûts indirects de la congestion nécessite de tenir compte de ces facteurs (ECMT, 2007).

Bruit Le bruit émis par les véhicules provient de deux sources distinctes : le moteur, qui domine aux faibles vitesses, et le mouvement du véhicule, en particulier le bruit de contact pneumatique/chaussée, qui devient prépondérant au-delà de 30 km/h (Leclercq, 2002). Autrement dit, la congestion aurait *a priori* tendance à réduire le bruit routier. Toutefois, deux éléments viennent perturber le schéma. D'une part, la densité des véhicules multiplie les sources de bruits ; d'autre part, les accélérations et décélérations associées aux niveaux de congestion élevés engendrent une irrégularité des émissions sonores préjudiciable, que l'usage immodéré du klaxon accentue encore.

Conséquences sur l'espace Les infrastructures de transport, ainsi que leur usage, ont des conséquences multiples sur les espaces environnant : consommation d'espace, effets de coupure et de fragmentation, effets de cicatrice pour les espaces

déshérités en bordure d'axes à forts trafics (Leurent, 2007). La congestion peut être de nature à accentuer certains de ces effets, ceux qui sont liés à l'intensité du trafic.

Productivité La congestion a des effets indirects sur la productivité d'une ville ou d'une région, en grande partie parce qu'elle engendre perte de temps et incertitudes sur les horaires, pour les travailleurs comme pour les marchandises. Nous avons abordé cette question dans le chapitre précédent, en montrant comment la congestion constituait une force de dispersion à l'échelle d'une agglomération. Nous ne reviendrons donc pas ici sur cet aspect des coûts de la congestion.

2.5. Conclusion

Nous avons vu, dans ce chapitre, que la congestion est un phénomène complexe, tant dans ses mécanismes que dans ses manifestations. Cette complexité se traduit par une grande difficulté à rendre compte de l'ensemble de ses modes d'action, et à mesurer l'ensemble des coûts économiques de la congestion. En particulier, nous avons insisté sur les multiples interactions possibles entre les différentes formes de congestion, psychologiques et physiques. Les phénomènes de remontées de file (*spillback*) peuvent avoir de sévères répercussions sur les trafics des arcs adjacents, et sont pourtant très rarement pris en compte. De même, les coûts indirects imputables à la congestion sont difficilement mesurables, et donc rarement mesurés.

Le panorama que nous avons présenté fait toutefois ressortir un certain nombre d'éléments permettant de fonder une étude spatialisée de la congestion, telle que nous nous proposons de mener dans les deux chapitres suivants. Ainsi, si la fiabilité revêt pour les usagers autant d'importance que le temps de trajet, le risque de variation imprévue du temps de trajet est fortement relié au niveau de saturation des infrastructures. Par ailleurs, les situations d'hypercongestion étant instables, une modélisation statique correctement calibrée permet de mesurer le coût de la congestion de manière relativement satisfaisante, et nous avons vu à quel point une telle approche est féconde en termes de calcul économique. Il demeure qu'une amélioration des capacités de modélisation du phénomène de congestion permettant de rendre compte des multiples interactions que nous avons mentionnées, dans un cadre suffisant simple pour s'adapter aux besoins du calcul économique, constituerait une avancée majeure.

Deuxième partie .

Les manifestations de la congestion du transport : une approche spécialisée

3

Méthodologie d'analyse de la congestion sur un réseau de transport

3.1. Introduction

L'objectif de ce chapitre est de déterminer et d'expliciter les dimensions à intégrer dans une démarche d'analyse territoriale de la congestion, de manière à faire progresser, sur cette question, la pratique des études de planification.

Le phénomène de congestion reste, par certains aspects, assez mal compris et analysé. Si le détail des mécanismes, tant physiques qu'économiques, qui en sont à l'origine, sont aujourd'hui bien connus et compris à un niveau microscopique (cf. chapitre 2), il n'en va pas nécessairement de même de ses aspects plus macroscopiques et de l'inscription spatiale et temporelle de ce phénomène. En effet, le transport est un bien localisé¹, tant du point de vue de l'offre, c'est-à-dire des infrastructures, et de l'accessibilité qu'elles permettent, que du point de vue de la demande, c'est-à-dire en ce qui concerne le besoin de mobilité des agents économiques, ménages ou entreprises. Une conséquence directe de ce constat est que les *effets* de la consommation de ce bien sont également localisés. En particulier, la congestion

1. Dans le temps comme dans l'espace.

des infrastructures de transport est sujette à une forte variabilité temporelle² et spatiale (Leurent, 2006 ; Leurent et Breteau, 2009). Ainsi, une infrastructure ne présentera pas les mêmes niveaux de congestion selon les moments de la journée (heure de pointe ou heure creuse), les jours de la semaine (jour ouvrable et week-end) et les périodes de l'année (vacances scolaires ou non). De même, deux infrastructures situées à proximité immédiate l'une de l'autre ne subiront pas les mêmes niveaux de congestion, selon la topologie du réseau, l'usage que les agents font de chacune des infrastructures et les caractéristiques propres à ces infrastructures. Il suffit de penser aux deux sens de circulation d'une seule infrastructure pour s'en convaincre.

Les effets de la variabilité temporelle de la congestion ont été largement étudiés, en particulier au niveau d'un arc par l'intermédiaire du modèle de goulot, à partir des travaux initiaux de Vickrey (1969), puis ceux de Arnott *et al.* (1993). Leurs analyses ont permis de mieux comprendre la nature fondamentalement dynamique de la congestion et d'en tirer des conséquences en terme de politique de correction. Ces éléments ont été présentés dans le chapitre 2.

En revanche, la variabilité spatiale de la congestion a été beaucoup moins systématiquement étudiée. Certains travaux sont toutefois à signaler, comme Riff Brems *et al.* (2002). Or, nous le verrons dans ce chapitre, le caractère fondamentalement non linéaire du phénomène de congestion a des effets importants sur la manifestation spatiale de la congestion sur un territoire. Etudier la variabilité spatiale de la congestion permet de mieux comprendre le fonctionnement du réseau dans son ensemble, d'identifier les difficultés et d'en prendre la mesure. Une approche spatialement différenciée permet en outre de mettre en évidence des problèmes de distribution spatiale des sources et des victimes de la congestion.

En améliorant la connaissance et la compréhension des manifestations spatiales de la congestion, nous visons à faire progresser la pratique des études de planification. Plus précisément, il s'agit de développer une méthodologie de diagnostic de la congestion sur un réseau routier, respectueuse à la fois de la dimension spatiale et de la dimension économique. Cela constitue le principal enjeu de ce travail et implique que la méthodologie que nous élaborons puisse concilier les impératifs pratiques de l'ingénieur-planificateur qui analyse un territoire et les méthodes et arguments théoriques des économistes. Autrement dit, notre démarche dans ce chapitre consiste à conjuguer les modèles épurés de l'économie des transports théorique avec les exigences d'une analyse territoriale de la congestion.

Nous avons donc pour objectif particulier dans ce chapitre d'explicitier les différentes dimensions à intégrer dans une telle analyse : la dimension spatiale (réseau,

2. Nous ne nous intéresserons pas ici directement à la question de la fiabilité. Si la fiabilité est clairement liée aux questions de variabilité, c'est une notion différente de variabilité. Il s'agit plutôt d'une variabilité imprévue, tandis que nous nous intéressons dans ce chapitre à une variabilité *en moyenne*.

territoire desservi), la dimension socio-économique (implantation de la demande, ses désirs et comportements), la dimension physique (formes physiques de la congestion, situation dans l'espace), la dimension économique (coûts et utilités pour les différents acteurs). Nous mobilisons par conséquent dans ce chapitre plusieurs disciplines, modélisation des transports, géographie et économie, et nous adoptons une approche globale constructive, au sens où elle vise à rapprocher pratique et théorie. Pour ce faire, il nous semble nécessaire de commencer par décrire de façon critique les différentes approches proposées par les disciplines en question, puis de les réunir à travers l'élaboration d'une méthodologie d'étude.

3.2. La représentation de l'ingénieur

3.2.1. Description à travers une plateforme de simulation

La représentation du fonctionnement d'un système de transport se fait couramment à partir d'une plateforme de simulation visant à affecter les flux de déplacements aux différents itinéraires possibles sur les réseaux. Cette plateforme de simulation constitue à ce titre un modèle du système de transport qui vise à approcher autant que possible la réalité des trafics observés tout en maintenant un niveau de complexité compatible avec les capacités de calcul d'une part, et avec les capacités des bases de données d'autre part.

Une plateforme de simulation de l'affectation sur un réseau de transport est constituée de deux jeux de données. Le premier, qui concerne la demande, est généralement issu d'un modèle de demande comportant les étapes de génération et distribution des déplacements et de choix modal. Ce jeu peut également directement provenir d'une enquête sur les déplacements des usagers³. En pratique, il s'agit d'une matrice⁴ fournissant, pour chaque couple Origine-Destination, le volume de flux sur le mode de transport considéré. La définition de cette matrice repose elle-même sur un zonage, autrement dit un découpage de la région représentée en zones de tailles variables, qui doivent correspondre le plus possible, à des unités élémentaires en termes d'usage des transports. Ainsi, en milieu urbain dense, un îlot peut constituer à lui seul une zone, tandis qu'en milieu périurbain, les zones seront vraisemblablement plus étendues. Par ailleurs, le découpage est généralement adapté aux besoins particuliers de l'étude. Ainsi, on procède à un redécoupage des zones autour d'un projet d'infrastructure, de manière à mieux

3. Sur les différents modes d'enquête, voir Ortúzar et Willumsen (2001) qui est l'ouvrage de référence pour l'ensemble des questions portant sur la plateforme de simulation.

4. Dans le cas où plusieurs catégories d'usagers sont considérées, ou plusieurs périodes temporelles, la demande est en fait représentée par un ensemble de matrices, autant que de segments de demande.

rendre compte des déplacements locaux (qui passent d'« intra-zonaux » donc non modélisés, à « inter-zonaux »).

L'autre jeu de données concerne l'offre. Il est principalement constitué d'une description des réseaux de transports. Cette description comporte trois aspects : physique, topologique, technologique. L'aspect physique vise à inscrire le réseau dans un cadre spatial le plus proche possible de la réalité géographique. Bien que non essentielle au fonctionnement des algorithmes d'affectation, une bonne représentation spatiale du réseau facilite l'étape d'interprétation des résultats et simplifie la tâche du chargé d'études pour les deux autres aspects. L'aspect topologique consiste en la définition des réseaux en tant graphes orientés, comportant donc des arcs orientés reliés entre eux par des nœuds. L'orientation des arcs et l'existence des nœuds autorise des parcours du graphe. Enfin l'aspect technologique consiste en la définition d'un coût de parcours du graphe, par l'intermédiaire d'une modélisation du fonctionnement technique d'un arc (ou d'un nœud). En pratique, ce coût est calculé en fonction d'attributs portés par les arcs et les nœuds : longueur, vitesse de parcours « à vide », capacité, présence d'un péage, etc. Ces attributs sont utilisés, au sein de l'algorithme d'affectation, pour calculer le coût de parcours « en charge », tenant compte de l'existence d'un volume de trafic parcourant les arcs et d'une modélisation de la congestion (cf. chapitre 2).

Dans le cas d'un réseau de transport collectif, une représentation des missions est nécessaire. Il s'agit d'une couche supplémentaire de modélisation, sous la forme de graphes extraits du graphe principal représentant le système de transport dans son ensemble. Des attributs supplémentaires, tels que la fréquence voire les horaires des services sont également ajoutés sur les nœuds.

La plateforme de simulation confronte ces deux jeux de données dans un algorithme d'affectation du trafic au réseau, qui diffère selon le mode considéré, la technologie de congestion retenue, le type d'équilibre recherché, etc. À titre d'exemple, l'algorithme de Frank et Wolfe est fréquemment employé pour résoudre le problème de l'équilibre de Wardrop (cf. chapitre 2).

3.2.2. Le traitement des dimensions

La représentation de l'affectation sur un réseau à travers une plateforme de modélisation, telle que nous venons de le décrire, souligne l'intérêt d'une telle approche pour prendre en compte la dimension spatiale de la congestion : la description détaillée du réseau qu'autorise cette plateforme permet un traitement satisfaisant de cette dimensions. Par ailleurs, la dimension socio-économique est également prise en compte par l'intermédiaire d'une part des étapes précédent l'affectation (génération et distribution des déplacements, choix modal) et d'autre part par la modélisation des comportements de choix d'itinéraire (et éventuellement d'horaire de départ).

Concernant les autres dimensions, physique et économique, nous avons, dans le chapitre précédent, abordé les deux courants principaux de modélisation de la congestion du trafic routier, à savoir la modélisation statique et la modélisation dynamique. Nous avons opté, pour nos analyses, pour une modélisation statique. Cette section vise, dans un premier temps et à travers une discussion critique, à justifier ce choix, puis celui de la forme fonctionnelle particulière retenue. Ensuite, nous justifierons l'approche économique retenue, tout en soulignant les limites dans un cadre opérationnel.

3.2.2.1. Dimension physique : avantages et inconvénients de la modélisation statique

La modélisation statique du trafic peut être critiquée pour son incapacité à traiter des aspects dynamiques de la congestion. En effet, les modèles statiques ne prennent pas en compte le temps de manière *explicite*, et s'il est courant de considérer plusieurs périodes temporelles permettant de rendre compte des variations de la demande au cours du temps, les volumes considérés sont donnés de manière exogène, alors qu'en réalité, ils sont interdépendants. Mais elle présente également des avantages, liés à sa simplicité. En particulier, elle s'adapte plus facilement à de grands réseaux⁵ et à l'étude des interactions entre le transport et les autres aspects du système économique (Verhoef, 1999). C'est l'une des raisons pour lesquelles les modèles statiques sont encore très souvent employés pour des besoins de recherche, et d'enseignement, mais aussi et surtout dans la très grande majorité des études opérationnelles.

L'autre raison réside dans le fait que la modélisation statique est suffisante pour rendre compte des situations de congestion non extrêmes en section courante ou pour les jonctions régulées, qui constitue une large partie des cas d'étude opérationnels. Lorsque les niveaux de congestion deviennent importants, une modélisation dynamique s'avère plus appropriée même si les difficultés inhérentes à sa mise en œuvre la réservent généralement à la recherche.

Par ailleurs, un modèle statique d'affectation statique repose sur des formes fonctionnelles liant débit et vitesse. Nous en comparons ici quelques unes, afin de mettre en évidence leurs différences, avantages et inconvénients. De nombreuses formes fonctionnelles différentes peuvent être employées : ainsi, Leurent (2001) en étudie et en compare huit. Toutefois, certaines formes ne sont pas adaptées à une utilisation dans un modèle opérationnel d'affectation. Il faut pour cela des formes fonctionnelles présentant des propriétés mathématiques adéquates. En particulier,

5. Nous signalons toutefois que les travaux présentés dans Leurent et Aguiléra (2009) sont une tentative de modélisation dynamique à l'échelle de l'Île-de-France.

la branche inférieure de la courbe débit-vitesse présentée sur le premier quadrant de la Figure 2.2 du chapitre précédent n'est pas modélisée dans ces formes fonctionnelles⁶. À l'instar de Ortúzar et Willumsen (2001), nous nous limiterons ici aux formes fonctionnelles couramment employées dans ces modèles, autrement dit celles qui sont directement utilisables dans des modèles d'affectation routière. Plus précisément, nous avons donc choisi de détailler les formes fonctionnelles suivantes : la fonction BPR, la fonction Davidson et la fonction Akçelik.

Nous reprenons les notations de Leurent (2001) qui propose une formulation générique d'une telle forme fonctionnelle :

$$t_{ac} = t_{ac}(x_a) = t_{ac}^0 F_{ac} \left(\frac{x_a}{\kappa_a} \right) = t_{ac}^0 F_{ac}(\xi_a) \quad (3.1)$$

où t_{ac} est le temps de franchissement unitaire, c'est-à-dire par unité de distance, de l'arc a pour la classe de véhicules c , t_{ac}^0 le temps de franchissement de l'arc à *vide* c'est-à-dire, en l'absence de tout véhicule, x_a le débit, κ_a la capacité de l'arc, ξ_a le ratio volume sur capacité, appelé taux de charge, et F_{ac} la fonction de congestion, au centre de notre intérêt. Même si le réalisme d'une telle forme fonctionnelle, dans laquelle le temps de parcours ne dépend que du volume de trafic sur l'arc et des caractéristiques physiques de l'arc, peut être critiquée (Ortúzar et Willumsen, 2001), cette hypothèse simplifie grandement la construction de modèles d'affectation du trafic.

La première forme fonctionnelle que nous étudions, par ailleurs historiquement l'une des premières formulations et également la plus employée, est connue sous le nom de BPR (1964)⁷. Il s'agit de la forme fonctionnelle suivante, qui fournit le temps unitaire de franchissement :

$$t_{ac} = t_{ac}^0 (1 + \alpha \xi_a^\beta) \quad (3.2)$$

où α et β sont deux paramètres d'ajustement. α peut s'interpréter comme le retard à capacité, c'est-à-dire lorsque le flux est égal à la capacité. β , quant à lui, est un paramètre de forme, qui agit sur la courbure de la fonction. Ce paramètre joue donc un rôle important dans la mesure du coût externe marginal de la congestion.

La deuxième forme fonctionnelle étudiée, couramment employée, en particulier

6. Comme l'a montré Verhoef (2001), cette branche inférieure n'est pas stable, cf. 2.4.1.1.

7. En réalité, le nom de BPR est réservé au cas où, dans la forme fonctionnelle, les valeurs des coefficients α et β sont respectivement 0,15 et 4, tandis que pour $\alpha = 0,2$ ou $0,05$ et $\beta = 10$, on parle de BPR actualisée.

en France⁸, est celle issue de Davidson (1966) :

$$t_{ac} = t_{ac}^0 \frac{b - a\xi_a}{b - \xi_a} \quad (3.3)$$

où a et b sont deux paramètres ajustables. On interprète b comme un débit asymptotique au voisinage duquel le temps tend vers l'infini, tandis que le paramètre a dose la vitesse de cette évolution rapide.

Enfin, la troisième forme fonctionnelle que nous avons choisie a été formulée par Akçelik (1991) :

$$t_{ac} = t_{ac}^0 + \frac{P}{4} \left[(\xi_a - 1) + \sqrt{(\xi_a - 1)^2 + \frac{8J\xi_a}{\kappa_a P}} \right] \quad (3.4)$$

où P est la durée de la période de pointe considérée et J un paramètre de retard, correspondant à une modélisation stochastique des files d'attente avec arrivées aléatoires.

Ces trois formes fonctionnelles peuvent être comparées entre elles, mettant en évidence des différences de comportement, en particulier lorsque le taux de charge prend des valeurs élevées. Mais il est tout d'abord à noter que la formulation d'Akçelik dépend de la longueur de l'arc ou, plus précisément, du temps à vide, et de sa capacité. En revanche, elle n'exige pas la détermination, souvent difficile, de paramètres spécifiques de l'arc. Néanmoins, cette dépendance de la formulation à la longueur de l'infrastructure est un sérieux handicap pour son utilisation pratique dans des modèles où la longueur des arcs est bien souvent artificielle. De plus, le paramètre P peut se révéler difficile à évaluer, car la définition d'une période de pointe peut être complexe.

La Figure 3.1 compare les trois relations (3.2), (3.3) et (3.4) précédentes pour les valeurs numériques des paramètres présentées dans la Table 3.1, issues, pour les fonctions BPR et Davidson, d'un calibrage sur l'Île-de-France réalisée par Cohen *et al.* (2001) :

La Figure 3.1 amène plusieurs commentaires. On constate que pour des taux de charge élevés, la pente des courbes Davidson est plus élevée que celle des courbes BPR. Cela est dû à la nature hyperbolique de la relation de Davidson. Cela a une influence importante sur l'évaluation des coûts de la congestion. Enfin, on constate que, d'une façon générale, les courbes se rapportant à des voies rapides présentent une croissance plus lente du temps de franchissement, en particulier pour les valeurs faibles du taux de charge. Les voies rapides ont donc une meilleure « résistance »

8. Du fait de son utilisation au sein du logiciel de modélisation DaVISUM de PTV.

| Type | Paramètres | Voie rapide | Voie de desserte |
|----------|------------------|-------------|------------------|
| BPR | α | 0,87 | 5 |
| | β | 4,96 | 5 |
| Davidson | a | 0,90 | 0,40 |
| | b | 1,1 | 1,1 |
| Akçelik | κ (véh/h) | 4000 | 600 |
| | t^0 (h) | 0,05 | 0,01 |
| | J | 0,1 | 1 |

TABLE 3.1.: Valeurs numériques employées dans la comparaison des formes fonctionnelles de congestion. *Source* : Estimations de l'auteur et Cohen *et al.* (2001)

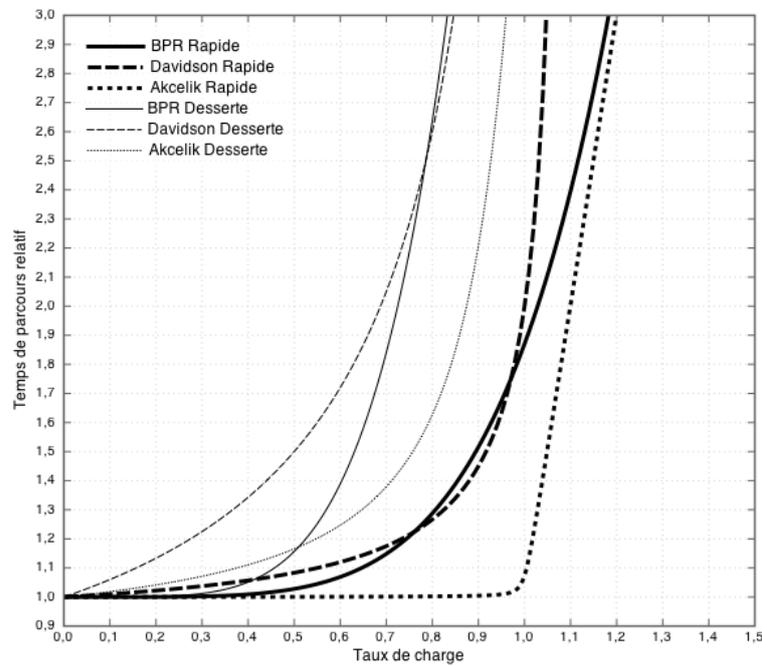


FIGURE 3.1.: Comparaison de trois types de formes fonctionnelles de technologie de congestion. *Source* : Calculs de l'auteur

à la congestion.

Pour nos simulations spatialisées de la congestion en Île-de-France, nous avons retenu la forme fonctionnelle BPR, pour sa simplicité d'usage et de calibration, ainsi que par son utilisation très répandue dans les modèles tant opérationnels que de recherche.

3.2.2.2. Dimension économique : quelle définition pour l'offreur ?

Le traitement de la dimension économique à travers la plateforme de simulation détaillée au 3.2.1 repose sur l'analogie explicitée au 2.4.1.1 du chapitre précédent : la courbe débit - temps de parcours constitue une courbe d'offre, à laquelle on peut confronter une courbe de demande provenant directement des usagers. Comme nous l'avons déjà mentionné, cette analogie laisse de côté un élément pourtant important : si la courbe débit - temps de parcours a la forme d'une courbe d'offre, elle ne permet pas d'identifier l'offreur. Plus précisément, alors qu'une courbe d'offre est sensée représenter le coût de production, pour un offreur, d'une certaine quantité d'un bien ou service, aucun offreur véritable ne subit le coût que représente le temps de parcours : seuls les usagers le subissent ! Ils sont donc à la fois individuellement demandeurs et collectivement offreurs.

Concernant la représentation de la demande, plusieurs options sont envisageables. Une segmentation de la demande permet de rendre compte de l'hétérogénéité des usagers, en termes de comportement, de valeur du temps, etc. Toutefois, considérer les usagers comme homogène, permet, outre la simplicité qu'offre cette option, d'assurer la continuité des fonctions de coût (Verhoef, 1999). De plus, la question que nous traitons ici ne portant pas spécifiquement sur l'hétérogénéité des comportements, l'hypothèse d'homogénéité est parfaitement valable.

3.3. Trois perspectives d'évaluation de la congestion

L'évaluation de la congestion, qu'il s'agisse de diagnostiquer le niveau de congestion auquel sont soumis une agglomération, un quartier ou un axe routier, dépend du point de vue choisi par l'évaluateur. En effet, la perspective du gestionnaire d'infrastructure routière et celui de l'utilisateur diffèrent sur bien des points. Il en va de même entre celui de l'élu et celui de l'économiste.

C'est la raison pour laquelle le rapport ECMT (2007) préconise l'utilisation simultanée de plusieurs indicateurs de congestion, de sorte à capturer de manière satisfaisante les différences de perception de la congestion par les acteurs. Le rapport propose en conséquence un inventaire des indicateurs couramment employés pour évaluer les manifestations de la congestion routière. Ils se classent en sept catégories : indicateurs basés sur les vitesses, sur les retards, indicateurs d'emprise spatiale, indicateurs d'efficacité technique, de fiabilité, d'efficacité économique, et enfin, les autres indicateurs, généralement liés à l'impact environnemental de la congestion.

Ces nombreux indicateurs n'ont cependant pas le même intérêt pour les différents acteurs concernés par la congestion. Ainsi, un usager sera principalement intéressé par les indicateurs de vitesses, de retard et de fiabilité, tandis qu'un gestionnaire

d'infrastructure privilégiera les indicateurs d'efficacité technique comme le taux de charge. De plus, chacun de ces acteurs s'appuie, explicitement ou implicitement, un critère d'optimalité différent.

C'est pourquoi nous nous proposons, dans cette section, de définir trois perspectives d'évaluation, chacune d'elles étant associée à un indicateur. Ces perspectives sont celles de l'utilisateur, du gestionnaire d'infrastructure et de l'économiste⁹. Nous avons élaboré ces perspectives et choisi ces indicateurs car ils répondent, selon nous, à notre objectif d'analyse spatialisée de la congestion sous différents angles, tout en maintenant un niveau de lisibilité suffisant. Autrement dit, nous avons estimé qu'elles permettraient de réaliser un compromis entre un dessin du contour de cette notion complexe qu'est la congestion et l'exigence de relative concision que suppose l'écriture d'un chapitre de thèse. Ce choix trouve également un écho dans les différentes mesures de la congestion envisagées par Boiteux *et al.* (1994) et Boiteux et Baumstark (2001). Nous restons néanmoins conscient du caractère caricatural et réducteur de cette démarche qui se limite à trois perspectives, chacune étant associée à un ou deux indicateurs.

3.3.1. Perspective de l'utilisateur

Pour l'utilisateur, les principales variables de choix — d'itinéraire, d'horaires, de mode — sont liées à la distance et au temps de parcours sur les réseaux de transport, qui engendrent des coûts directs ou indirects pour les utilisateurs. Comme nous l'avons mis en évidence dans le chapitre précédent, le temps mis pour se rendre d'un lieu d'origine à un lieu de destination est d'une importance primordiale. Les différents coûts monétaires, liés à l'utilisation du véhicule et aux différents péages ou tarifs de titres de transport, peuvent être associés aux différents coûts temporels qui varient selon qu'il s'agit d'un temps d'attente, d'un temps de marche ou d'un temps en véhicule, par l'intermédiaire de valeurs monétaires du temps, pour former un coût généralisé de transport (Ortúzar et Willumsen, 2001). Dans un contexte routier urbain en l'absence de péage (comme c'est le cas en Île-de-France, très majoritairement), le temps de parcours apparaît comme la variable de choix d'itinéraire essentielle. La variabilité temporelle de ce temps de trajet, en ce qu'elle implique pour la fiabilité de ce temps, représente également, pour l'utilisateur, un critère de choix important (ECMT, 2007 ; Small et Verhoef, 2007).

En ce qui concerne la référence, autrement dit le critère d'optimalité correspondant à la perspective de l'utilisateur, il semble raisonnable, en concordance avec l'une des définitions proposées par le rapport ECMT (2007), de choisir le temps à vide.

9. Prud'homme (1999a) n'en cite que deux, mais il nous semble, et nous tenterons de le justifier, que ces trois perspectives existent et jouent un rôle bien distinct dans l'analyse de la congestion.

Ainsi, l'écart entre le temps réel et le temps à vide devient une mesure de l'écart à l'optimum, *pour l'utilisateur*¹⁰. L'indicateur retenu pour la perspective de l'utilisateur est donc l'écart relatif au temps de parcours à vide :

$$I_{ac}^u = \frac{t_{ac} - t_{ac}^0}{t_{ac}^0}. \quad (3.5)$$

Notre propos ne porte pas directement sur les questions de fiabilité du temps de trajet, qui constituent à elles seules un champ complet de recherche encore relativement peu exploré (*cf.* Small *et al.*, 1999). Ce choix se justifie en partie par le fait qu'en Île-de-France, la congestion récurrente représente près de 86 % du temps perdu lié à la congestion (ECMT, 2007). Toutefois, comme nous avons pu le noter dans le chapitre précédent, retard et fiabilité sont liés : un axe fortement congestionné, occasionnant un retard important pour les usagers, présente également un risque plus élevé d'être sujet à une forme sévère de congestion incidente. C'est pourquoi nous avons choisi de considérer le temps de trajet comme base pour l'indicateur caractéristique de la perspective de l'utilisateur. Toutefois, nous compléterons l'analyse par un indicateur reflétant la sensibilité du temps de parcours à un accroissement du volume (ce qui, en première approximation, correspond également à une baisse de la capacité). L'élasticité temps-volume, indicateur de risque pour l'utilisateur, est fournie par l'expression suivante :

$$I_{ac}^r = \frac{x_a}{t_{ac}} \frac{dt_{ac}}{dx_a}. \quad (3.6)$$

Plus cet indicateur est élevé, plus le risque est grand qu'une variation ponctuelle à la hausse du nombre d'utilisateurs ou une baisse de la capacité (accident, etc.) entraînent une forte augmentation du temps de parcours sur l'arc.

3.3.2. Perspective du gestionnaire d'infrastructure

Le gestionnaire d'infrastructure est, quant à lui, surtout intéressé par le fait d'utiliser au mieux son réseau, autrement dit d'en optimiser l'usage, d'un point de vue technique. Il n'a, *a priori*, pas de raison particulière de chercher à savoir quel est le temps de parcours d'un utilisateur donné, voire d'une catégorie d'utilisateurs. En revanche, il doit veiller à ce que le système de transport dont il a la charge, qu'il s'agisse d'un réseau routier complet, d'un sous-ensemble de ce réseau, voire d'une seule infrastructure, soit utilisé de la manière la plus rentable, techniquement. Or, comme nous l'avons évoqué précédemment, c'est lorsque l'infrastructure fonctionne

10. Ce critère n'est pas du tout satisfaisant dans une perspective d'optimalité économique, comme nous l'avons déjà mentionné au chapitre précédent.

à *capacité* que le débit de véhicules est maximal. Plus précisément, pendant une période δh , N_h usagers circulant à la vitesse v_h correspondent à une production $N_h v_h \delta h$, tandis que le facteur de production pendant la même période est $L \kappa \delta h$, où L est la longueur de la route et κ la capacité. Le rapport de ces deux grandeurs fournit la productivité de l'infrastructure : $N_h v_h \delta h / L \kappa \delta h = x_h / \kappa$. Le taux de charge fournit donc la *productivité* d'une infrastructure et elle est maximale à la capacité (Leurent, 2009). Autrement dit, si l'exploitant routier a comme objectif de maximiser le nombre d'usagers circulant chaque heure sur son infrastructure, il cherchera vraisemblablement à se rapprocher de la capacité de l'infrastructure.

Il veillera néanmoins à ne pas s'en approcher trop, car cette zone à proximité de la saturation est caractérisée par une grande instabilité : lorsque la capacité est atteinte, la moindre perturbation du flux de véhicules peut faire passer dans la zone hypercongestionnée, avec création d'une file d'attente, entraînant un phénomène d'hystérésis, avec un difficile et lent retour à une situation non saturée (Leurent, 2005). Par conséquent, le taux de charge est également une mesure du risque de saturation liée à la congestion incidente. De plus, compte tenu de la croissance constatée du trafic routier sur le long terme, une infrastructure fonctionnant à proximité de la capacité, c'est-à-dire avec un taux de charge légèrement inférieur à 1, rencontrera vraisemblablement des problèmes de saturation récurrents dans un avenir proche.

Il ressort de cela qu'un indicateur pertinent pour le gestionnaire d'infrastructure est avant tout le taux de charge sur les arcs du réseau, qui mesure l'efficacité technique du réseau. La situation de référence correspondante est le fonctionnement à capacité, autrement dit le taux de charge unité. Nous retiendrons donc le taux de charge (et son écart à l'unité) comme indicateur de la congestion routière pour le gestionnaire d'infrastructure :

$$I_a^g = \frac{x_a}{\kappa_a}. \quad (3.7)$$

3.3.3. Perspective de l'économiste

La perspective de l'économiste se distingue assez nettement des deux précédentes, en particulier parce que la définition de l'optimum, ou situation de référence, est dans ce cas plus complexe. L'objectif de l'économiste lorsqu'il s'intéresse au fonctionnement d'un réseau routier, est de le rendre efficient¹¹. Ainsi, comme le rappelle le rapport ECMT (2007), ce point de vue permet de définir une congestion excessive de la manière suivante :

La congestion peut être définie comme excessive lorsque les coûts mar-

11. L'efficacité est le rapport entre le résultat obtenu et le résultat attendu, tandis que l'efficience est le rapport entre le résultat obtenu et les moyens mis en œuvre pour l'obtenir. Le terme est à rapprocher du rendement.

ginaux sociaux de la congestion dépassent les gains marginaux sociaux des efforts visant à la réduire (tels que l'extension de la capacité routière ou d'autres modes de transport)¹².

On peut également reprendre la définition proposée par Boiteux et Baumstark (2001) :

Le coût économique de la congestion peut être défini en théorie comme la différence entre l'utilité effectivement retirée de l'usage de l'infrastructure et l'utilité qui en serait retirée si elle était utilisée de façon optimale.

La situation de référence de la perspective de l'économiste est donc l'optimum social. La principale différence avec les perspectives précédentes réside dans le fait que cette situation de référence n'est pas uniquement liée aux caractéristiques physiques de l'infrastructure considérée. Si la capacité de l'arc ou le temps de parcours à vide ne dépendent que des caractéristiques physiques de l'arc, l'optimum économique dépend également de la demande et de sa structure. Ainsi, la Figure 3.2, qui reprend celle du chapitre précédent, illustre la dépendance de la situation de référence à la courbe de demande. Plus précisément, on constate ici que la situation optimale dans le cas d'une demande plus forte et moins élastique correspond à un volume de véhicules plus élevé sur l'arc. Autrement dit, si la demande est forte et peu sensible au coût, un niveau plus élevé de congestion est acceptable.

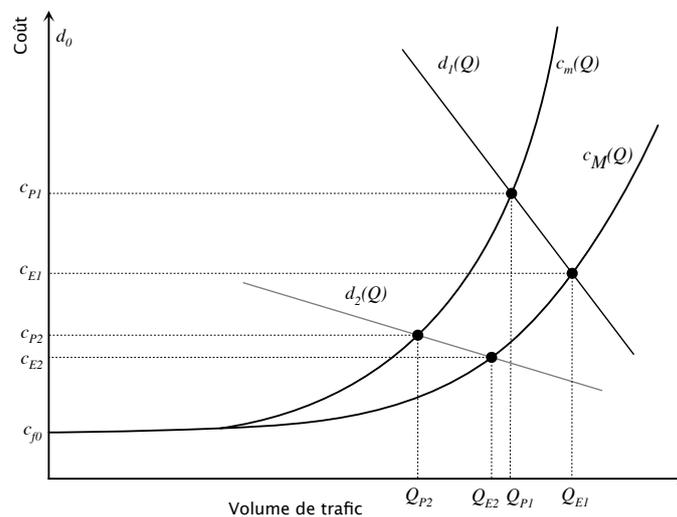


FIGURE 3.2.: Equilibre et optimum pour deux motifs distincts. *Source* : Auteur

12. Traduction par l'auteur de la définition de la Victorian Competition and Efficiency Commission australienne.

De manière à refléter la perte de surplus due à la congestion, l'indicateur que nous retiendrons est le coût externe marginal, donné par :

$$I_a^e = x_a \frac{dt_a}{dx_a}. \quad (3.8)$$

Les paragraphes suivants visent à mettre en évidence le rôle joué par ce coût externe dans l'évaluation de la perte de surplus des usagers.

3.4. Évaluation locale de la congestion

Avant d'aborder les questions liées aux manifestations spatiales de la congestion et aux manières d'appréhender sa variabilité dans l'espace, il nous semble essentiel d'explicitier, au niveau d'un arc du réseau, l'évaluation économique de la congestion. Celle-ci passe par l'évaluation d'une perte de surplus, elle-même directement liée au coût externe marginal de congestion.

3.4.1. Perte de surplus due à la congestion

Dans le cas d'une fonction de temps de parcours BPR, et pour une demande parfaitement élastique, la perte sèche de congestion sur un axe routier se calcule de la manière suivante¹³, avec les notations employées auparavant :

$$H = \int_{Q_S}^{Q_E} (C_m(q) - D^{-1}(q)) dq. \quad (3.9)$$

En supposant que la fonction de demande inverse est telle que $\forall q, D^{-1}(q) = C_E = t_0[1 + \alpha(\frac{Q_E}{\kappa})^\beta]$ et par ailleurs que $C_m(q) = t_0[1 + \alpha(\frac{q}{\kappa})^\beta(1 + \beta)]$, on obtient :

$$H = \frac{t_0 \alpha Q_S}{\kappa^\beta} (Q_E^\beta - Q_S^\beta). \quad (3.10)$$

Q_S peut être déterminé comme l'intersection de la courbe de demande et de la courbe de coût marginal :

$$Q_S = \kappa \left[\frac{C_E - t_0}{\alpha t_0 (1 + \beta)} \right]^\beta. \quad (3.11)$$

13. Compte tenu de l'hypothèse relativement forte de parfaite élasticité, cette évaluation constitue une borne supérieure de la perte sèche qui aurait lieu avec une élasticité finie.

Si bien que la borne supérieure de la perte de surplus associée à la congestion sur un axe peut être évaluée par :

$$H(Q_E) = \frac{\alpha\beta t_0 Q_E^{1+\beta}}{\kappa^\beta (1+\beta)^{1+1/\beta}}. \quad (3.12)$$

À titre d'application numérique et d'illustration, prenons le cas d'un arc collecteur de capacité $\kappa = 1000$ véh/h, avec $\alpha = \beta = 5$, une vitesse à vide de 60 km/h, et un flux d'équilibre de 700 véh/h. La perte de surplus liée à la congestion est majorée par 5,7 heures pour chaque heure d'écoulement à ce régime.

Cette analyse montre que l'évaluation de la grandeur pertinente pour l'économiste, le coût économique de la congestion, suppose la caractérisation de la situation d'équilibre, et donc la connaissance précise de la demande. Par ailleurs, nous avons vu que, par l'intermédiaire du triangle de Harberger, ce coût économique de la congestion est intimement lié au coût marginal de congestion, et plus précisément à la différence entre coût marginal et coût moyen, autrement dit le coût externe. Bien entendu, cet indicateur ne suffit pas à déterminer l'écart à l'optimum ou la perte sèche. Néanmoins, en l'absence d'une information précise sur la situation optimale¹⁴, l'étude du coût marginal fournit une première approche de la question. Cela justifie en partie, *a posteriori*, notre choix de cet indicateur dans la perspective de l'économiste.

3.4.2. Évaluation du coût externe marginal de congestion routière

Dans le cas qui nous préoccupe ici d'un arc routier, on peut définir le coût marginal de congestion de la manière suivante¹⁵ : sur un arc, noté a , de longueur L_a , qui écoule un débit de x_a véhicules pendant une durée H , le passage d'un véhicule supplémentaire induit une distance parcourue supplémentaire L_a , et une perte de temps pour chacun des autres usagers. En reprenant l'équation (3.1), on note $T_a = T_a(x_a)$ le temps de franchissement de l'arc a par usager lorsque le volume est x_a . Le passage du véhicule supplémentaire cause à chaque usager une perte de temps de $\delta T_a = T_a(x_a + 1) - T_a(x_a)$.

La perte de temps pour l'ensemble du flux local est la somme des gênes subies

14. Le rapport de Didier et Prud'homme (2007) évoque des méthodes visant à déterminer cette valeur optimale mais l'évaluation de la perte sèche nécessite également de faire des hypothèses sur la courbe de demande.

15. On reprend ici la présentation de Leurent et Breteau (2009).

individuellement, soit, pour les $x_a + 1$ usagers :

$$\Delta T_a = (x_a + 1)T_a(x_a + 1) - x_a T_a(x_a) = x_a \delta T_a + T_a(x_a + 1) = x_a \frac{dT_a}{dx_a} + T_a(x_a + 1). \quad (3.13)$$

Ramené au supplément de distance parcourue L_a , le coût marginal d'une unité de trafic supplémentaire est :

$$c_{ma} = \frac{\Delta T_a}{L_a} = x_a \frac{dt_a}{dx_a} + t_a(x_a + 1), \quad (3.14)$$

avec $t_a(x_a) = T_a(x_a)/L_a$.

Comme nous l'avons évoqué précédemment, ce coût marginal se décompose en un coût marginal subit par l'utilisateur supplémentaire, c'est-à-dire le coût moyen $t_a(x_a + 1)$, et un coût imposé aux autres usagers, le coût externe. Ce dernier se résume donc à :

$$\chi_a = x_a \frac{dt_a}{dx_a}. \quad (3.15)$$

Il s'agit là d'un indicateur physique, en unité de temps, pour le coût externe marginal. On peut en déduire aisément un indicateur économique en unité monétaire, en le multipliant par une valeur unitaire du temps passé en transport. Nous conserverons, pour notre part, l'indicateur temporel, pour sa simplicité d'usage et son indépendance aux évaluations des valeurs du temps.

Comme le suggèrent les courbes présentées plus haut, l'évaluation des coûts externes marginaux de congestion, de même que l'évaluation des temps de parcours, dépendent généralement du choix de la forme fonctionnelle de la relation temps-débit. Plus précisément, les trois types de relations évoquées plus haut, BPR, Davidson et Akçelik, fournissent les coûts externes marginaux suivants :

$$\chi_a^{BPR} = t_a^0 \alpha \beta \xi_a^\beta, \quad (3.16)$$

$$\chi_a^{Dav} = t_a^0 \xi_a \frac{b(1-a)}{(b-\xi_a)^2}, \quad (3.17)$$

$$\chi_a^{Akc} = \frac{P}{4} \left[\xi_a + \frac{\xi_a(\xi_a - 1) + \frac{4J\xi_a}{\kappa P}}{\sqrt{(\xi_a - 1)^2 + \frac{8J\xi_a}{\kappa P}}} \right]. \quad (3.18)$$

On constate que les coûts externes marginaux sont croissants en ξ_a , ce qui était visible sur la Figure 3.1, où les courbes sont clairement convexes.

Par ailleurs, pour les valeurs de ξ_a possibles en régime stationnaire, c'est-à-dire $\xi_a \leq 1$, le coût supplémentaire imposé à chaque usager présent sur l'arc par un

usager supplémentaire est majoré, dans chacun des cas, par les valeurs suivantes :

$$\overline{\delta\chi_a^{BPR}} = \frac{t_a^0 \alpha \beta}{\kappa}, \quad (3.19)$$

$$\overline{\delta\chi_a^{Dav}} = t_a^0 \frac{b(1-a)}{\kappa(b-1)^2}, \quad (3.20)$$

$$\overline{\delta\chi_a^{Akc}} = \frac{P}{4\kappa} \left(1 + \sqrt{\frac{2J}{\kappa P}} \right). \quad (3.21)$$

On constate, avec ces formules, que ces majorants dépendent de la capacité κ de l'arc étudié. Cette propriété met en évidence une caractéristique du coût externe marginal de la congestion routière de flux : le coût imposé par un usager supplémentaire (ou un véhicule-kilomètre) diminue lorsque la capacité de l'infrastructure augmente. On peut parler d'un *effet capacité* pour désigner cette dépendance inverse du coût externe marginal par usager supplémentaire à la capacité.

3.5. Appréhender la variabilité spatiale de la congestion

Si au niveau local d'un élément de réseau pris isolément, l'analyse peut être menée entièrement à l'aide des outils de la modélisation économique, l'analyse n'est pas directement transposable à l'étude d'un réseau de transport complexe, comprenant de nombreuses origines et destinations, et d'encore plus nombreux itinéraires les reliant. Nous proposons, dans cette section, deux approches complémentaires visant à mieux appréhender la congestion dans un contexte spatial. La première s'appuie sur une typologie et des traitements statistiques, la seconde développe une approche de nature plutôt géographique visant à rendre compte des logiques spatiales sous-jacentes à la congestion.

3.5.1. Approche statistique

Il s'agit, dans cette approche, de décrire, sur la base d'analyses statistiques exploratoires, différentes classifications visant à appréhender, à différents niveaux d'agrégation, la variabilité spatiale (et temporelle) de la congestion. Plus précisément, nous avons opéré une segmentation du réseau, qui a nécessité un choix de critères de regroupement, permettant de définir un certain nombre de sous-réseaux. Cette méthode est d'une certaine manière celle qu'ont employée Prud'homme et Sun (2000) et Koning (2009), en ne s'intéressant qu'à un segment très particulier du réseau francilien, le Boulevard Périphérique parisien. Par ailleurs, des travaux tels

que ceux de Geroliminis et Daganzo (2007, 2008) et Buisson et Ladier (2009) ont montré qu'il était possible d'établir des diagrammes fondamentaux agrégés, à partir de segments du réseau. Nous visons, pour notre part, un traitement plus systématique de l'ensemble du réseau, avec la contre-partie d'un traitement moins complet, en particulier par le fait que, comme nous l'avons déjà signalé, nous n'estimerons pas de courbes de demande.

Afin de définir les segments, nous avons croisé deux critères de classification qui nous sont apparus à la fois suffisants pour décrire relativement finement les sous-réseaux et suffisamment généraux pour s'appliquer à la majorité des réseaux de transport auxquels un ingénieur-planificateur peut être confronté : le type d'infrastructure routière et l'environnement urbain dans lequel se situe l'infrastructure. Ces deux critères nous sont en effet apparus comme les plus centraux et aisément identifiables pour expliquer la variabilité spatiale de la congestion. Le croisement de ces deux critères permet de remplir notre double objectif : disposer d'une méthode efficace pour analyser la variabilité de la congestion, et faire en sorte que cette méthode soit aisément applicable par un chargé d'étude (Leurent *et al.*, 2009).

3.5.1.1. Deux axes de distinction : la hiérarchie du réseau et l'environnement urbain

Des deux axes de distinction, le type routier s'impose naturellement, puisque ce critère, par l'intermédiaire de la capacité et des coefficients de la fonction de temps de parcours BPR, détermine en grande partie la manière dont une infrastructure réagit à la congestion. Cet axe de distinction correspond à une segmentation de l'*offre* fournie par le réseau.

La hiérarchie d'un réseau distingue classiquement quatre niveaux, comme l'illustre la Figure 3.3. Les deux niveaux supérieurs correspondent aux réseaux structurant, tandis que les deux niveaux inférieurs sont des niveaux de desserte. La capacité des voies augmente à mesure qu'on s'élève dans la hiérarchie. La hiérarchisation de la voirie en termes de capacité des voies est d'ailleurs un élément important de planification des transports, car elle assure, lorsqu'elle convenablement choisie, une adéquation entre demande et offre. En d'autres mots, si les voies structurantes du côté de la demande ne sont pas dotées d'une capacité suffisante, les niveaux de congestion risquent d'être très élevés. En retour, le développement urbain se faisant en relation aux réseaux existants, une hiérarchisation correctement définie est de nature à limiter un développement urbain anarchique (étalement urbain).

Face à un réseau-modèle réel de transport, la détermination du groupe d'appartenance d'une voie donnée n'est pas toujours faisable avec certitude. Elle nécessite *a priori* une connaissance approfondie du contexte local, en particulier pour déterminer si une voie appartient à la catégorie des artérielles ou des collectrices, celle



FIGURE 3.3.: Hiérarchie d'un réseau. *Source* : ECMT (2007)

des voies rapides étant beaucoup plus clairement définie.

Le second axe de distinction que nous avons retenu est lié au type d'environnement urbain des arcs du réseau. Un critère de ce type vise à obtenir une forme de segmentation de la *demande*. La typologie peut dépendre du contexte particulier du réseau considéré ou des besoins de l'étude. Toutefois, pour qu'une telle typologie ait un intérêt dans l'appréhension de la congestion, il faut que le critère rende compte d'un déterminant de la demande. Ainsi, un environnement urbain dense présente une structure de demande différente d'un environnement périurbain peu dense. Autrement dit, la densité constitue, à notre sens, un bon critère de distinction.

3.5.1.2. Analyses statistiques exploratoires

À partir du croisement des typologies que nous venons de définir, l'analyse statistique que nous nous proposons de mettre en œuvre consiste à comparer les valeurs des principaux moments statistiques et quantiles remarquables des indicateurs correspondant aux trois perspectives définies au 3.3 afin d'apporter des éléments de compréhension de l'organisation spatiale et structurelle de la congestion. L'analyse peut également montrer les faiblesses des typologies choisies, et ainsi guider l'évaluateur vers une typologie plus adaptée au territoire étudié.

Les indicateurs ainsi que les populations statistiques varient selon la perspective adoptée. Par exemple, l'individu statistique intéressant avant tout le gestionnaire d'infrastructure est l'unité de longueur de linéaire routier, typiquement le kilomètre d'infrastructure. L'économiste s'intéressera plutôt, pour sa part, à la population qui crée ou qui subit les coûts de la congestion ; autrement dit, il considérera les unités

de trafic routier, typiquement les véhicules-kilomètres¹⁶. Nous précisons donc ici pour chacune des grandeurs étudiées, la population statistique considérée, et par conséquent les formules fournissant les statistiques souhaitées.

Dans la perspective de l'utilisateur, La population statistique considérée est celle des éléments unitaires de trafic, autrement dit des véhicules-kilomètres¹⁷ et la variable aléatoire considérée est l'écart relatif au temps de parcours unitaire $\Delta t_{ra} = \frac{t_a - t_a^0}{t_a^0}$. La moyenne, sur la population des véhicules-kilomètres, de l'écart de temps de parcours relatif est fournie par :

$$\mathbb{E}(\Delta t_r) = \frac{1}{\sum_{a \in A} L_a x_a} \sum_{a \in A} L_a x_a \left(\frac{t_a - t_a^0}{t_a^0} \right), \quad (3.22)$$

où L_a est la longueur de l'arc a , x_a le nombre de véhicules sur l'arc, t_a et t_a^0 respectivement le temps de parcours et le temps de parcours à vide par unité de distance sur l'arc. De même, on obtient la variance du temps de parcours relatif :

$$\mathbb{V}(\Delta t_r) = \left[\frac{1}{\sum_{a \in A} L_a x_a} \sum_{a \in A} L_a x_a \left(\frac{t_a - t_a^0}{t_a^0} \right)^2 \right] - \mathbb{E}(\Delta t_r)^2. \quad (3.23)$$

Comme nous l'avons mentionné, les questions de fiabilité peuvent être en partie appréhendées par un indicateur d'élasticité temps-volume. Celle-ci s'écrit $\chi_a/t_a = \frac{x_a}{t_a} \frac{dt_a}{dx_a}$; sa moyenne et sa variance sont fournies par :

$$\mathbb{E}(\chi/t) = \frac{1}{\sum_{a \in A} L_a x_a} \sum_{a \in A} \left(L_a x_a \frac{x_a}{t_a} \frac{dt_a}{dx_a} \right) = \frac{1}{\sum_{a \in A} L_a x_a} \sum_{a \in A} L_a \frac{x_a^2}{t_a} \frac{dt_a}{dx_a}, \quad (3.24)$$

$$\mathbb{V}(\chi/t) = \frac{1}{\sum_{a \in A} L_a x_a} \sum_{a \in A} L_a \frac{x_a^3}{t_a^2} \left(\frac{dt_a}{dx_a} \right)^2 - \mathbb{E}(\chi/t)^2. \quad (3.25)$$

Dans la perspective du gestionnaire d'infrastructure, la population statistique à considérer est celle des unités de longueur d'infrastructure. En effet, le taux de charge est une grandeur se rapportant au linéaire d'infrastructure. Ce qui intéresse

16. La distinction entre véhicules et usagers par l'intermédiaire du taux d'occupation des véhicules pourrait être intéressante puisqu'elle permet de mettre en évidence l'influence du covoiturage sur le coût externe imposé par chaque usager. Toutefois, en première approximation, nous assimilons véhicule et usager, autrement dit, nous considérons implicitement un taux d'occupation de 1. Les matrices de la DRIEA-IF, que nous avons utilisées au chapitre 4, sont fournies en véhicules et non en usagers, si bien que notre hypothèse ne surcharge par le réseau, mais sous-estime le coût social total.

17. Il convient de noter qu'une autre population statistique d'intérêt serait celle des déplacements. Malheureusement, les informations nécessaires, et en particulier l'ensemble des chemins suivis par les usagers, ne sont pas disponibles, pour une utilisation systématique et automatisée, dans le logiciel TransCAD utilisé au LVMT. Le lecteur intéressé par la méthode statistique à employer avec cette population peut se reporter à l'annexe 3.A.1.

le gestionnaire, c'est par exemple de savoir la proportion du réseau soumis à des taux de charge élevés. L'espace de travail est donc bien celui du réseau et non pas celui des usagers ou des déplacements. La variable aléatoire considérée est quant à elle tout simplement le taux de charge $\xi_a = x_a/\kappa_a$. Par conséquent, la moyenne et la variance des taux de charge sur le réseau sont fournies par :

$$\mathbb{E}(\xi) = \frac{1}{\sum_{a \in A} L_a} \sum_{a \in A} L_a \xi_a, \quad (3.26)$$

$$\mathbb{V}(\xi) = \left[\frac{1}{\sum_{a \in A} L_a} \sum_{a \in A} L_a \xi_a^2 \right] - \mathbb{E}(\xi)^2, \quad (3.27)$$

où ξ_a est le taux de charge de l'arc a , et L_a sa longueur.

Dans la perspective de l'économiste, enfin, la population statistique considérée est, comme dans la perspective de l'utilisateur, celle des unités de trafic. En effet, l'unité d'intérêt est bien l'élément unitaire de trafic, car c'est lui qui induit le coût marginal subi par les autres usagers. Or, cet élément unitaire de trafic est bien le véhicule-kilomètre, c'est-à-dire une distance unitaire parcourue sur le réseau par un véhicule équivalent à une voiture particulière. La variable aléatoire à analyser est le coût externe marginal associé à un élément de trafic, qui s'écrit, comme nous l'avons vu précédemment, $\chi_a = x_a \frac{dt_a}{dx_a}$. Les formules fournissant la moyenne et la variance du coût externe sur le réseau sont les suivantes :

$$\mathbb{E}(\chi) = \frac{1}{\sum_{a \in A} L_a x_a} \sum_{a \in A} \left(L_a x_a \cdot x_a \frac{dt_a}{dx_a} \right) = \frac{1}{\sum_{a \in A} L_a x_a} \sum_{a \in A} L_a x_a^2 \frac{dt_a}{dx_a}, \quad (3.28)$$

$$\mathbb{V}(\chi) = \frac{1}{\sum_{a \in A} L_a x_a} \sum_{a \in A} L_a x_a^3 \left(\frac{dt_a}{dx_a} \right)^2 - \mathbb{E}(\chi)^2. \quad (3.29)$$

3.5.2. Approche géographique

Au delà de l'approche statistique, qui, tout en offrant une méthode d'analyse de façon à mieux appréhender les niveaux de congestion et les répercussions qu'ils peuvent avoir du point de vue des usagers, du gestionnaire d'infrastructure ou de l'économiste, il semble donc essentiel de développer des méthodes d'analyse adaptée à la variabilité, en particulier spatiale, de la congestion. La section 4.4 suggère que le facteur spatial revêt une importance considérable dans le phénomène de congestion. Or, la simple segmentation du réseau par couronne ne semble pas suffisante pour faire émerger une structure spatiale. Cela s'explique par le fait que si la localisation géographique d'un arc peut influencer son usage, c'est vraisemblablement sa place dans les cheminements qui importe le plus. Plus précisément, la segmentation employée jusque là néglige l'environnement urbain des zones d'origine ou de desti-

nation des déplacements. Or, cette variable influe fortement sur les comportements de déplacements, en particulier au niveau du choix du mode de transport (Ortúzar et Willumsen, 2001).

Nous proposons donc ici une méthode permettant une analyse spatiale de la congestion (Leurent et Breteau, 2009). Cette méthode consiste en une agrégation par zones géographiques sur la base des relations origine-destination empruntées par les usagers. L'objectif d'une telle méthode est de tenir compte de la structure réelle des déplacements dans le diagnostic de la congestion, alors que les analyses précédentes se sont limitées à une étude des flux, sans tenir compte de la place des arcs dans les itinéraires suivis par les usagers.

3.5.2.1. Principe et limites de la méthode

Le principe de cette agrégation réside dans la sommation pondérée, le long des chemins suivis par les usagers, des indicateurs voulus. La méthode consiste donc à rattacher à chaque zone les déplacements qui y prennent leur origine ou y trouvent leur destination pendant la période d'analyse. Ainsi, en notant q_{od} le volume de trafic entre une origine o et une destination d , I_{od} l'indicateur par déplacement sur la relation OD considérée (obtenu lui-même par agrégation le long de l'itinéraire), et D_{od} l'indicateur de distance moyenne par déplacement sur la relation OD , l'indicateur agrégé par unité de distance est fourni par :

$$I_o = \frac{\sum_d q_{od} I_{od}}{\sum_d q_{od} D_{od}} \quad (3.30)$$

où $q_{od} I_{od} = \sum_r x_r I_r$, x_r est le volume de trafic empruntant le chemin r de O à D , et I_r la valeur de l'indicateur le long d'un chemin. Par exemple, dans le cas de l'indicateur de temps de parcours :

$$T_r = \sum_{a \in r} L_a t_a$$

Ce mode d'agrégation fournit la valeur de l'indicateur vu par les usagers partant ou souhaitant se rendre dans une zone. Ainsi, en ce qui concerne le temps de parcours moyen par zone d'origine, il fournit, en tenant compte des trajets effectivement suivis par les usagers provenant d'une zone donnée, le temps de parcours unitaire moyen de ces usagers.

Si le principe est simple, la mise en œuvre se heurte au problème de la non unicité des volumes sur les chemins suivis, à l'équilibre de Wardrop. L'unicité des volumes sur les arcs est en effet assurée à l'équilibre, mais elle ne l'est pas pour les volumes sur les itinéraires. L'Annexe 3.A.2 propose une formalisation mathématique de ce

problème, mettant en évidence la non-unicité, mais l'exemple schématisé sur la Figure 3.4 l'illustre également. Les usagers peuvent, pour se rendre de l'origine O à la destination D , prendre quatre itinéraires différents : arc 1 puis 3 (C_1), arc 1 puis 4 (C_2), arc 2 puis 3 (C_3) ou arc 2 puis 4 (C_4). Si les arcs ont des caractéristiques identiques (longueur, capacité, fonction de temps de parcours), le flux total V se répartira à hauteur de $V/2$ sur les 4 arcs. En revanche, on ne peut rien dire, *a priori* de la répartition des usagers sur les différents chemins. Il est possible que la moitié d'entre eux emprunte C_1 , tandis que l'autre moitié emprunte C_4 , ou qu'ils se répartissent uniformément sur les 4 chemins, ou qu'ils se répartissent selon une autre combinaison linéaire de ces deux situations.

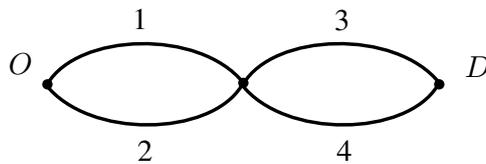


FIGURE 3.4.: Schéma d'un réseau simple

Cette difficulté peut être contournée de différentes manières. En différenciant, de manière continue, les usagers par une caractéristique telle que leur valeur du temps, il est possible, en dehors du cas de stricte équivalence des chemins, d'aboutir à un équilibre garantissant l'unicité des flux sur les itinéraires empruntés. C'est le principe sous-tendant le modèle prix-temps employé au Sétra¹⁸, par exemple. De même, on peut obtenir l'unicité en ajoutant une contrainte supplémentaire dans l'algorithme d'obtention des flux, par exemple un critère de maximisation de l'entropie (maximum de vraisemblance) (Lu et Nie, 2010).

En l'absence de ces méthodes de contournement, comme c'est le cas ici¹⁹ puisque nous ne discernons pas les usagers selon leur valeur du temps, il est peu légitime de s'attacher aux volumes par itinéraire fournis par le modèle d'affectation, qui n'ont pas la robustesse suffisante pour être le support d'analyses poussées²⁰.

Néanmoins, la structure du réseau routier, de par sa hiérarchisation, a pour conséquence de rendre, généralement, un itinéraire plus avantageux que les autres pour les usagers, au moins sur la majeure partie du trajet. Concrètement, la majorité

18. Service d'études sur les transports, les routes et leurs aménagements.

19. Une difficulté supplémentaire a émergé, pour nous, du fait que le traitement systématique des différents itinéraires empruntés par les usagers sur l'ensemble des OD n'est pas possible avec le logiciel TransCAD que nous avons employé, comme avec la très grande majorité des logiciels de modélisation du trafic.

20. Comme le signale Gaudry (2007), certains chercheurs utilisent les résultats des dernières itérations de l'algorithme de Frank et Wolfe pour obtenir une approximation des flux sur les chemins. Mais cette méthode est purement *ad hoc* et non corroborée.

des chemins empruntés sont souvent, en pratique, de la forme illustrée sur la Figure 3.5. Autrement dit, les itinéraires alternatifs ne se distinguent bien souvent qu'au niveau des extrémités. Par conséquent, l'utilisation, à des fins d'agrégation le long des chemins, d'un seul des chemins empruntés, le plus court (en temps à vide) en pratique, est une solution relativement satisfaisante. Elle pourra être entachée d'un certain niveau d'erreur, notamment en milieu urbain, mais fournira néanmoins, dans la plupart des cas, une bonne approximation de la réalité.



FIGURE 3.5.: Schéma des itinéraires dans un réseau routier

3.5.2.2. Mise en œuvre pratique

La méthode consiste donc à remplacer, dans l'équation (3.30), I_{od} par I_{od}^{pcc} , autrement dit la valeur de l'indicateur le long de ce chemin représentatif, qu'on considère égal au plus court chemin à vide :

$$I_o = \frac{\sum_d q_{od} I_{od}^{pcc}}{\sum_d q_{od} D_{od}^{pcc}}. \quad (3.31)$$

Les formules donnant I_{od}^{pcc} varient selon la grandeur étudiée. Pour les grandeurs retenues pour les trois perspectives, elles s'écrivent :

$$T_{od}^{pcc} = \sum_{a \in pcc} L_a t_a, \quad (3.32)$$

$$\xi_{od}^{pcc} = \sum_{a \in pcc} L_a \frac{x_a}{\kappa_a}, \quad (3.33)$$

$$\chi_{od}^{pcc} = \sum_{a \in pcc} L_a x_a \frac{dt_a}{dx_a}. \quad (3.34)$$

La première équation fournit la formule d'agrégation pour le temps de parcours absolu, autrement dit le coût privé pour les usagers. La dernière équation concerne le coût externe marginal. Associées, ces deux formules permettront de comparer coût privé et coût externe par zone.

En appliquant la formule (3.31) avec les formes ci-dessus pour I_{od}^{pcc} , on obtient, pour chaque zone d'origine des déplacements²¹, un indicateur agrégé pour chacune des grandeurs d'intérêt.

21. la formule s'adapte immédiatement pour les zones de destination.

3.6. Les indicateurs de congestion : du local au global

3.6.1. Quelle agrégation ?

Les approches statistique et géographique constituent des formes particulières d'agrégation des indicateurs, correspondant à des échelles et des logiques spatiales différentes. Ainsi, la segmentation introduite dans l'approche statistique, mêlant caractéristique de l'offre et de la demande, fournit une agrégation à ce niveau. Un des axes de distinction correspondant à des caractéristiques spatiales (environnement urbain), l'agrégation statistique présente une dimension spatiale. Plus précisément, à supposer que le critère de segmentation de l'environnement urbain corresponde à l'éloignement au centre urbain, l'agrégation statistique qui en résulte est bien de nature spatiale.

Toutefois, ce mode d'agrégation ne rend pas véritablement compte des logiques spatiales à l'œuvre dans le phénomène de congestion. En d'autres mots, regrouper les arcs de type autoroutier situés dans une zone peu dense ne renseigne pas sur les liens spatiaux auxquels répondent ces infrastructures. L'approche géographique que nous avons proposée fournit une méthode permettant de prendre en compte les liaisons effectives entre zones. Ce mode d'agrégation semble donc plus approprié pour rendre compte de cet aspect central des manifestations spatiales de la congestion, même si c'est au prix d'une perte de segmentation de l'offre de transport.

Ces deux approches d'agrégation, malgré leurs différences, ont un point commun essentiel : elles reposent toutes les deux sur une représentation cartographique. Par conséquent, elles sont de nature à conserver un certain niveau de détail spatial, selon le degré d'agrégation retenu. En contrepartie, elles n'ont que peu de chance de pouvoir apporter un niveau satisfaisant d'information au niveau d'agrégation le plus élevé, celui de la zone d'étude dans sa totalité. Au contraire, une méthode d'agrégation spécifiquement conçue pour rendre compte de la réponse d'un réseau dans sa globalité, en termes de congestion, peut apporter des informations précieuses à un évaluateur.

3.6.2. Agrégation d'ensemble

L'approche agrégée que nous avons retenue, employée notamment par Prud'homme (1999b), revient à considérer le réseau routier de la région comme un seul et même arc routier, sur lequel l'ensemble des véhicules présents pendant une période temporelle circuleraient à la même vitesse. C'est également l'approche retenue par Mohring (1999). Elle est bien évidemment critiquable sur bien des points, en particulier du fait de la grande variabilité spatiale de la congestion, comme le précisent

Prud'homme et Sun (2000). Par ailleurs, comme nous l'avons déjà indiqué lorsque nous avons présenté l'analyse économique de la congestion au niveau d'un arc au chapitre précédent, une telle approche repose sur l'analogie entre courbe débit - temps de parcours et courbe d'offre. Or, cette analogie néglige le fait qu'il n'y a, dans le cas du transport routier, pas véritablement d'offreur, et donc pas véritablement de courbe d'offre. En réalité, les usagers sont à la fois offreurs (collectivement) et demandeurs (individuellement).

Néanmoins, cette méthode fournit des résultats intéressants. Il s'agit en fait de fournir, à partir des données correspondant aux différentes périodes temporelles, une courbe de *réponse* du réseau à la demande sur ce réseau. On la construit en estimant la réponse par le temps total passé sur le réseau et par la demande par le nombre d'unités de trafic, en véhicules-kilomètres²². Cette méthode permet en outre de déterminer l'effet, sur le temps passé total, de l'arrivée d'une unité de trafic supplémentaire, et donc de déterminer un coût externe marginal. On attend la croissance de cette réponse lorsque la demande augmente. De plus, du fait de la présence de congestion, cette croissance se fera vraisemblablement de façon non linéaire.

La courbe de réponse est construite en utilisant les points pour lesquels nous disposons des informations de temps passé et de demande. Autrement dit, on construit la courbe « d'offre²³ » à partir des différents points d'intersection dont nous disposons entre cette courbe d'offre et les différentes courbes de demande correspondant aux différentes périodes horaires. Cela revient à considérer que cette courbe d'offre ne dépend pas de la demande, mais uniquement des caractéristiques physiques et topologiques du réseau. Généralement, on dispose des points correspondant aux heures de pointe (matin et soir) et à une heure creuse caractéristique. À ces trois points s'ajoute le point (0, 0) correspondant à une demande nulle. Ces points sont ensuite utilisés pour estimer une courbe de réponse sous la forme d'une somme d'une fonction puissance et d'une fonction linéaire, de façon à conserver une forme pseudo-BPR à cette fonction. Plus précisément :

$$R(D) = \tau_0 D + \tau_1 D^{\beta+1}, \quad (3.35)$$

où τ_0 correspond au temps de parcours à vide, τ_1 et β étant des coefficients de courbure de la fonction de réponse.

L'obtention des points de mesure du temps passé et de la demande se fait en sommant, sur l'ensemble des arcs du réseau, les unités de trafic, ce qui fournit la

22. Représenter la demande par le nombre d'usagers pourrait constituer une alternative, mais rend le résultat dépendant de la distance moyenne parcourue par les usagers, qui peut varier d'une période temporelle à une autre.

23. Avec les limite que nous avons rappelées plus haut.

demande, et les unités de temps passés, ce qui fournit la réponse du réseau. Plus précisément, si D et R sont respectivement la demande et la réponse d'un réseau A :

$$D = \sum_{a \in A} L_a x_a, \quad (3.36)$$

$$R = \sum_{a \in A} L_a x_a t_a. \quad (3.37)$$

Les valeurs fournies par ces formules permettent, à l'aide d'une régression, d'obtenir la courbe de réponse complète. Cette courbe peut ensuite fournir des résultats sur l'évolution de la somme des temps de parcours avec le volume, et donc sur les coûts externes marginaux de congestion à l'échelle du réseau régional.

On peut également définir une capacité agrégée K du réseau :

$$K = \sum_{a \in A} L_a \kappa_a \quad (3.38)$$

À partir de ces formulations, il est donc possible d'analyser les résultats de l'agrégation au niveau régional, selon les trois postures que nous avons précédemment identifiées : pour la posture du gestionnaire d'infrastructure, nous utiliserons le rapport D/K ; pour la posture de l'économiste, nous analyserons $R(D)$, c'est-à-dire le temps total passé sur le réseau, ou $R(D)/D$, le coût privé moyen et $R'(D)$, le coût marginal social. Enfin, pour la posture de l'utilisateur, nous estimerons le temps unitaire (kilométrique) moyen par usager pour chaque période temporelle, afin de constater l'évolution du temps de parcours.

3.7. Conclusion

Ce chapitre avait pour visée de proposer un cadre méthodologique d'analyse spatialisée de la congestion. Nous avons cherché, à travers la description de cette méthodologie, à respecter les différentes dimensions permettant de cerner cette question des manifestations spatiales de la congestion. Par ailleurs, trois perspectives d'acteurs ont été définies et explicitées (usager, gestionnaire, économiste), deux approches ont été décrites (statistique, géographique), enfin, une méthode d'agrégation à l'échelle régionale a été définie, donnant une vision d'ensemble que ne permettent pas les autres approches. Cette outillage méthodologique n'a pas de prétention d'exhaustivité, mais il nous semble qu'il permet des analyses suffisamment vastes du phénomène de congestion du transport, respectueux de la dimension spatiale. Nous nous sommes donc dotés, dans ce chapitre, d'un cadre d'analyse à la fois complet et suffisamment simple pour être mise en œuvre sur une région telle que

l'Île-de-France. C'est ce que nous nous proposons de faire dans le chapitre suivant.

Annexe 3.A Annexes

3.A.1 Les déplacements comme population statistique

On appelle O l'ensemble des zones d'origine des déplacements, et D celui des zones de destination. $O \cap D$ peut être non vide, O et D peuvent même être confondus. Par ailleurs, on note C_{od} l'ensemble des chemins empruntés par les usagers pour se rendre de la zone origine o à la zone destination d . Enfin, un chemin c est un ensemble d'arcs a .

En notant X le nombre total de déplacements et x_c le nombre de déplacements empruntant le chemin c , avec $\sum_c x_c = X$, on obtient la formule suivante, donnant l'écart relatif de temps de parcours moyen pour l'ensemble des déplacements :

$$\mathbb{E}(\Delta t_r) = \frac{1}{X} \sum_{o \in O} \sum_{\substack{d \in D \\ c \in C_{od}}} x_c \left(\frac{\sum_{a \in c} (t_a - t_a^0)}{\sum_{a \in c} t_a^0} \right) \quad (3.39)$$

Le même raisonnement peut être appliqué pour obtenir la variance.

3.A.2 Le problème d'unicité des résultats de l'affectation statique

On peut représenter le résultat, à l'équilibre, de l'étape d'affectation par, d'une part, une matrice indicatrice M , contenant a lignes et r colonnes, où a est le nombre d'arcs du réseau, et r le nombre d'itinéraires empruntés par les usagers, telle que si l'itinéraire j emprunte l'arc i , $M_{ij} = 1$, et d'autre part un vecteur A des flux sur les arcs. Les algorithmes d'affectation généralement employés garantissent l'unicité de ce vecteur A et de la matrice M . En revanche, le vecteur R des flux sur les itinéraires, n'est pas déterminé.

En effet, pour connaître R , il convient de résoudre le système linéaire suivant :

$$MR = A$$

Or, dans les réseaux fortement maillés tels ceux que l'on rencontre en milieu urbain, on a $r > a$, si bien que le système est sous-déterminé, entraînant une infinité de solutions.

4

Analyse de la congestion routière en Île-de-France : l'émergence de logiques spatiales

4.1. Introduction

Les objectifs de ce chapitre sont d'analyser, avec la méthodologie présentée au chapitre précédent, la variabilité spatiale de la congestion routière dans le contexte francilien, d'en faire émerger certaines logiques spatiales et de proposer une évaluation économique du coût agrégé de congestion à l'échelle régionale.

Ce chapitre prend l'Île-de-France comme terrain d'expérimentation de la méthodologie élaborée au chapitre précédent. Le choix de ce territoire vaste et complexe s'explique par différents éléments. Tout d'abord, l'aménagement et les infrastructures de transport de la région Île-de-France constituent des enjeux souvent nationaux, comme en témoigne le récent projet de Grand Paris ou, plus anciennement, le SDAURP¹ adopté par le gouvernement en 1965. En témoignent également les huit OIN² actuellement en cours dans la seule région Île-de-France, contre huit

1. Schéma Directeur d'Aménagement et d'Urbanisme de la Région Parisienne.

2. Opération d'Intérêt National, régime particulier s'appliquant à des zones d'intérêt majeur dans lesquelles L'État conserve la maîtrise de la politique d'urbanisme.

autres pour tout le reste du territoire français. L'étude du fonctionnement des systèmes de transport de cette région revêt donc un enjeu important. De plus, les volumes de déplacements qui parcourent la région parisienne constituent un enjeu pour le diagnostic de la congestion : caractérisée par des niveaux élevés de congestion, tout au long de l'année et sur des portions importantes de son réseau routier, l'Île-de-France présente également l'un des « plus gros » bouchons d'Europe sur la portion commune A4-A86. Ensuite, le Laboratoire Ville Mobilité Transport, au sein duquel nous avons effectué ce travail, cultive depuis sa création un ancrage fort dans la région Île-de-France, à travers une expertise poussée de son fonctionnement métropolitain et une maîtrise des bases de données à disposition. Enfin, le point de départ de ce travail étant une recherche réalisée pour le compte du Ministère des Transports³ au cours de laquelle l'ensemble des acteurs ayant participé étaient franciliens, le choix de cette région était difficilement évitable.

Appliquer la méthodologie d'analyse du chapitre précédent à ce contexte territorial constitue, pour notre part, un fort enjeu de recherche. En effet, la complexité du fonctionnement territorial de cette région engendre des difficultés d'interprétation des résultats, donnant tout sens aux approches complémentaires que nous avons développées. De plus, l'importance des enjeux que nous avons mentionnés donne tout leur intérêt aux résultats que nous avons obtenus. Après avoir présenté le contexte francilien, en soulignant ses enjeux et décrit la plateforme de simulation employée dans les deuxième et troisième sections, la quatrième section présente, sous forme cartographique, les résultats des simulations. Les trois dernières sections reprennent les différentes approches, développées au chapitre précédent, dans le cadre francilien.

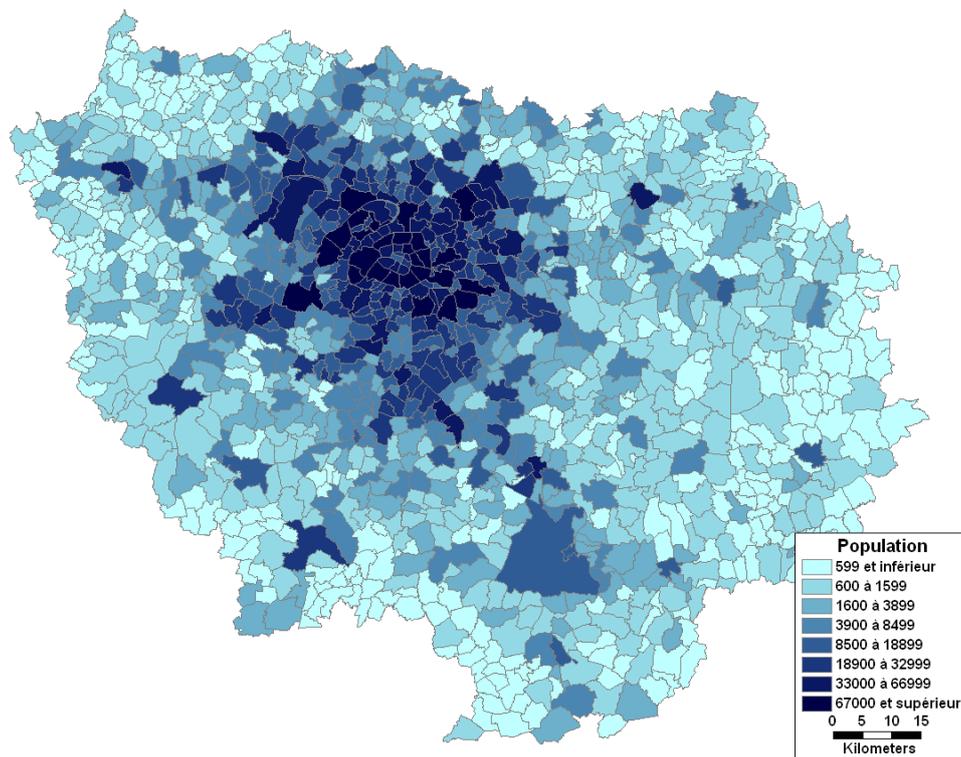
4.2. Présentation du contexte francilien

4.2.1. L'agglomération parisienne

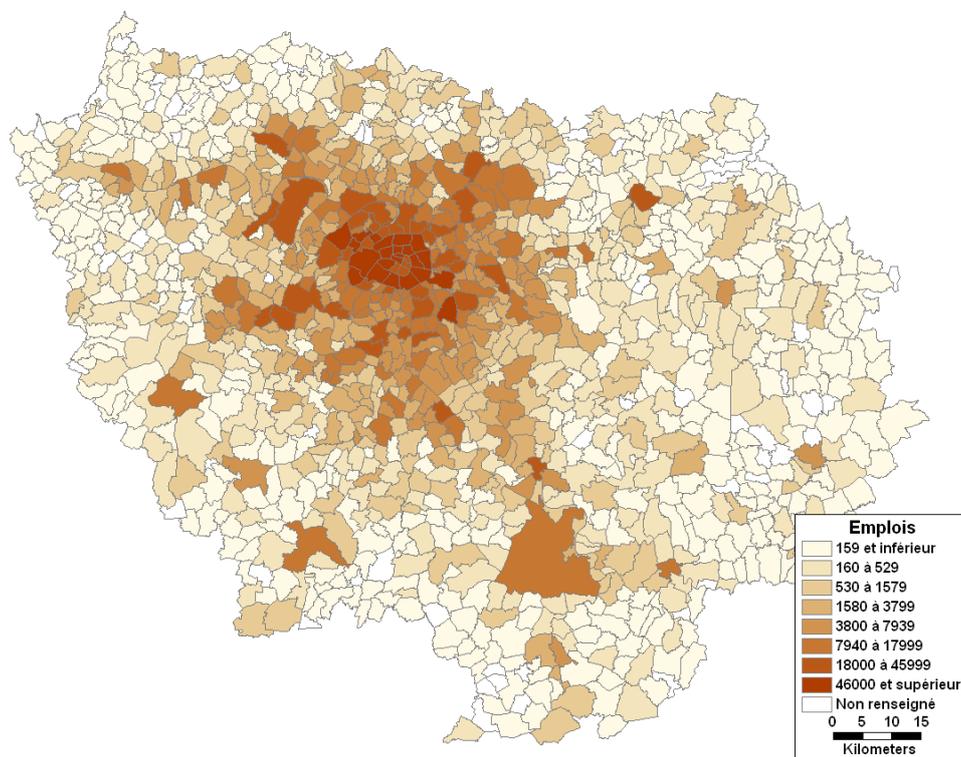
La région Île-de-France rassemblait 11,7 millions d'habitants en 2008, sur une surface urbanisée d'environ 2500 km², représentant environ 20 % de la surface totale de la région (IAU-IDF, 2009). Cette dernière valeur cache toutefois la très forte variabilité spatiale de la densité d'occupation du sol. Ainsi, la Figure 4.1 illustre cet aspect, en distinguant la population [4.1(a)] et les emplois [4.1(b)]. Paris et ses environs proches concentrent une très grande partie de la population et des emplois. Le phénomène est encore plus marqué pour l'emploi.

Cette structure monocentrique de la région parisienne est confirmée, avec des

3. Le chapitre suivant présentera les résultats de cette recherche en la replaçant dans son contexte.



(a) Population des communes en 1999



(b) Emplois des communes en 1999

FIGURE 4.1.: Population et emplois des communes d'Île-de-France en 1999. *Source* : RGP 1999 et auteur

nuances, par l'analyse de l'équilibre entre emplois et actifs, et des déplacements qui en résultent. Ainsi, en 1999, avec 1,6 millions d'emplois, Paris représentait la première destination des navettes domicile-travail : 32 % des actifs travaillant en Île-de-France s'y rendent tous les jours. Toutefois, ce chiffre baisse progressivement, puisqu'il s'élevait à 38 % en 1982 et 36 % en 1990 (Iaurif-Insee, 2003). Cette baisse s'explique par le fait que l'emploi se desserre vers la périphérie, proche couronne d'abord, grande couronne ensuite, même si ce desserrement ne concerne pas toutes les catégories d'emplois de la même manière.

En termes de déplacement, et à l'échelle régionale, en 2001, les Franciliens effectuaient, chaque jour ouvrable, environ 35 millions de déplacements, dont 44 % en voiture particulière, pour une distance parcourue totale en voiture de près de 100 millions de véhicules-kilomètres. En ce qui concerne plus spécifiquement les déplacements domicile-travail, 55 % des déplacements mécanisés sont effectués en voiture, contre 40 % pour les transports collectifs (DREIF, 2004a). Ces constats soulignent l'ampleur des enjeux de la mobilité en Île-de-France, au rang desquels la maîtrise de la congestion est au tout premier plan.

4.2.2. Le réseau routier francilien : maillé mais fortement radiocentrique

Le réseau routier francilien comprend aujourd'hui 800 km d'autoroutes et environ 200 km de voies rapides, dont un peu plus de la moitié en zone agglomérée. Ce réseau principal est très attractif, puisqu'il écoule 35 % du trafic aux heures de pointe. En comparaison, le réseau secondaire, qui dispose d'un linéaire de 10 500 km, est beaucoup moins sollicité. Le contraste est encore plus fort si l'on se limite aux déplacements liés à Paris, dont 60 % utilisent le réseau de voiries rapides urbaines (Sdrif, 2008). Ce constat est à nuancer par le fait, nous le verrons dans le chapitre suivant, que les voies rapides sont moins sujettes à la congestion, si bien que le réseau secondaire, bien que moins intensément utilisé, subit de très forts niveaux de congestion. Dans l'ensemble, la congestion routière sur le réseau francilien est importante, avec des durées de saturation qui atteignent 14 heures par jour sur le Boulevard Périphérique parisien.

En termes de structure du réseau, la Figure 4.2 illustre le caractère fortement radiocentrique du réseau magistral, qui comprend une dizaine de radiales et trois rocares concentriques (Boulevard Périphérique, A86, Francilienne). Les autres niveaux hiérarchiques sont en revanche beaucoup plus maillés, si bien qu'il est délicat d'y voir une logique structurelle claire⁴.

4. Le réseau représenté est issu du modèle Modus de la Direction régionale et interdépartementale de l'équipement et de l'aménagement d'Île-de-France (Driea-IF). La totalité de la voirie

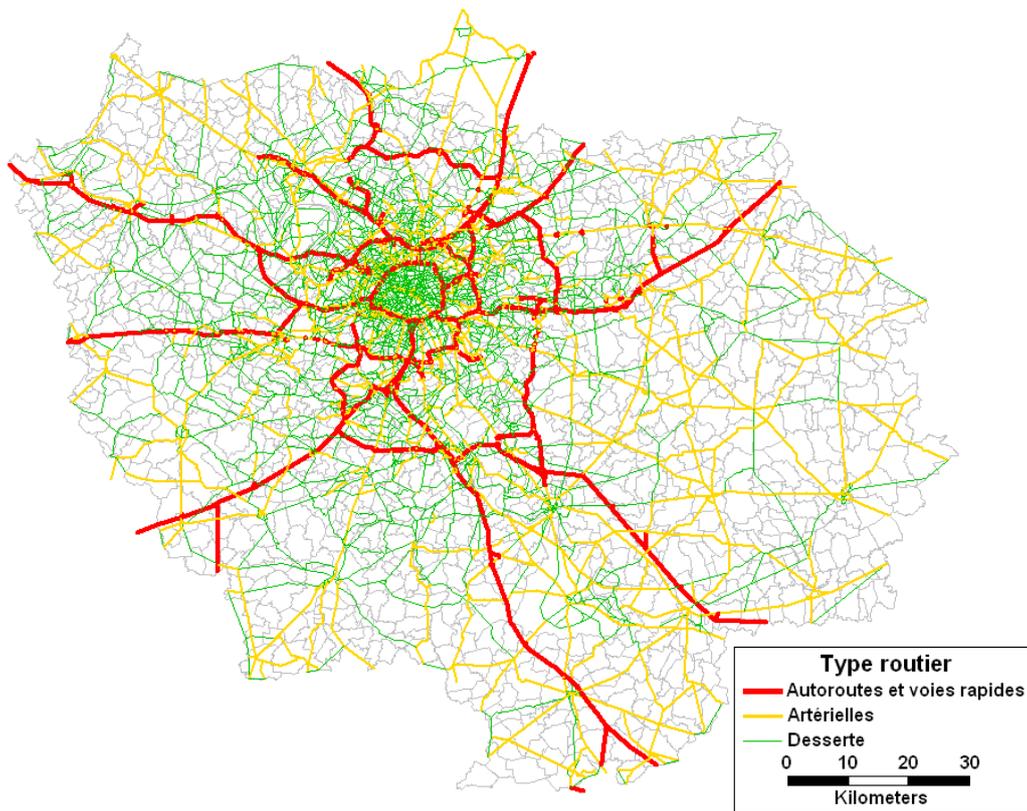


FIGURE 4.2.: Hiérarchie du réseau routier de l'Île-de-France. *Source* : auteur et DREIF (2004b)

4.2.3. Un important potentiel de variabilité

Avant d'aborder la question de la variabilité temporelle et spatiale de la congestion routière dans le contexte de l'Île-de-France, il convient de noter que cette variabilité transparaît explicitement à travers la forme fonctionnelle choisie pour cette étude, à savoir la fonction BPR⁵.

Ainsi, pour une même route, entre une période de pointe avec un taux de charge typique de $\xi_a = 1$ et une période creuse de taux de charge typique $\xi_a = 0,5$ et en supposant un exposant $\beta = 5$, alors le rapport des coûts externes marginaux de congestion atteint $2^5 = 32$. De même, pour une période donnée, entre deux routes de même fonction temps-débit mais situées l'une au centre du réseau, où l'intensité d'utilisation du réseau est élevée, et l'autre à la périphérie, avec une intensité d'utilisation beaucoup plus faible, les taux de charge respectifs seront nettement différents, de même que les temps de parcours, et plus encore les coûts externes marginaux de congestion.

4.3. La plateforme de simulation employée

Le Ministère chargé du transport a développé et maintient, au sein de la Direction régionale de et interdépartementale de l'équipement et de l'aménagement d'Île-de-France (Driea-IF, anciennement Dreif), un modèle de trafic pour aider la planification des réseaux de transport en région Île-de-France, qui englobe l'agglomération parisienne : ce modèle, appelé Modus, est un modèle de formation des déplacements en cinq étapes, respectivement de génération des déplacements, de distribution spatiale, de répartition horaire, de choix modal et d'affectation au réseau modal (DREIF, 2008).

C'est un souhait de rationalisation des choix budgétaires qui pousse la Dreif à lancer, au début des années 1970, ses premiers développements de modélisation des déplacements. Le modèle est ensuite mis à jour après le RGP⁶ 1982 et l'EGT⁷ 1983 pour prendre le nom de « Chaîne 85 », modèle avant tout adapté à l'étude du système de transport routier. Au début des années 1990, des développements significatifs sont réalisés : MODUS, acronyme signifiant « MODèle de Déplacements Urbains et Suburbains », voit le jour, avec des avancées théoriques importantes : structure en quatre étapes, modèle de choix discrets. La seconde moitié des années 1990 est consacrée à l'amélioration de MODUS : développement de la partie

n'est donc pas présente, les niveaux hiérarchiques les plus bas (réseau capillaire en particulier) n'étant pas modélisés.

5. Des résultats similaires seraient obtenus avec les autres formes fonctionnelles évoquées au chapitre méthodologique précédent.

6. Recensement Général de la Population.

7. Enquête Globale Transport.

transports collectifs, perfectionnement de la partie routière grâce aux travaux de l'INRETS. Les années 2000 voient la mise à jour de MODUS à partir de données plus récentes (RGP 1999 et EGT 2001), ainsi qu'une amélioration du calibrage, la mise en place d'un bouclage entre affectation et choix modal, etc. Ce chantier s'achève en 2008.

Le modèle d'affectation du trafic au réseau routier est un modèle statique, utilisé pour les heures de pointe du matin et du soir en jour ouvrable et pour l'heure interpointe. Le réseau routier est représenté par 15 000 nœuds et 40 000 arcs environ, avec des fonctions temps-débit de type Davidson modifié⁸. Concernant la demande de déplacement, le territoire d'étude est divisé en plus de 1300 zones d'affectation du trafic. Les matrices origine-destination ont été évaluées par motif de déplacement à partir de l'EGT de 2001-2002, et régulièrement mises à jour à partir de prévision de la demande. Ainsi, la matrice routière globale que nous avons employée a été calibrée sur les déplacements de 2003, et ajustée pour représenter la demande de 2007. Elle contient, pour l'heure de pointe du matin (une heure moyenne de 7h à 9h) environ 1,6 millions de déplacements.

En termes de logiciel-support, nous avons simulé le trafic routier avec le logiciel TransCAD, en approchant par des fonctions BPR les fonctions Davidson du logiciel DaVISUM utilisé par la Dreif pour le modèle Modus. Nous avons estimé des fonctions temps-débit de type BPR, avec un exposant β de l'ordre de 5 (entre 4 et 7) et un facteur α s'étalant de 0,1 à 1, des voies les plus urbaines aux voies rapides. Nous appellerons ce modèle, différent du modèle utilisé par la Dreif, *Quasi-Modus*.

L'algorithme d'affectation, utilisé par TransCAD pour les affectations à l'équilibre de l'utilisateur que nous avons employées, est celui de Frank-Wolfe (Caliper, 2007). L'avantage de cet algorithme, outre sa simplicité et sa relative rapidité de convergence, est surtout sa stabilité satisfaisante (Bar-Gera *et al.*, 2010), qui autorise, avec précaution néanmoins, des analyses le long des chemins suivis par les usagers. En effet, dans une affectation statique à l'équilibre de l'utilisateur, seuls les volumes de trafic sur les arcs sont uniques ; les chemins empruntés, quant à eux, ne le sont pas.

La Figure 4.3 présente le réseau modélisé au sein de Quasi-Modus. On constate que la finesse du zonage est très variable, passant d'un détail très fin au centre de

8. Plus précisément, les courbes utilisées sont de type Davidson pour les taux de charge inférieurs à 1, et de type parabolique au-delà de la saturation (DREIF, 2004b). L'emploi de telles courbes permet de pallier le fait que la pente des courbes Davidson devient infinie lorsque le taux de charge s'approche du paramètre b . De plus, la Dreif met en œuvre la méthode de la « demande écrêtée », qui rend compte des temps d'attente à l'entrée d'un arc saturé. Toutefois, comme nous l'avons vu dans chapitre précédent, et rappelé dans ce chapitre, l'utilisation de relations moyennées dans le temps permet déjà, à elle seule, de rendre compte des temps d'attente. En particulier, la pente très élevée, sans être infinie, des courbes BPR autour de la saturation est une manière de prendre en compte le très fort allongement des temps, lié au temps d'attente, lorsque la demande dépasse la capacité pendant une période donnée.

l'agglomération, à des zones plus agrégées à mesure que la distance au centre augmente. En ce qui concerne la modélisation du réseau routier, l'ensemble du réseau de voiries rapides urbaines est présent, ainsi que les voies artérielles principales. Le réseau de voiries collectrices et de desserte locale, en revanche, n'est pas entièrement modélisé. Nous verrons comment ce manque de détail, qui permet de réduire les temps de calculs, complique, voire biaise en partie, l'analyse des résultats du modèle, en termes de congestion routière.

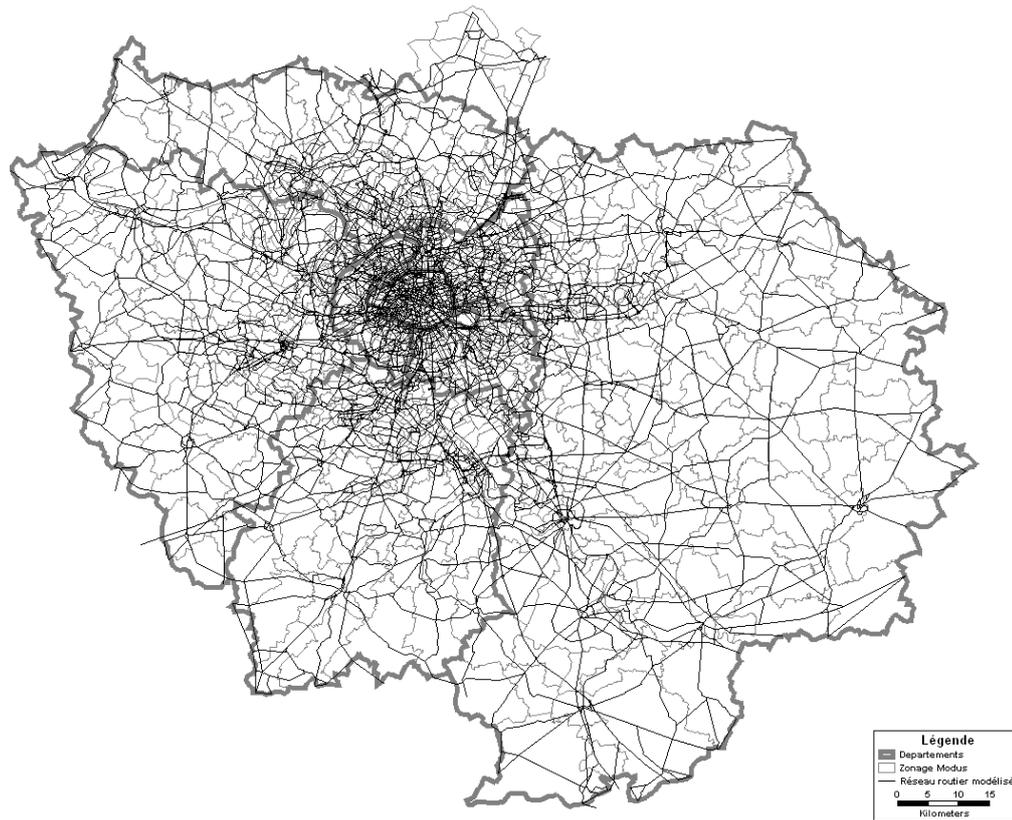


FIGURE 4.3.: Le réseau modélisé au sein de Quasi-Modus. *Source* : DREIF (2004b)

4.4. Les états simulés du trafic : une première approche de la variabilité de la congestion en Île-de-France

Les cartes présentées dans cette section sont issues d'affectations routières, à partir des matrices fournies par la DRIEA-IF, correspondant à l'heure de pointe du matin, à l'heure de pointe du soir, et à une heure creuse, typiquement de milieu de journée. Elles permettent de rendre compte visuellement de la variabilité, tant

spatiale que temporelle, des différents indicateurs de congestion considérés. Pour l'ensemble des indicateurs, nous avons choisi de ne faire figurer, dans le corps du texte, que des cartes centrées sur la zone dense de l'Île-de-France, car c'est dans cette zone que se situent la majorité des enjeux liés à la congestion. Néanmoins, les cartes de l'ensemble de la région figurent dans l'Appendice A.2.

Comme exposé au chapitre 3, l'indicateur retenu dans la perspective de l'utilisateur est le temps de parcours relatif au temps à vide. Cet indicateur est représenté, sur la Figure 4.4 pour la zone centrale de l'Île-de-France à l'heure de pointe du matin.

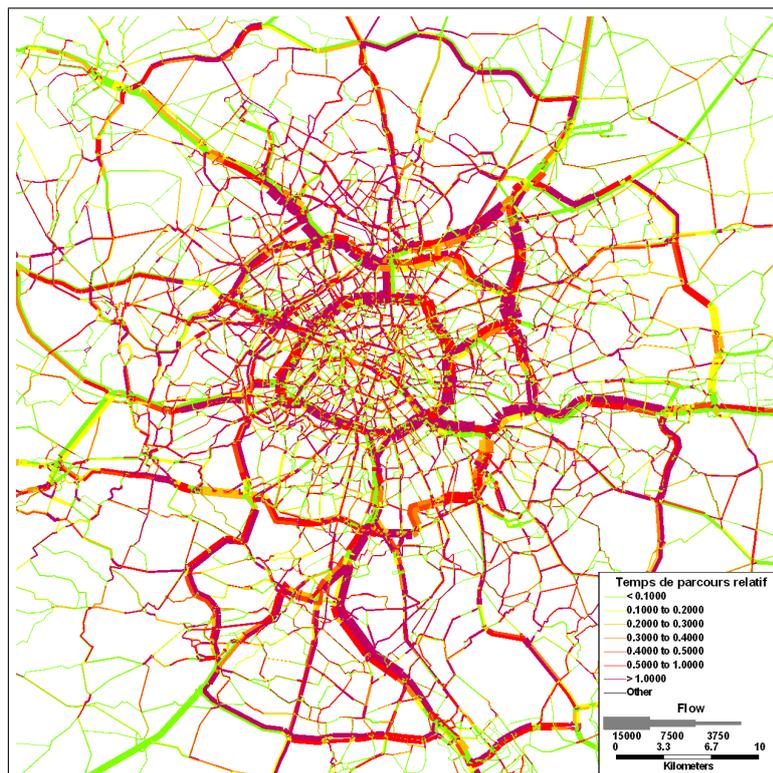


FIGURE 4.4.: Les écarts de temps relatifs à l'heure de pointe du matin. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

On constate en premier lieu que pendant l'heure de pointe du matin, le réseau subit une forte dégradation de la qualité de service offerte. Néanmoins, cette dégradation de la qualité de service est très variable, selon le lieu et le type d'infrastructure. Ainsi, on observe des écarts relatifs importants sur les grands axes, massivement employés, pendant les heures de pointe, en particulier en direction de Paris. La voirie secondaire semble toutefois relativement épargnée. Mais là encore, il convient de nuancer : les résultats fournis par le modèle sont très contrastés, puisque certaines zones sont, dans leur ensemble, sujettes à des écarts relatifs de temps de parcours élevés — les environs immédiats de La Défense, en particulier

— tandis qu'à d'autres endroits, seules certaines infrastructures sont touchées — l'autoroute A4 entre Marne-la-Vallée et Paris, par exemple.

Le même type d'analyse est possible pour la grandeur identifiée précédemment comme intéressant de façon prioritaire le gestionnaire d'infrastructure. La Figure 4.5 présente, pour l'heure de pointe du matin, le taux de charge sur le réseau routier de l'Île-de-France.

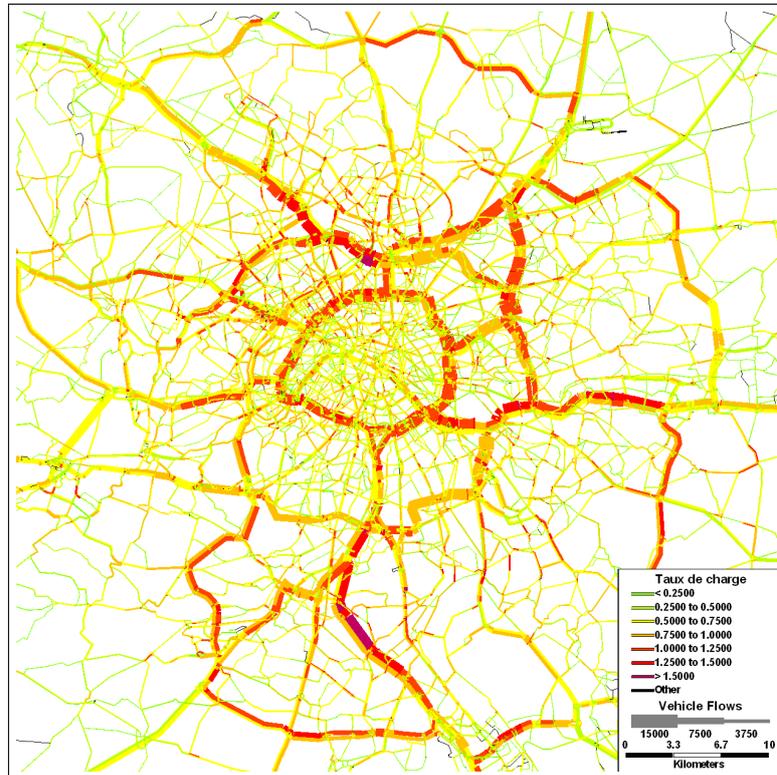


FIGURE 4.5.: Les taux de charge à l'heure de pointe du matin. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

La carte fait apparaître une variabilité spatiale plus faible que dans le cas des temps de parcours relatifs. Certains grands axes, fortement chargés à l'HPM, se dégagent nettement, mais l'allure de la carte demeure néanmoins plus uniforme. Cette différence de comportement entre le taux de charge et le temps de parcours peut s'expliquer par la non-linéarité du temps de parcours en fonction du taux de charge : une faible différence de taux de charge peut entraîner, suivant la plage où se situe le taux de charge, ou bien une très faible différence de temps de parcours — dans la zone des taux de charge faibles — ou bien au contraire un très forte différence de temps de parcours — dans la zone des taux de charge élevés.

Comme nous l'avons indiqué précédemment, la posture de l'économiste est celle

qui nous intéresse plus particulièrement ici. L'indicateur que nous avons retenu pour cette posture est le coût social marginal de congestion, et plus précisément, la différence entre ce coût social marginal et le coût moyen subi par l'utilisateur marginal, autrement dit le coût externe. La Figure 4.6 présente, pour l'HPM, l'estimation de ce coût externe sur le réseau routier francilien.

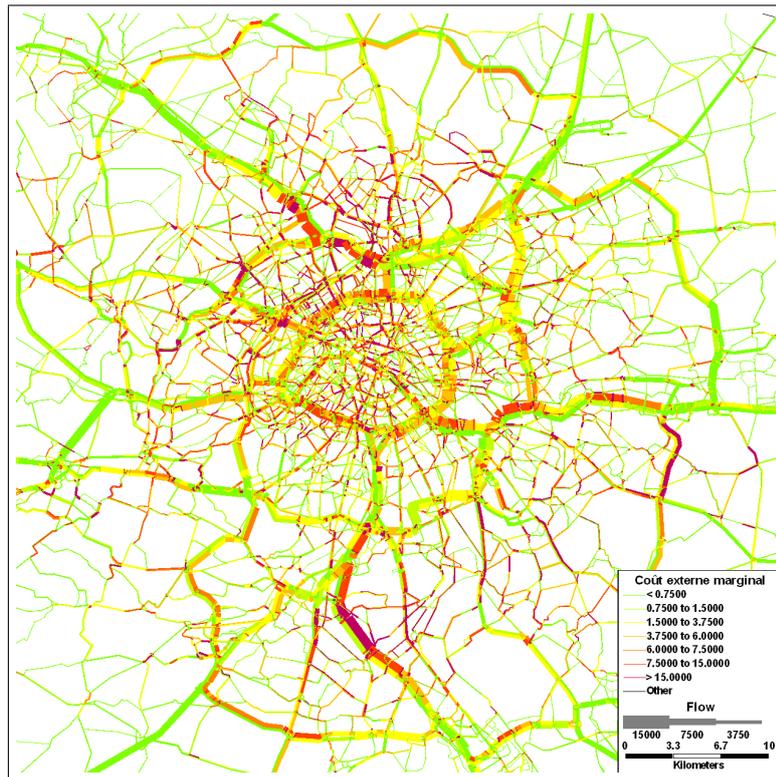


FIGURE 4.6.: Le coût externe marginal à l'heure de pointe du matin. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

Elle met en évidence une très grande variabilité spatiale de ce coût externe marginal. Cela s'explique par les caractéristiques des infrastructures et par la structure spatiale des volumes de trafic, mais surtout par la très forte non linéarité du coût marginal, en particulier son augmentation très rapide pour les taux de charge élevés. La variabilité du coût externe reflète en partie celle des coûts moyens, puisque, nous l'avons vu, la courbe de coût marginal croît partout plus vite que celle de coût moyen.

4.5. Diagnostic statistique de la congestion en Île-de-France

Ce diagnostic repose sur les analyses exposées au chapitre précédent dans le cadre de l'approche statistique.

4.5.1. Mise en œuvre de la segmentation

Nous avons employé, dans cette étude, une classification semi-automatique basée sur le type routier codé dans le réseau du modèle Modus, qui distingue 13 types différents. Nous les avons regroupés pour former trois types. En pratique, ils sont construits de manière similaires à Riff Brems *et al.* (2002) :

- Les voies rapides urbaines. Il s'agit des autoroutes et voies assimilées, qu'elles soient de disposition radiale, telles A1, A4, A6, A13, ou en rocade, telles la Francilienne, A86 et le Boulevard Périphérique. Toutes ces voies sont à accès réservé et ne comportent pas de feux de signalisation.
- Les voies artérielles. Il s'agit principalement de voies d'assez grande capacité, reliant des pôles urbains. En pratique, les (anciennes) routes nationales entrent dans cette catégorie.
- Les voies de desserte. Cette catégorie regroupe les autres voies, dont la capacité est généralement plus réduite. Leur rôle est également plus local.

Nous avons distingué trois zones, en fonction de leur éloignement au centre de Paris. Il s'agit d'une segmentation relativement frustrante, puisqu'elle ne distingue *a priori* que les zones de forte demande de celles où la demande est plus faible. En réalité, cette segmentation fait plus que cela, car compte tenu de la structure à tendance monocentrique de l'Île-de-France (Pottier *et al.*, 2007), la zone la plus éloignée correspond également à des distances parcourues par les usagers plus grandes, tandis que la zone centrale est caractérisée par des trajets plus courts. Or, la structure d'usage des infrastructures diffère, du fait des caractéristiques de ces infrastructures, selon la portée des déplacements.

Les trois zones sont les suivantes :

- La zone centrale. Elle correspond à Paris. Elle est caractérisée par une très forte densité, aussi bien des emplois que des ménages.
- La proche périphérie. Il s'agit des trois départements limitrophes à Paris, à savoir les Hauts-de-Seine, la Seine-Saint-Denis et le Val-de-Marne, autrement dit la petite couronne. Cette zone est caractérisée par une densité plus faible que la précédente, mais néanmoins assez élevée.
- La grande périphérie. Elle correspond aux quatre départements de la grande couronne de l'Île-de-France, à savoir l'Essonne, la Seine-et-Marne, le Val-d'Oise

et les Yvelines. La zone est caractérisée par des densités faibles voire très faibles aux frontières de la région, et en dehors de quelques pôles urbains isolés (pôles d'équilibre).

Cette décomposition par milieu géographique a bien évidemment ses limites, en particulier parce que pour des raisons de simplicité nous avons suivi les délimitations administratives. Bien que ce choix puisse apparaître réducteur, nous avons volontairement restreint la typologie, de manière à rendre plus lisibles et donc plus facilement interprétables les résultats issus de la segmentation.

Les segments ont été finalement définis par croisement des deux dimensions présentées ci-dessus. Nous avons donc obtenu, pour chaque période temporelle, neuf segments, ce qui donne, en considérant également la période temporelle comme une dimension de segmentation, un total de 27 segments.

Les Tables 4.1 et 4.2 présentent les effectifs de chacun des segments d'étude. Ces effectifs sont fournis d'une part en véhicules-kilomètres et d'autre part en longueur d'infrastructures, puisque ces deux populations statistiques sont employées, suivant les perspectives que nous présenterons à la section suivante. Bien évidemment, si la population des véhicules-kilomètres varie fortement d'une période temporelle à une autre, le réseau étant fixe, la population des kilomètres d'infrastructures est la même pour les trois périodes temporelles. Par ailleurs, le tableau présente le linéaire d'infrastructures *deux sens confondus*.

| HPM | Rapides | Artérielles | Desserte | Total |
|----------------|-----------|-------------|-----------|------------|
| Zone centrale | 524 848 | 193 959 | 600 550 | 1 319 357 |
| Proche périph. | 1 772 868 | 992 776 | 1 097 429 | 3 863 072 |
| Grande périph. | 3 779 929 | 3 256 163 | 2 816 048 | 9 852 140 |
| Total | 6 077 645 | 4 442 898 | 4 514 026 | 15 034 569 |
| HC | | | | |
| Zone centrale | 335 304 | 81 885 | 262 851 | 680 040 |
| Proche périph. | 839 351 | 494 142 | 433 184 | 1 766 678 |
| Grande périph. | 1 746 357 | 1 417 005 | 1 158 107 | 4 321 469 |
| Total | 2 921 011 | 1 993 032 | 1 854 143 | 6 768 186 |
| HPS | | | | |
| Zone centrale | 510 143 | 187 430 | 604 401 | 1 301 974 |
| Proche périph. | 1 471 273 | 852 991 | 949 858 | 3 274 121 |
| Grande périph. | 3 298 033 | 2 671 467 | 2 421 347 | 8 390 847 |
| Total | 5 279 449 | 3 711 887 | 3 975 606 | 12 966 942 |

TABLE 4.1.: Effectifs des segments d'analyse pour la population des véhicules-kilomètres. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

| | Rapides | Artérielles | Desserte | Total |
|----------------|---------|-------------|----------|-------|
| Zone centrale | 54 | 70 | 408 | 532 |
| Proche périph. | 198 | 411 | 1 101 | 1 710 |
| Grande périph. | 748 | 2 784 | 4 181 | 7 713 |
| Total | 1 000 | 3 265 | 5 690 | 9 955 |

TABLE 4.2.: Effectifs des segments d'analyse pour la population des kilomètres d'infrastructures. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

4.5.2. La variabilité à travers les trois perspectives d'analyse

Nous complétons ici la première approche, cartographique, de la variabilité de la congestion, par une analyse statistique exploratoire des indicateurs étudiés.

4.5.2.1. Perspective de l'utilisateur

L'analyse statistique exploratoire fournit les valeurs présentées dans la Table 4.3, pour les trois périodes précédemment étudiées.

| Période temporelle | Moyenne | Ecart-type | Dispersion rel. |
|--------------------|---------|------------|-----------------|
| HPM | 0,85 | 1,97 | 2,31 |
| HPS | 0,49 | 1,02 | 2,09 |
| HC | 0,06 | 0,20 | 3,24 |
| Période temporelle | Médiane | 90 % | Max |
| HPM | 0,28 | 2,21 | 198 |
| HPS | 0,16 | 1,33 | 47,9 |
| HC | 0,01 | 0,16 | 16,3 |

Source : calculs de l'auteur

TABLE 4.3.: Moyennes, écarts-types, dispersions relatives et quantiles de la distribution des temps de parcours relatifs, pour trois périodes temporelles (heure de pointe du matin - HPM, heure de pointe du soir - HPS, heure creuse - HC). *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

Ces résultats confirment la variabilité temporelle de l'indicateur de temps de parcours, puisque les moyennes, pour les trois périodes temporelles, sont fortement différentes. Le temps réellement mis par les usagers est en moyenne supérieur au temps nominal de 6 % en heure creuse et de près de 85 % à l'heure de pointe du matin. L'ordre de ces valeurs moyennes correspond par ailleurs bien aux attentes : le temps de parcours relatif est plus élevé à l'HPM qu'à l'HPS, et celui de l'HC est proche du temps de parcours à vide.

Ensuite, au sein d'une période temporelle donnée, la variabilité spatiale est également très importante, comme l'indiquent les valeurs de dispersion relative trouvées. On constate de plus que les dispersions relatives croissent avec la moyenne, suggérant que lorsque le trafic s'intensifie, les écarts entre infrastructures s'intensifient également (l'heure creuse fait exception). La Figure 4.7 illustre, par un diagramme Boxplot C_5/C_{95} , la structure des distributions des temps de parcours relatifs. Elle montre clairement la croissance de la dispersion avec la croissance de la moyenne (et de la médiane).

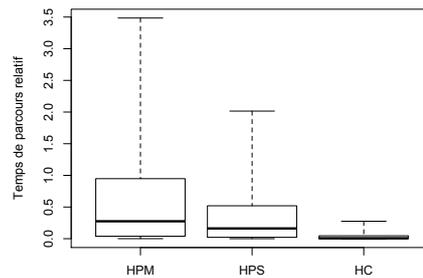


FIGURE 4.7.: Diagramme Boxplot C_5/C_{95} de la distributions des temps de parcours relatifs. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

Les valeurs des quantiles remarquables fournissent une information importante : la médiane est beaucoup plus faible que la moyenne. Même la valeur du quantile à 90 % est inférieure à la moyenne. Cela suggère qu'en réalité, la moyenne est très fortement déplacée par la présence de quelques arcs extrêmement congestionnés, qui engendrent des temps de parcours très élevés. Nous verrons, dans la suite de l'analyse, que ces valeurs très élevées peuvent trouver une part d'explication dans la modélisation trop grossière du réseau de voiries de faible capacité.

La mise en œuvre de la segmentation proposée au chapitre précédent fournit les résultats illustrés, pour l'heure de pointe du matin, sur la Figure 4.8. Les segments y sont désignés par un numéro allant de 11 à 33, selon le codage suivant : le premier chiffre indique la zone géographique (1 pour Paris, 2 pour la proche périphérie, 3 pour la grande périphérie), le second chiffre indique le type routier (1 pour les voies rapides, 2 pour les voies collectrices, 3 pour les voies de desserte). Par ailleurs, la valeur moyenne y est indiquée, pour chaque segment, par un symbole circulaire.

On constate tout d'abord que la dispersion des temps de parcours est très variable selon les segments. En pratique, les valeurs sont très dispersées pour les segments se situant en zone 3, ainsi que sur les infrastructures de type collectrice en zone 2, et d'une façon générale, sur les voies de desserte.

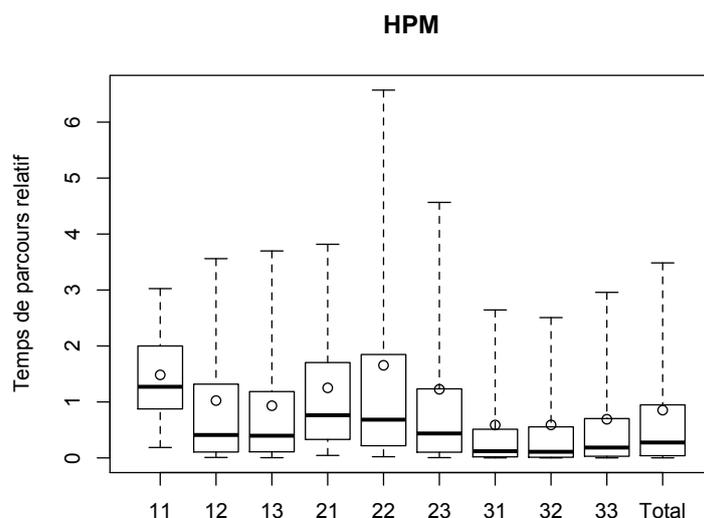


FIGURE 4.8.: Diagramme Boxplot C_5/C_{95} des distributions des temps de parcours relatifs par segment. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

Cette remarque s'applique également aux autres indicateurs et aux autres périodes temporelles, comme nous le constaterons par la suite. Plusieurs explications de ce phénomène peuvent être envisagées :

- Le segment « voies de desserte » regroupe en réalité une très grande diversité de voies, aussi bien en terme de capacité d'écoulement que d'usage ou d'environnement (urbain, périurbain, rural). La segmentation par couronne permet de prendre en compte une partie seulement de la diversité liée à l'environnement, en particulier pour les proches et lointaines périphéries.
- La modélisation des voies de desserte n'est pas aussi détaillée que celle des autres types de voies. Cela signifie d'une part que de nombreuses voies ne sont pas codées dans le modèle, pour des raisons de temps de calcul et d'occupation en mémoire, mais également que les caractéristiques codées au sein du modèle pour ces voies ne correspondent pas toujours finement à la réalité. Cet état de fait provient du caractère faiblement stratégique de ces voies pour la DRIEA-IF, dont l'attention se porte plutôt sur les voiries principales structurantes et dimensionnantes du réseau.
- La grande couronne est beaucoup moins finement codée que Paris et la proche couronne, que ce soit au niveau du réseau qu'au niveau des zones de demande.
- La localisation des connecteurs, qui s'effectue fréquemment sur les voies de desserte, entraîne, au niveau du réseau local entourant le nœud d'insertion du

connecteur, des niveaux de trafic parfois très élevés, assez éloignés de la réalité. Ce problème, fréquemment rencontré dans les modèles d'affectation n'a pas de solution simple (Qian et Zhang, 2010). Nous avons tenté de le pallier en reliant chaque zone à l'aide de plusieurs connecteurs, mais cette solution n'est pas idéale, car elle entraîne une sous-estimation des trafics locaux sur les réseaux, ceux-ci empruntant plus volontiers les connecteurs — très peu sensibles à la congestion — plutôt que le réseau local.

En réalité, l'ensemble de ces explications peut être résumé par le constat de Bovy et Jansen (1983) : ces auteurs ont montré que les erreurs les plus importantes sur les volumes de trafic ont toujours lieu pour les niveaux de hiérarchie les plus bas. Autrement dit, dans notre cas où la hiérarchie comporte trois niveaux, seuls les deux niveaux supérieurs sont susceptibles d'apporter des résultats peu erronés.

En revanche, le diagramme indique que pour les segments se situant dans les zones 1 et 2, c'est-à-dire à Paris ou en petite couronne, le temps de parcours relatif moyen se situe dans l'intervalle entre les quantiles à 25 % et 75 %, voire encore plus proche de la médiane. Cela signifie que pour ces segments, définir un temps de parcours relatif moyen n'est pas dénué de sens. Une telle valeur fournit une représentation satisfaisante de la situation effectivement rencontrée par les usagers sur ces infrastructures.

En ce qui concerne la segmentation par type d'environnement urbain, il est assez difficile d'interpréter les résultats, car aucune zone ne se dégage par des temps relatifs particulièrement élevés ou faibles. On peut néanmoins noter que la grande périphérie connaît des temps de parcours relatifs sensiblement plus faibles que les deux autres zones. Les diagrammes Boxplot pour l'heure creuse et l'heure du pointe du soir figurent en Annexe 4.A.2. Ils amènent les mêmes commentaires, atténués bien sûr pour l'heure creuse.

Grâce aux analyses présentées sur la Figure 4.8 mais surtout dans la Table 4.12 en Annexe 4.A.1, nous avons constaté que les voies rapides sont moins sujettes à la variabilité spatiale. Autrement dit, pendant une période temporelle donnée, les voies rapides offrent les conditions de circulation les plus homogènes. Ce résultat montre qu'en l'absence de connaissance précise de son itinéraire, un usager a intérêt à privilégier les voies rapides. En effet, c'est sur celles-ci qu'il peut le mieux prévoir son temps de parcours.

Plus précisément, concernant ces questions de fiabilité et de risque pour l'usager, l'exploration statistique de l'élasticité temps-volume, qui L'application de ces formules fournit les résultats présentés dans la Table 4.4.

On constate que la dispersion relative du coût externe relatif est plus faible à l'HPM qu'à l'HPS, et encore bien supérieure en HC. Il s'agit d'un résultat contre-intuitif. Il tend à montrer que plus le réseau est congestionné, plus l'impact potentiel

| Période temporelle | Moyenne | Ecart-type | Dispersion rel. |
|--------------------|---------|------------|-----------------|
| HPM | 1,41 | 1,30 | 0,92 |
| HPS | 1,06 | 1,10 | 1,03 |
| HC | 0,22 | 0,42 | 1,93 |
| Période temporelle | Médiane | 90% | Max |
| HPM | 1,07 | 3,42 | 4,97 |
| HPS | 0,69 | 2,83 | 4,90 |
| HC | 0,04 | 0,69 | 4,68 |

Source : calculs de l'auteur

TABLE 4.4.: Moyennes, écarts-types, dispersions relatives et quantiles de la distribution des élasticités temps-volume, pour trois périodes temporelles.

Source : DRIEA/SCEP/DPAT et calculs de l'auteur.

d'un incident s'uniformise. En d'autres mots, les itinéraires tendent à devenir semblables en termes de risque de fiabilité. Il s'agit d'une différence assez marquée avec l'indicateur de temps de parcours, dont la variabilité spatiale augmente avec le niveau de congestion du réseau.

4.5.2.2. Perspective du gestionnaire d'infrastructure

L'analyse statistique exploratoire, menée pour les taux de charge, fournit les valeurs présentées dans la Table 4.5, pour les trois périodes précédemment étudiées :

| Période temporelle | Moyenne | Ecart-type | Dispersion rel. |
|--------------------|---------|------------|-----------------|
| HPM | 0,39 | 0,29 | 0,75 |
| HPS | 0,33 | 0,27 | 0,81 |
| HC | 0,17 | 0,16 | 0,99 |
| Période temporelle | Médiane | 90 % | Max |
| HPM | 0,35 | 0,78 | 2,29 |
| HPS | 0,29 | 0,71 | 1,73 |
| HC | 0,12 | 0,41 | 1,31 |

Source : calculs de l'auteur

TABLE 4.5.: Moyennes, écarts-types, dispersions relatives et quantiles de la distribution des taux de charge, pour trois périodes temporelles. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

La Figure 4.9 illustre, par un diagramme Boxplot C_5/C_{95} , la structure des distributions des taux de charge.

On constate que le taux de charge moyen sur le réseau est plus élevé à l'heure de

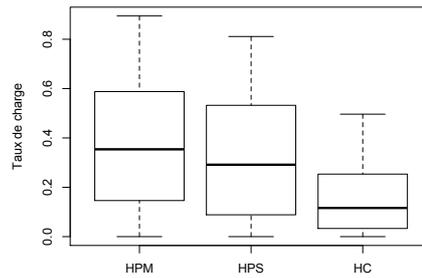


FIGURE 4.9.: Diagramme Boxplot C_5/C_{95} de la distributions des taux de charge.
Source : DRIEA/SCEP/DPAT et calculs de l'auteur.

pointe du matin qu'à celle du soir, et que les heures creuses sont nettement moins chargées encore. La variabilité temporelle est donc bien présente. Comme la carte le faisait pressentir, la dispersion des valeurs de taux de charge est beaucoup plus faible que dans le cas des temps de parcours relatifs. Elle reste néanmoins significative. Il est par ailleurs intéressant de constater que cette dispersion relative tend à diminuer lorsque les taux de charge augmentent. Cela provient du fait que l'écart-type augmente moins vite que la moyenne. On assiste donc à une uniformisation des taux de charge lorsque le trafic s'intensifie.

En ce qui concerne les quantiles remarquables de la distribution, on remarque que la médiane et la moyenne sont très proches, et ce pour les trois périodes temporelles. Par ailleurs, on constate que les arcs dont le taux de charge est très proche de 1 représentent, dans tous les cas, moins de 10 % du linéaire d'infrastructure, même dans le cas le plus chargé de l'heure de pointe du matin.

Du point de vue du gestionnaire d'infrastructure, il ressort de cette analyse que si la variabilité spatiale des taux de charge existe bel et bien, son amplitude est relativement modérée. Autrement dit, la valeur moyenne du taux de charge sur le réseau constitue un bon indicateur du fonctionnement du réseau dans son ensemble. La variabilité temporelle est, par contre, assez forte, ce qui justifie l'étude de plusieurs périodes temporelles.

Les résultats par segment sont illustrés sur la Figure 4.10 avec les mêmes notations que précédemment.

Le caractère quasi normal des distributions des taux de charge se vérifie également pour les distributions par segment, à l'exception du segment des voies rapides parisiennes, c'est-à-dire principalement le Boulevard Périphérique. C'est également le segment le plus chargé en moyenne, comme en valeur médiane.

Contrairement à ce que nous avons noté pour les temps de parcours relatifs,

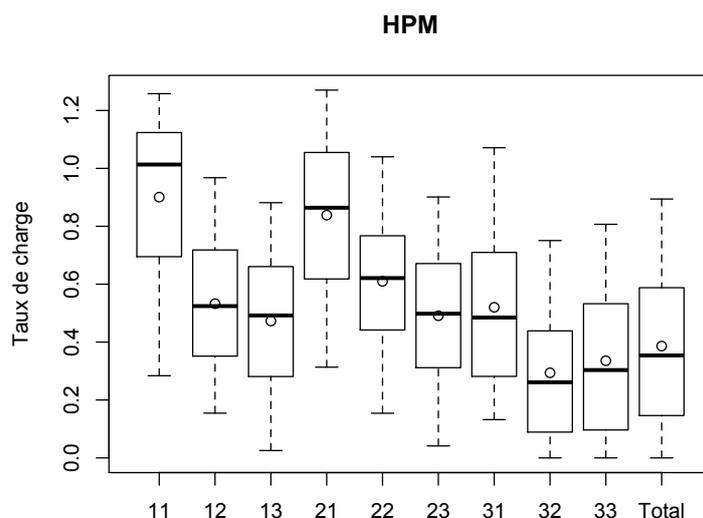


FIGURE 4.10.: Diagramme Boxplot C_5/C_{95} des distributions des taux de charge par segment. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

on peut, ici, observer plus clairement un ordre se dégager : les taux de charge ont tendance à décroître avec l'éloignement au centre, et, au sein d'une même zone géographique, des voies rapides aux voies de desserte. Cela reflète clairement d'une part la structure urbaine de l'Île-de-France — très dense au centre et assez fortement monocentrique — et d'autre part les effets des caractéristiques des infrastructures — avec un taux de charge plus élevé, une voie rapide offrira un temps de parcours plus faible qu'une petite voie de desserte.

Il ressort de ces constats que pour l'exploitant routier, la segmentation que nous avons proposée offre une bonne cohérence et permet de mieux appréhender la réalité des conditions de trafic sur le réseau. En effet, on le constate sur le diagramme, le taux de charge moyen sur le réseau, s'il fournit une première indication, est très imparfait, tandis que la simple segmentation que nous proposons améliore fortement le diagnostic. Ainsi, elle met clairement en évidence le cas, préoccupant, du Boulevard Périphérique.

4.5.2.3. Perspective de l'économiste

L'analyse statistique exploratoire que nous avons menée repose sur Le Tableau 4.6 présente les résultats, sur le réseau francilien, de l'étude statistique exploratoire utilisant les formules précédentes.

La Figure 4.11 illustre, par des diagrammes Boxplot C_5/C_{95} , la structure de la

| Période temporelle | Moyenne | Ecart-type | Dispersion rel. |
|--------------------|---------|------------|-----------------|
| HPM | 4,04 | 12,4 | 3,06 |
| HPS | 2,51 | 6,81 | 2,72 |
| HC | 0,31 | 1,24 | 3,97 |
| Période temporelle | Médiane | 90 % | Max |
| HPM | 1,08 | 9,63 | 1 469 |
| HPS | 0,63 | 6,41 | 951 |
| HC | 0,03 | 0,75 | 101 |

Source : calculs de l'auteur

TABLE 4.6.: Moyennes, écarts-types, dispersions relatives et quantiles de la distribution des coûts externes marginaux, pour trois périodes temporelles. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

distribution des coûts externes marginaux.

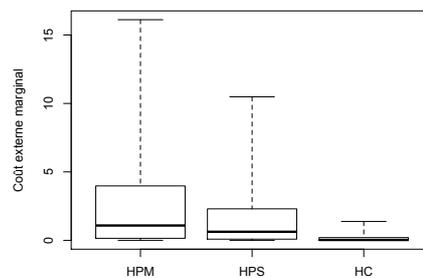


FIGURE 4.11.: Diagramme Boxplot C_5/C_{95} des distributions des coûts externes marginaux. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

Les résultats montrent une extrême variabilité du coût externe marginal. Les indices de dispersion relative sont en effet très élevés. La significativité de la moyenne lorsque la dispersion est à ce point élevée est à nuancer fortement. Néanmoins, on remarque que les résultats statistiques confirment les constats réalisés précédemment sur les cartes : les coûts externes marginaux sont faibles en heure creuse, plus élevés pendant l'heure de pointe du soir, et encore plus pendant l'heure de pointe du matin.

Une conséquence à signaler est la difficulté apparente, dans un tel contexte de variabilité des coûts externes, d'imposer une tarification au coût marginal (sous la forme d'un péage kilométrique, par exemple), telle que le recommande la théorie économique. En effet, les coûts externes dépendent très fortement de la période considérée et des itinéraires suivis par les usagers.

En ce qui concerne les quantiles remarquables, il s'avère que la distribution des coûts externes marginaux est très déformée, encore plus que celle des temps de parcours relatifs. En effet, en dehors de l'heure creuse, la moyenne est très supérieure à la valeur du quantile à 90 %. Les raisons de cette déformation sont, on le verra un peu plus loin, assez semblables à celles à l'origine de la déformation de la distribution des temps de parcours.

Il est intéressant, selon nous, de noter que l'élasticité temps-volume étudiée dans le cadre de la posture de l'utilisateur est également le coût externe relatif au coût moyen. Une seconde analyse, au regard de ce constat, de la Table 4.4 permet de voir que le coût externe relatif moyen est, pendant les heures de pointe, supérieur à 1. Bien sûr, le coût marginal total, somme du coût moyen et du coût externe, est dans tous les cas supérieur au coût moyen. Néanmoins, les résultats indiquent qu'en moyenne, pendant les heures de pointe, un usager marginal sur le réseau impose aux autres usagers déjà présents un coût supplémentaire supérieur au coût qu'il subit lui-même.

De manière semblable aux deux indicateurs précédents, la Figure 4.12 illustre les résultats de la segmentation pour le coût externe marginal, à l'heure de pointe du matin.

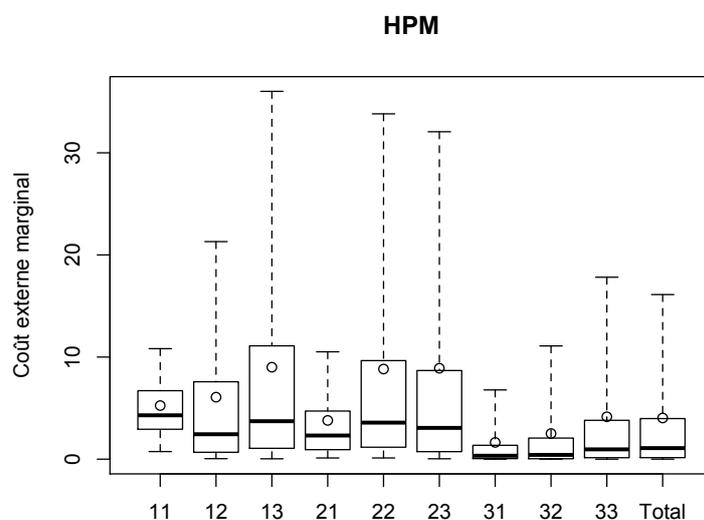


FIGURE 4.12.: Diagramme Boxplot C_5/C_{95} des distributions des coûts externes marginaux par segment. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

Le diagramme correspondant aux coûts externes marginaux fait apparaître une

très grande variabilité, et la segmentation permet d'en réduire partiellement l'amplitude (*cf.* Annexe 4.A.1) puisque la dispersion relative au sein des segments est toujours plus faible que celle sur la totalité du réseau. Néanmoins, comme dans le cas des temps de parcours relatifs, les segments de voies de desserte sont caractérisés par une dispersion très élevée, ainsi que par un écart très important de la moyenne et de la médiane, avec des moyennes au-delà du quantile à 95 %. En pratique, cela signifie que quelques arcs (moins de 5 %, et même moins de 1 % en volume) présentent des coûts externes marginaux extrêmement élevés, déplaçant ainsi la moyenne, sans que la médiane ne soit fortement affectée.

Les conséquences sur la moyenne à l'échelle du réseau sont très importantes. L'indicateur de coût externe marginal moyen est donc relativement peu fiable à l'échelle de l'agglomération, mais fournit en revanche, pour les segments des voies rapides et collectrices, une assez bonne indication. Le cas des voies de desserte, lié là encore à la modélisation peu fine du réseau à cette échelle, nécessitera, pour un traitement opérationnel, une méthode particulière que nous proposerons dans le chapitre suivant.

Malgré ces limites, il est possible de mettre en évidence un ordre entre les segments : le coût externe marginal moyen croît des voies rapides aux voies de desserte, et en règle générale, de la périphérie lointaine à la zone centrale. On peut y voir d'une part une conséquence de l'effet capacité déjà évoqué plus haut — une voie rapide est plus à même de recevoir un usager supplémentaire sans faire subir un coût trop élevé aux autres usagers, et d'autre part un effet de l'intensité d'usage — un usager supplémentaire imposera un coût plus important s'il s'insère dans un réseau déjà intensément utilisé.

4.6. Diagnostic géographique de la congestion en Île-de-France

L'application de la méthode d'agrégation précédente fournit des résultats que nous proposons de visualiser sous forme de cartes, car elles permettent de faire émerger les structures spatiales.

Avant de détailler l'analyse des cartes, il convient de préciser l'interprétation à leur donner. Ainsi, concernant le temps de parcours, une zone pour laquelle l'indicateur de temps de parcours à l'origine est élevé est une zone au départ de laquelle les usagers subissent, le long de leurs itinéraires, des niveaux élevés de congestion. Cela peut avoir deux origines : ou bien la zone est située dans un environnement fortement sujet à la congestion, si bien que tous les itinéraires partant de cette zone sont congestionnés ; ou bien les usagers de cette zone ont une destination privilégiée, et empruntent donc très majoritairement un itinéraire, qui se trouve être conges-

tionné. Il est souvent difficile de trancher entre les deux hypothèses. Néanmoins, pour certaines situations particulières, il est possible d'en avoir une idée, et nous le précisons le cas échéant.

Nous présentons les cartes de l'heure de pointe du matin, et ce pour une raison principale : cette thèse s'intéresse plus particulièrement aux déplacements domicile-travail qui représentent, à l'heure de pointe du matin, la très grande majorité des déplacements (Dreyfus, 2005). C'est en revanche moins clair pour l'heure du pointe du soir. La conséquence est double. Tout d'abord, les cartes de l'heure de pointe du soir sont plus délicates à interpréter, car il est moins aisé de comprendre pourquoi une zone subit des temps de parcours longs, ou produit des coûts externes marginaux élevés. Par ailleurs, les cartes de l'heure de pointe du matin permettent d'apporter des éléments pour notre analyse des déplacements domicile-travail. Cependant, les cartes correspondant à l'heure de pointe du soir sont présentées dans l'Appendice A.3.

4.6.1. Les victimes de la congestion

La Figure 4.13 présente le coût temporel unitaire privé, autrement dit le temps de parcours moyen d'un kilomètre pour les usagers se rendant dans les différentes zones de destination, à l'heure de pointe du matin. Cet indicateur spatialisé permet de visualiser les zones de destination dont les usagers s'y rendant le matin subissent le plus de congestion. En d'autres mots, la carte présente les lieux d'emplois où les usagers se rendant à leur travail subissent, le long de leur itinéraire, le plus de congestion.

On y constate encore une fois la grande variabilité spatiale des temps de parcours moyens. En revanche, on voit apparaître des structures spatiales à grande échelle, qui mettent en évidence une structure sous-jacente de fonctionnement de l'agglomération. En particulier, on notera les temps de parcours plus élevés en moyenne au centre de l'agglomération, ainsi qu'autour de certains pôles d'emplois secondaires, comme Saint-Denis ou Orly (cela est lié à la présence à la fois de l'aéroport d'Orly et du marché de Rungis). Le fait que les destinations les plus congestionnées sont généralement situées à l'ouest de Paris est une traduction de la concentration des emplois (en particulier très qualifiés) dans la cette partie de la région.

La Figure 4.14 présente, pour la même période temporelle, le même indicateur de temps de parcours moyen, mais cette fois par zone d'origine des déplacements. Là également, des structures spatiales émergent. Elles sont localisées en proche périphérie de Paris et sont globalement plus dispersées que dans le cas des zones de destination. Cela est lié au fait que la dispersion des lieux d'habitation est plus importante que celle des emplois. Autrement dit, les zones pour lesquelles le temps de parcours unitaire à l'origine est élevé sont à la fois plus nombreuses et

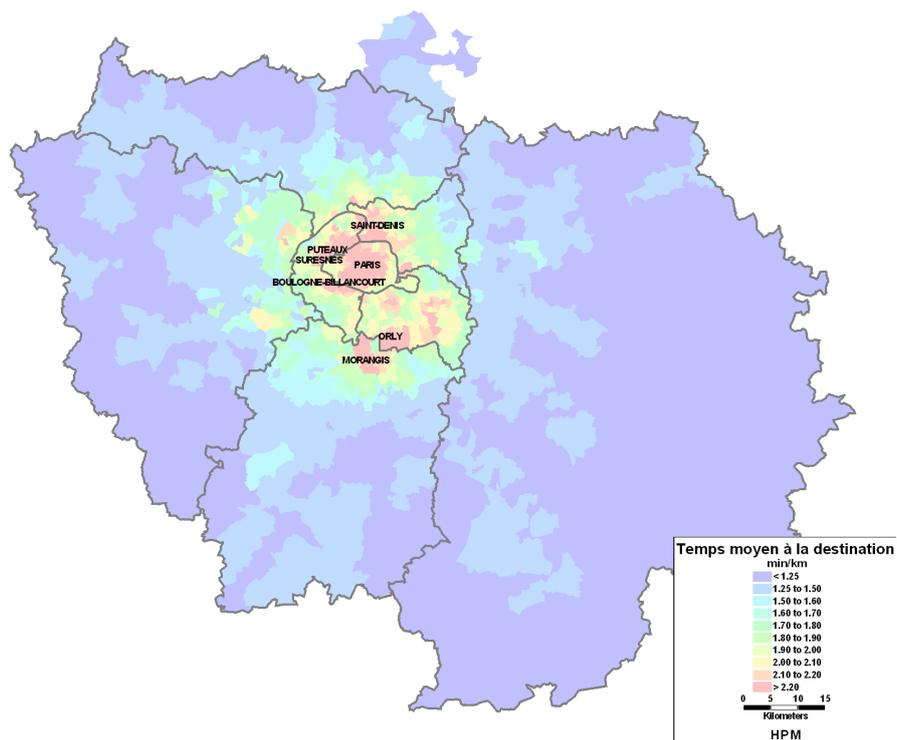


FIGURE 4.13.: Temps de parcours unitaire moyen par zone de destination des déplacements, à l'HPM. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

moins concentrées que les zones semblables à la destination. On peut néanmoins noter que Paris et la proche périphérie se distinguent d'une façon générale par des temps de parcours unitaires relativement élevés en moyenne, même si le centre de Paris semble relativement épargné. Enfin, on constate que la périphérie ouest est plus touchée que l'est. La périphérie plus lointaine semble également globalement épargnée. De manière générale, on constate, en dehors de Paris, une baisse du coût unitaire moyen avec la distance au centre de l'agglomération.

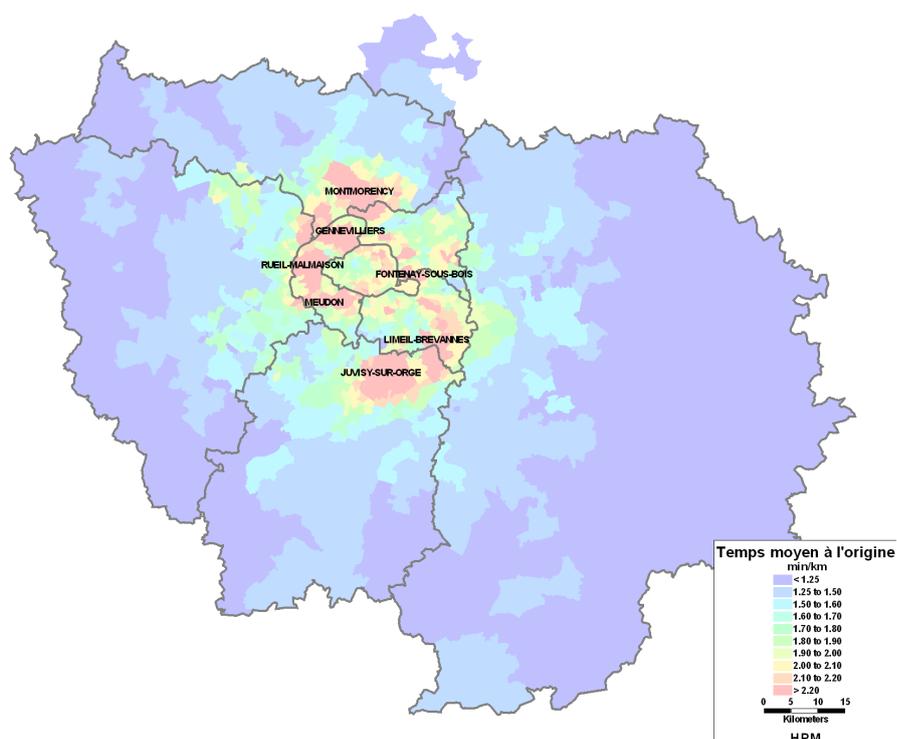


FIGURE 4.14.: Temps de parcours unitaire moyen par zone d'origine des déplacements, à l'HPM. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

4.6.2. L'occupation de la voirie

En ce qui concerne les taux de charge, la Figure 4.15 met en évidence un phénomène assez remarquable : à l'heure de pointe du matin, les habitants situés non loin de l'A86, autrement dit autour d'un rayon de 10 à 15 km de Paris, font face à des taux de charge très élevés en moyenne.

Comme nous l'avons précisé en introduction de cette analyse spatiale, deux causes peuvent être à l'origine de ce constat : l'environnement immédiat de ces zones est de manière générale très congestionné, ou la mobilité des usagers de ces zones est principalement orientée le long d'itinéraires congestionnés, éventuellement à longue

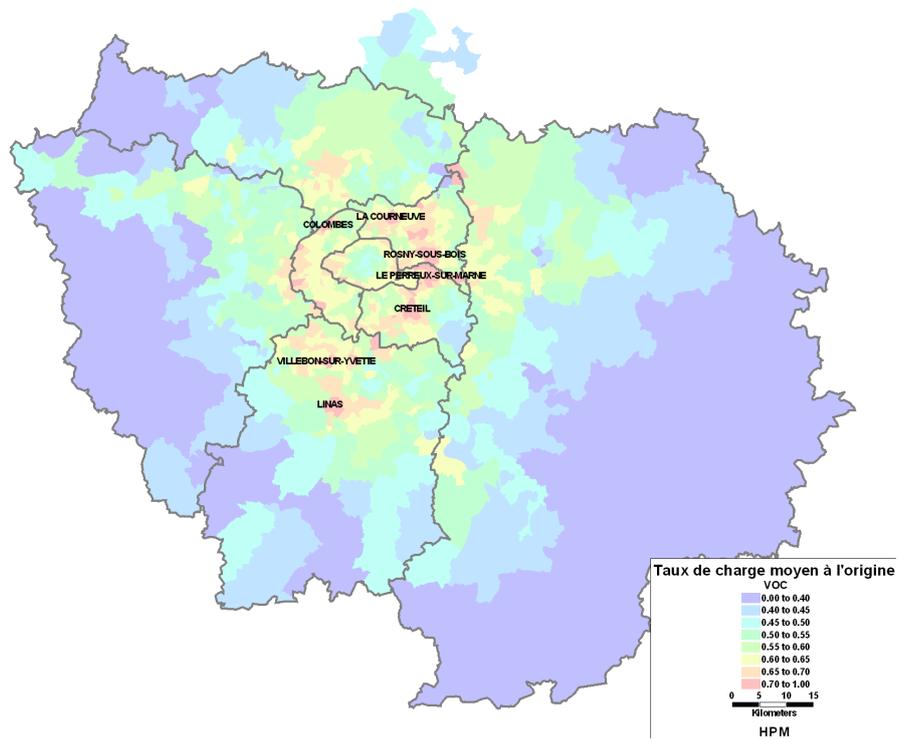


FIGURE 4.15.: Taux de charge moyen par zone d'origine des déplacements, à l'HPM.
Source : DRIEA/SCEP/DPAT et calculs de l'auteur.

distance. Dans le cas présent, il est probable que les deux phénomènes se renforcent l'un l'autre. En effet la proche périphérie est très fortement congestionnée, comme on a pu le constater au cours de l'analyse par segment du réseau. On peut penser que ce constat résulte du fait que son réseau, de capacité souvent moyenne, est emprunté par tous les usagers se rendant à Paris depuis l'ensemble de la périphérie, petite et grande couronne, ainsi que par l'ensemble des usagers se rendant en périphérie lointaine, depuis Paris ou la proche couronne. En conséquence, vu de cette frange de la proche couronne, le réseau semble subir de forts taux de charge. De plus, les grands axes de cette zone étant soumis à des niveaux élevés de congestion, la mobilité de ses habitants, orientée le long de ces axes, subit cette congestion de plein fouet.

4.6.3. Les usagers à l'origine de la congestion

Si les analyses qui précèdent ont permis de déterminer les zones où les usagers subissent le plus de congestion lors de leurs déplacements routiers, les analyses qui suivent visent à identifier les zones les plus émettrices de congestion, et ce en tant qu'origine ou destination des déplacements. Ainsi, la Figure 4.16 présente le coût externe kilométrique moyen par zone de destination, à l'heure de pointe du matin, tandis que la Figure 4.17 fait de même par zone d'origine des déplacements.

Dans les deux cas, des structures spatiales se dégagent clairement. Elles sont localisées de manière assez semblable à celles des Figures 4.13 et 4.14. En revanche, les valeurs se différencient nettement, puisque les valeurs élevées du coût externe sont près de 5 fois plus élevées que celles du temps de parcours. Cette différence d'ordre de grandeur est due au fait, déjà évoqué, en particulier au 4.5.2.3, que le coût externe est fréquemment plus élevé que le coût moyen.

La similitude constatée entre les structures spatiales présentées par les deux cartes est due au caractère extrêmement local de l'externalité de congestion. Si une externalité telle que l'émission de gaz à effet de serre est caractérisée par le fait que les émetteurs ne subissent qu'une infime partie de ce qu'ils émettent, la congestion est quant à elle très majoritairement subie par ceux qui l'émettent, et ce à plusieurs niveaux. Tout d'abord, en première approximation, seuls les usagers de la route subissent la congestion en termes de perte de temps. Il s'agit donc d'une externalité de club. Ensuite, la congestion est spatialement localisée au niveau des arcs congestionnés, si bien que ce sont les usagers qui créent le plus de congestion qui en subissent également le plus, puisqu'ils sont sur la route qu'ils congestionnent. On ne doit donc pas s'étonner d'une telle similitude de distribution. Ce résultat suggère toutefois que la congestion ne pose pas la question de l'équité spatiale de manière aussi aiguës que peuvent le faire d'autres externalités (émission de gaz à effet de serre).

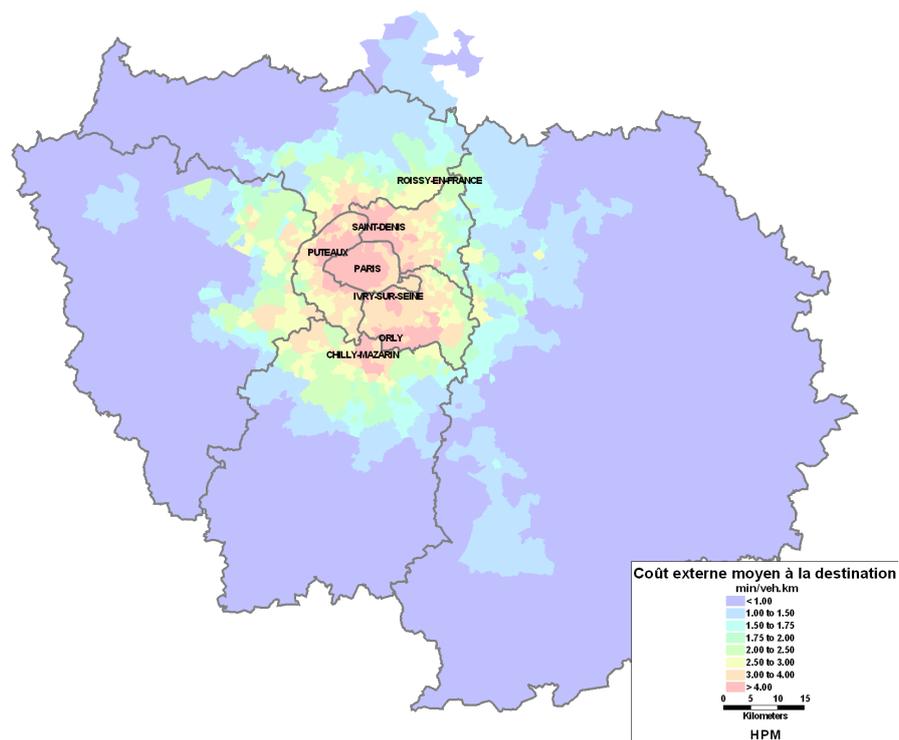


FIGURE 4.16.: Coût externe moyen kilométrique par zone de destination des déplacements, à l'HPM. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

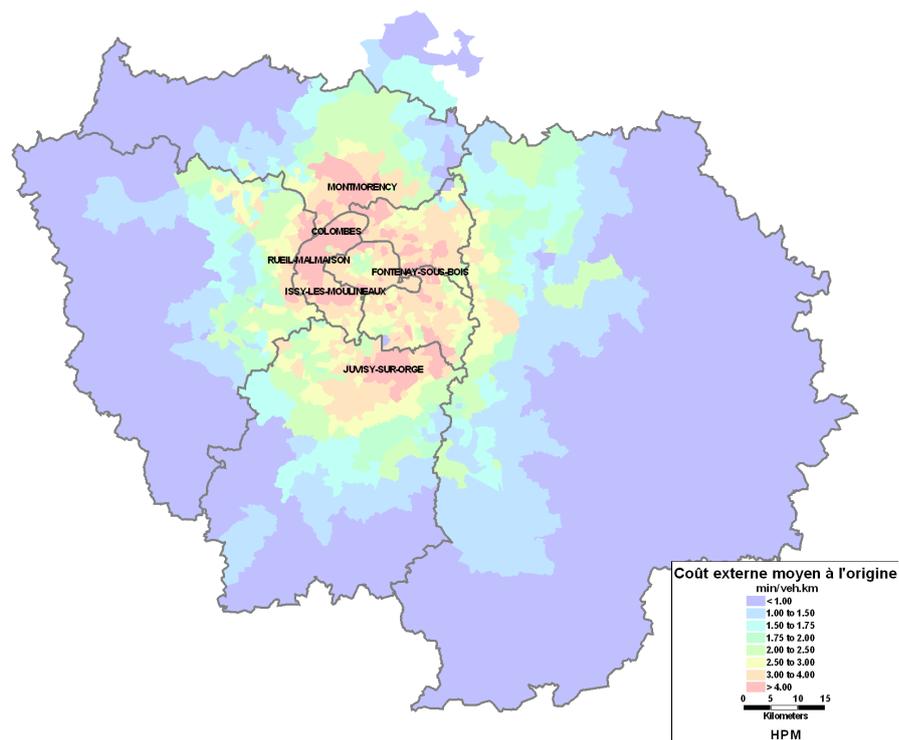


FIGURE 4.17.: Coût externe moyen kilométrique par zone d'origine des déplacements, à l'HPM. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

L'indicateur de coût externe kilométrique moyen n'est cependant pas suffisant pour un diagnostic complet des émissions de congestion. En effet, la longueur des trajets effectués par les usagers est tout aussi importante. Le coût social doit en effet être évalué dans sa totalité. Les Figures 4.18 et 4.19 présentent les coûts externes moyens *par déplacement*, vus des zones de destination ou des zones d'origine, respectivement.

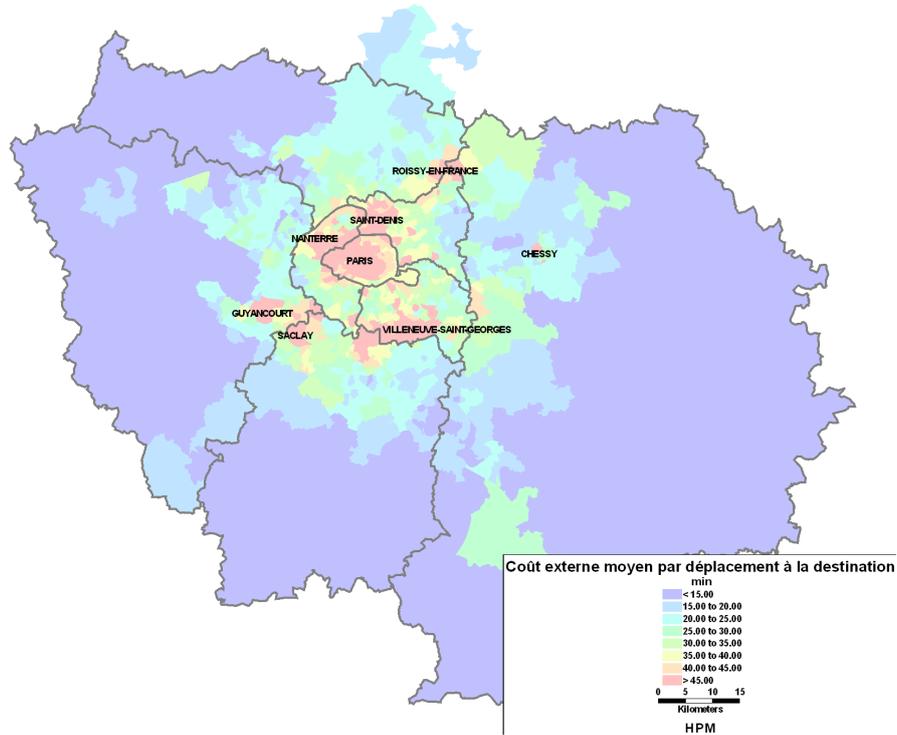


FIGURE 4.18.: Coût externe moyen par déplacement par zone de destination des déplacements, à l'HPM. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

La prise en compte de la longueur des déplacements joue donc peu dans le cas des zones de destination. Cela est probablement lié à la très forte concentration des emplois. En revanche, il en va tout autrement pour les zones d'origine. Pour ces dernières, on constate une dispersion beaucoup plus forte des zones fortement émettrices de congestion. Cela est la conséquence directe des grandes distances parcourues par ces usagers de la grande périphérie parisienne, en particulier à l'est et au nord, pour se rendre sur leur lieu de travail. Ce constat vient donc nuancer le résultat précédent : si, en termes d'intensité de dommages la proche périphérie peut être incriminée, en termes absolus, ce sont les zones situées en grande couronne à la limite de la proche couronne qui produisent le plus de congestion.

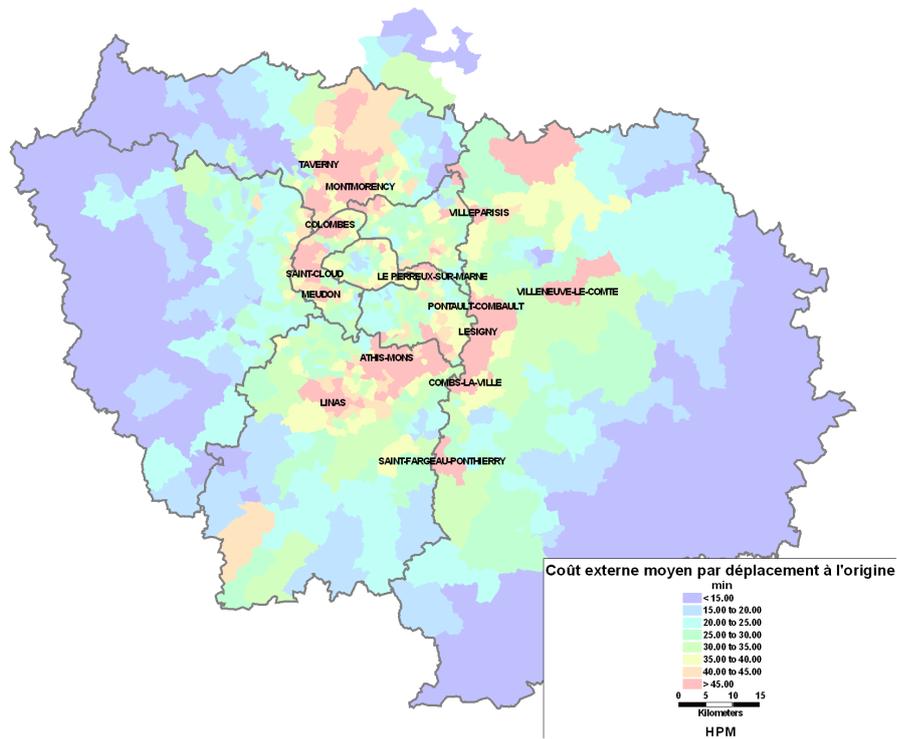


FIGURE 4.19.: Coût externe moyen par déplacement par zone d'origine des déplacements, à l'HPM. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

4.7. Diagnostic économique agrégé de la congestion en Île-de-France

Nous avons cherché, dans les deux sections précédentes, à rendre compte et à comprendre la variabilité spatiale de la congestion. Cette section est consacrée à une évaluation de la congestion à une échelle régionale agrégée.

4.7.1. Temps passé et coût marginal social

L'application de la méthode présentée au 3.6.2 fournit les résultats de la Table 4.7, issus d'une régression linéaire⁹ du type $\ln(R - \tau_0 D) = \ln(\tau_1) + (\beta + 1) \ln D$. Nous avons évalué τ_0 de manière à minimiser la somme des écarts quadratiques, ce qui a fourni : $\tau_0 = 0,014$ h/km, correspondant à une vitesse à *vide* d'environ 71,5 km/h. La Figure 4.20 fournit une illustration du résultat du modèle sur les données initiales (sans passage aux logarithmes).

| | Estimation | Erreur | Student | p-valeur |
|---------------|--------------------|-----------------------|---------|---------------|
| $\ln(\tau_1)$ | -51,71 | 0,01461 | -3539 | 0,000180 *** |
| $\beta + 1$ | 3,872 | $9,011 \cdot 10^{-4}$ | 4297 | 0,000148 *** |
| R^2 ajusté | 1 | | | |
| Fischer | $1,846 \cdot 10^7$ | | | 0,0001482 *** |

TABLE 4.7.: Résumés statistiques des résultats du modèle. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

La valeur trouvée pour β amène quelques commentaires. Ainsi, en poursuivant l'analogie avec une fonction BPR au niveau des arcs, il s'agit d'une valeur significativement plus basse que les coefficients β utilisés dans le modèle d'affectation. Cette valeur agrégée de β ne correspond pas non plus à la moyenne, à l'équilibre pour chaque période temporelle, des β des arcs pondérés par les véhicules-kilomètres. Cette fonction agrégée met donc en évidence un effet de réseau : si le réseau régional était remplacé par un seul arc avec une fonction de temps de parcours BPR, son exposant serait bien inférieur aux exposants des BPR de chaque arc du réseau initial. Ces différentes valeurs sont comparées dans la Table 4.8.

En d'autres mots, l'arrivée, sur le réseau dans son ensemble, d'une unité de trafic supplémentaire se manifeste moins fortement que si cette arrivée se faisait sur un

9. Cette régression étant fondée sur quatre points uniquement, nous n'avons pas la prétention de la présenter comme extrêmement robuste. D'autres points de mesure permettraient d'en confirmer la robustesse. Toutefois, elle permet de définir une courbe de réponse du réseau utile à la détermination d'un coût global de congestion.

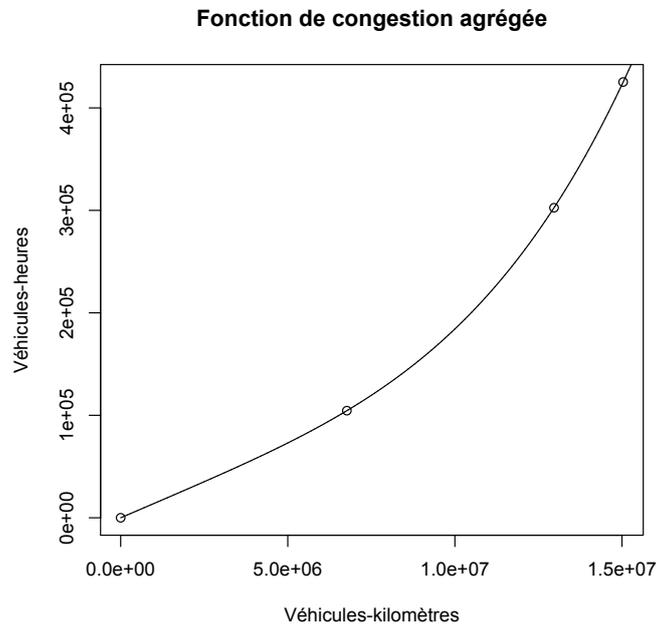


FIGURE 4.20.: Résultat du modèle de régression. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

| Méthode | β |
|---------------------|---------|
| Régression | 2,872 |
| Moyenne /km | 4,9646 |
| Moyenne /véh.km HPM | 4,9638 |
| Moyenne /véh.km HC | 4,9636 |
| Moyenne /véh.km HPS | 4,9641 |

TABLE 4.8.: Valeurs d'un β moyen par différentes méthodes. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

seul arc du réseau par un effet de « dilution » sur l'ensemble des arcs sur réseau. Cet effet est lié au comportement optimisateur des usagers, qui choisissent leur itinéraire, c'est-à-dire l'enchaînement d'arcs qui minimise leur temps de parcours. Si un arc présente un temps de parcours important, un usager supplémentaire pourra avoir tendance à préférer un itinéraire alternatif, diminuant ainsi son effet sur le temps passé global.

Avec la relation de la forme $R = \tau_0(D + \tau_1/\tau_0 D^{\beta+1})$ que nous avons obtenue, on peut déterminer un coût social marginal de congestion à l'échelle de l'agglomération. Il est fourni par la dérivée de $R(D)$: $R'(D) = \tau_0(1 + \tau_1/\tau_0(\beta+1)D^\beta)$. En effet, cette dérivée représente l'effet d'un véhicule-kilomètre supplémentaire sur le temps total passé par l'ensemble des usagers de l'agglomération. La Table 4.9 donne la valeur de ce coût marginal social pour les trois périodes temporelles étudiées, ainsi que le coût moyen kilométrique, c'est-à-dire le temps passé en moyenne par un usager pour parcourir un kilomètre du réseau, fourni par $R(D)/D$, et le coût externe marginal, différence entre les deux grandeurs précédentes.

| Coût | HPM | HC | HPS |
|-------------------------------|------|------|------|
| Moyen (min/km) | 1,70 | 0,93 | 1,40 |
| Marginal social (min/véh.km) | 4,16 | 1,18 | 3,01 |
| Externe marginal (min/véh.km) | 2,46 | 0,25 | 1,61 |

TABLE 4.9.: Coût moyen kilométrique et coût marginal social aux trois périodes temporelles considérées. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

On constate la croissance du coût moyen kilométrique lorsque le trafic s'intensifie. De même, le coût marginal social augmente, encore plus fortement. On remarque également que les valeurs trouvées pour le coût externe marginal sont sensiblement différentes des moyennes obtenues à partir des moyennes sur le réseau précédemment, tout en restant dans le même ordre de grandeur.

La Figure 4.21 présente les courbes de coût moyen et de coût marginal pour l'agglomération francilienne dans son ensemble.

À partir des fonctions de coût moyen et de coût marginal représentées sur ce graphique, il est possible, à l'aide de l'équation (3.12) présentée au chapitre précédent, de calculer une borne supérieure de la perte sèche associée à la congestion en Île-de-France, et ce pour les différentes périodes temporelles considérées. L'équation (3.12) se réécrit de la façon suivante :

$$H(D) = \tau_1 \frac{\beta}{(1 + \beta)^{1+1/\beta}} D^{\beta+1} \quad (4.1)$$

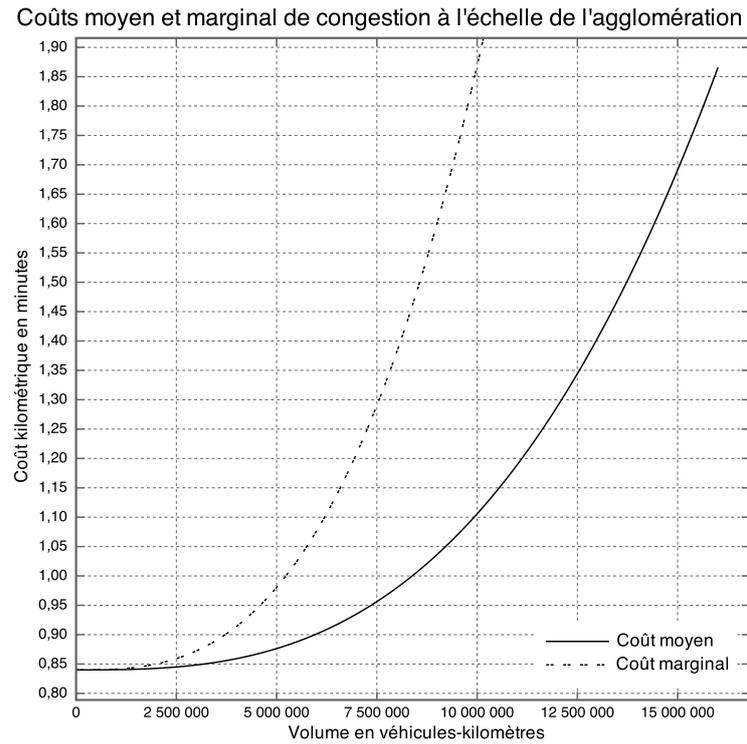


FIGURE 4.21.: Coûts moyen et marginal de congestion en Île-de-France. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

La Table 4.10 présente une évaluation, basée sur le modèle que nous avons développé, du coût horaire de la congestion¹⁰, pour la région dans son ensemble, pour les trois périodes temporelles considérées. Ces coûts sont donnés en termes temporels, autrement dit en unité d'heures pour une heure de la période considérée, ainsi qu'en euros, en supposant une valeur du temps de transport de 16 €₂₀₀₈/h.

| | HPM | HC | HPS |
|------------------------------------------|--------|-------|--------|
| Coût de congestion (h) | 99 400 | 4 500 | 56 000 |
| Coût de congestion (k€ ₂₀₀₈) | 1 590 | 72 | 896 |

TABLE 4.10.: Borne supérieure du coût horaire de la congestion pour l'agglomération parisienne. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

Nous avons employé une méthode d'expansion temporelle simple permettant de passer dans un premier temps de la segmentation en trois périodes à une valeur journalière, puis à une année : un jour standard = 2 heures de pointe du matin + 5 heures creuses + 3 heures de pointe du soir ; une année = 280 jours standards. Il est clair que cette méthode néglige certaines pointes saisonnières de trafic, et réduit fortement la complexité réelle des aspects temporels de la congestion. Toutefois, ce type d'expansion est couramment pratiqué, pour des raisons de disponibilité des données, et donc de reproductibilité. Avec cette méthode, on obtient une estimation de la borne supérieure du coût annuel, en termes de temps perdu, de la congestion en Île-de-France :

$$C_{\text{congestion}} = 1,74 \text{ G€}_{2008}$$

Rapportée au PIB de la région, qui s'élevait, en 2008, à 552,7 G€ (INSEE), cette valeur représente 0,3 % du PIB. Nous confirmons donc par notre analyse les résultats obtenus, par des méthodes différentes, par Prud'homme et Sun (1999, 2000) et résumés dans Prud'homme (1999a), qui donnent un coût horaire de la congestion de l'ordre de 0,2 % du PIB.

4.7.2. Taux de charge

L'analyse agrégée des taux de charge fournit les résultats présentés dans la Table 4.11.

On retrouve des résultats assez semblables à ceux obtenus pour la moyenne des taux de charge (4.5.2.2). Ils diffèrent néanmoins, puisqu'ils n'ont pas été obtenus

10. Nous avons fait l'hypothèse d'une demande parfaitement élastique, ce qui a tendance à sur-estimer le coût de congestion, mais dans le même temps, compte tenu de la non linéarité du phénomène de congestion, la méthode d'agrégation retenue peut sous-estimer le coût de congestion. Le résultat fournit néanmoins une approximation valable du coût de congestion.

| | HPM | HC | HPS |
|----------------|------|------|------|
| Taux de charge | 0,44 | 0,20 | 0,38 |

TABLE 4.11.: Taux de charge pour l'agglomération. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

de la même manière, et que :

$$\frac{\sum_a L_a \frac{x_a}{\kappa_a}}{\sum_a L_a} \neq \frac{\sum_a L_a x_a}{\sum_a L_a \kappa_a}$$

L'intérêt de ces résultats est de mettre en évidence la très forte concentration spatiale de la congestion. En effet, les taux de charge constatés aux heures de pointe demeurent, au niveau de l'agglomération, relativement faibles, alors même que le temps de parcours moyen y est assez élevé. C'est la conséquence du fait que la répartition des véhicules-kilomètres n'est pas du tout homogène dans l'espace, et que le temps de parcours est une fonction non linéaire du taux de charge.

4.8. Conclusion

Les différentes analyses menées, de natures statistique ou cartographique, convergent toutes pour diagnostiquer une variabilité, tant temporelle que spatiale, très importante de la congestion en Île-de-France. Les postures que nous avons adoptées, et pour lesquelles nous avons développé des indicateurs spécifiques, ont permis de constater cette variabilité, en particulier spatiale, qui bien souvent rend l'interprétation difficile.

Toutefois, nous avons également pu mettre en évidence l'émergence de structures spatiales plus nettes en utilisant une méthode d'agrégation par zones d'origine ou de destination. Autrement dit, la prise en compte de la structure spatiale des déplacements redonne une intelligibilité aux manifestations spatiales de la congestion. Par ailleurs, l'agrégation à l'échelle régionale nous a permis d'évaluer un coût économique de la congestion en Île-de-France, et a également mis en évidence un effet réseau qui atténue l'intensité de la congestion lorsque la demande augmente.

Il demeure que les manifestations spatiales de la congestion restent un phénomène complexe, délicat à analyser et à appréhender. Néanmoins, l'étude présentée dans ce chapitre, en plus d'en fournir une analyse relativement détaillée, amène à suggérer quelques pistes pour en améliorer notre compréhension. Ainsi, l'analyse agrégée par zone souligne l'importance de comprendre les causes de la variabilité spatiale de la congestion : c'est la répartition spatiale des activités et des ménages, et plus précisément leurs répartitions conjointes, qui détermine en grande partie la

demande de transport, et donc la congestion des réseaux. Autrement dit, c'est en comprenant mieux comment les ménages choisissent leur localisation, compte tenu de la localisation des activités (et inversement) que nous pourrons mieux appréhender la question de la congestion et de sa répartition spatiale. Ce sera l'objet principal des chapitres 6 à 8 de cette thèse.

Annexe 4.A Annexes

4.A.1 Tableaux détaillés des résultats par segment

Les Tables 4.12 à 4.15 présentent les résultats détaillés des moyennes par segment des différents indicateurs. La dispersion relative autour de cette moyenne figure entre parenthèses.

| HPM | Rapides | Artérielles | Desserte | Total |
|----------------|-------------|-------------|-------------|-------------|
| Zone centrale | 1,48 (0,71) | 1,02 (1,53) | 0,93 (1,62) | 1,17 (1,18) |
| Proche périph. | 1,25 (1,55) | 1,65 (1,87) | 1,23 (2,91) | 1,35 (2,08) |
| Grande périph. | 0,59 (2,19) | 0,59 (2,67) | 0,69 (2,70) | 0,62 (2,54) |
| Total | 0,86 (1,79) | 0,85 (2,44) | 0,85 (2,78) | 0,85 (2,31) |
| HC | | | | |
| Zone centrale | 0,20 (1,03) | 0,15 (1,34) | 0,08 (2,41) | 0,15 (1,42) |
| Proche périph. | 0,04 (2,17) | 0,11 (1,82) | 0,09 (2,51) | 0,07 (2,39) |
| Grande périph. | 0,01 (3,55) | 0,04 (2,65) | 0,09 (3,84) | 0,04 (4,72) |
| Total | 0,04 (2,57) | 0,06 (2,35) | 0,09 (3,42) | 0,06 (3,24) |
| HPS | | | | |
| Zone centrale | 1,38 (0,76) | 0,81 (1,48) | 0,75 (1,60) | 1,01 (1,18) |
| Proche périph. | 0,46 (1,21) | 0,76 (1,43) | 0,75 (2,05) | 0,62 (1,73) |
| Grande périph. | 0,21 (2,60) | 0,34 (2,31) | 0,57 (2,39) | 0,36 (2,62) |
| Total | 0,39 (1,79) | 0,46 (1,98) | 0,64 (2,16) | 0,49 (2,09) |

TABLE 4.12.: Temps de parcours relatif moyen par segment, dispersion relative entre parenthèses. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

| HPM | Rapides | Artérielles | Desserte | Total |
|----------------|-------------|-------------|-------------|-------------|
| Zone centrale | 0,90 (0,35) | 0,53 (0,47) | 0,47 (0,54) | 0,52 (0,56) |
| Proche périph. | 0,84 (0,35) | 0,61 (0,42) | 0,49 (0,52) | 0,56 (0,51) |
| Grande périph. | 0,52 (0,57) | 0,29 (0,82) | 0,34 (0,80) | 0,34 (0,80) |
| Total | 0,60 (0,55) | 0,34 (0,79) | 0,38 (0,73) | 0,39 (0,75) |
| HC | | | | |
| Zone centrale | 0,58 (0,39) | 0,21 (1,02) | 0,21 (0,82) | 0,24 (0,88) |
| Proche périph. | 0,40 (0,37) | 0,30 (0,56) | 0,19 (0,86) | 0,24 (0,74) |
| Grande périph. | 0,24 (0,56) | 0,12 (1,03) | 0,14 (1,12) | 0,14 (1,03) |
| Total | 0,29 (0,59) | 0,15 (1,00) | 0,15 (1,04) | 0,17 (0,99) |
| HPS | | | | |
| Zone centrale | 0,89 (0,35) | 0,52 (0,46) | 0,47 (0,51) | 0,52 (0,53) |
| Proche périph. | 0,69 (0,35) | 0,52 (0,44) | 0,43 (0,58) | 0,48 (0,54) |
| Grande périph. | 0,45 (0,54) | 0,24 (0,95) | 0,29 (0,90) | 0,29 (0,89) |
| Total | 0,52 (0,53) | 0,28 (0,89) | 0,33 (0,81) | 0,33 (0,81) |

TABLE 4.13.: Taux de charge moyen par segment, dispersion relative entre parenthèses. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

| HPM | Rapides | Artérielles | Desserte | Total |
|----------------|-------------|-------------|-------------|-------------|
| Zone centrale | 5,25 (0,88) | 6,07 (1,57) | 9,01 (1,78) | 7,08 (1,69) |
| Proche périph. | 3,79 (1,86) | 8,83 (1,88) | 8,91 (3,29) | 6,54 (2,83) |
| Grande périph. | 1,64 (2,35) | 2,51 (3,22) | 4,15 (3,03) | 2,65 (3,24) |
| Total | 2,58 (2,02) | 4,08 (2,69) | 5,95 (3,13) | 4,04 (3,06) |
| HC | | | | |
| Zone centrale | 0,71 (1,28) | 0,79 (1,38) | 0,77 (2,40) | 0,75 (1,85) |
| Proche périph. | 0,13 (3,29) | 0,57 (1,97) | 0,67 (2,79) | 0,39 (3,03) |
| Grande périph. | 0,04 (5,59) | 0,15 (3,08) | 0,54 (4,14) | 0,21 (5,73) |
| Total | 0,14 (3,28) | 0,28 (2,66) | 0,61 (3,49) | 0,31 (3,97) |
| HPS | | | | |
| Zone centrale | 4,85 (0,84) | 4,93 (1,59) | 7,20 (1,68) | 5,95 (1,54) |
| Proche périph. | 1,42 (1,51) | 4,09 (1,52) | 5,48 (2,23) | 3,29 (2,33) |
| Grande périph. | 0,64 (3,07) | 1,49 (2,81) | 3,25 (2,84) | 1,66 (3,45) |
| Total | 1,26 (2,06) | 2,26 (2,26) | 4,39 (2,42) | 2,51 (2,72) |

TABLE 4.14.: Coût externe marginal moyen par segment en min/véh.km, dispersion relative entre parenthèses. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

| HPM | Rapides | Artérielles | Desserte | Total |
|----------------|-------------|-------------|--------------|-------------|
| Zone centrale | 2,67 (0,32) | 1,70 (0,77) | 1,65 (0,78) | 2,07 (0,60) |
| Proche périph. | 2,13 (0,55) | 2,08 (0,65) | 1,70 (0,78) | 2,00 (0,64) |
| Grande périph. | 1,06 (1,14) | 1,03 (1,15) | 1,21 (1,839) | 1,10 (1,41) |
| Total | 1,51 (0,87) | 1,29 (1,01) | 1,39 (0,92) | 1,41 (0,92) |
| HC | | | | |
| Zone centrale | 0,75 (0,66) | 0,55 (1,08) | 0,30 (1,60) | 0,55 (0,99) |
| Proche périph. | 0,16 (1,45) | 0,39 (1,35) | 0,33 (1,68) | 0,27 (1,63) |
| Grande périph. | 0,06 (2,56) | 0,15 (2,10) | 0,28 (1,98) | 0,15 (2,47) |
| Total | 0,17 (1,92) | 0,22 (1,82) | 0,30 (1,85) | 0,22 (1,93) |
| HPS | | | | |
| Zone centrale | 2,52 (0,37) | 1,55 (0,79) | 1,55 (0,76) | 1,93 (0,61) |
| Proche périph. | 1,27 (0,68) | 1,56 (0,74) | 1,39 (0,87) | 1,38 (0,77) |
| Grande périph. | 0,59 (1,30) | 0,82 (1,21) | 1,08 (1,08) | 0,80 (1,23) |
| Total | 0,96 (1,04) | 1,03 (1,06) | 1,22 (0,97) | 1,06 (1,03) |

TABLE 4.15.: Elasticité temps-volume moyenne par segment, dispersion relative entre parenthèses. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

4.A.2 Diagrammes Boxplot complémentaires par segment

Les Figures 4.22 à 4.24 fournissent les diagrammes Boxplot pour les périodes temporelles non présentées dans le corps du texte.

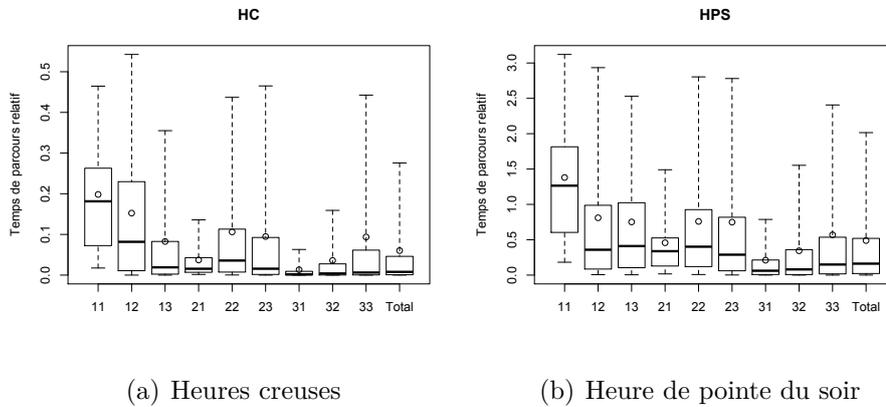


FIGURE 4.22.: Diagramme Boxplot C_5/C_{95} des distributions des temps de parcours relatifs par segment. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

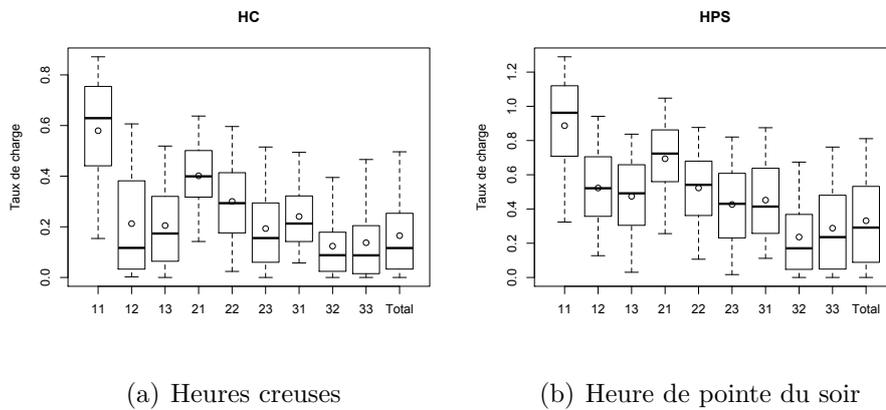


FIGURE 4.23.: Diagramme Boxplot C_5/C_{95} des distributions des taux de charge par segment. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

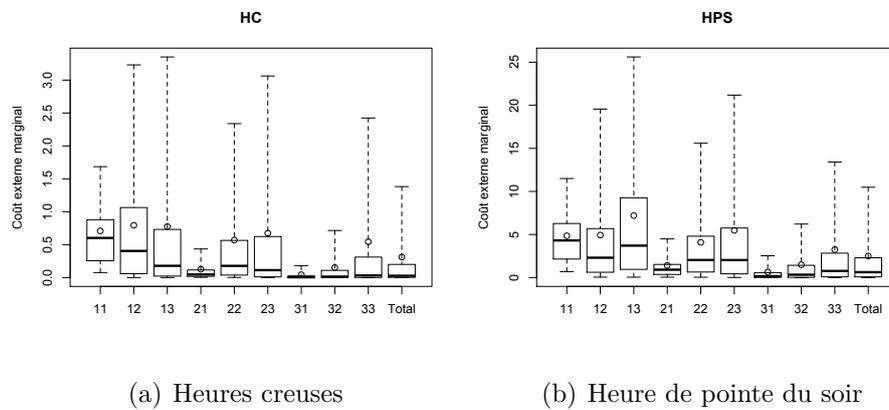


FIGURE 4.24.: Diagramme Boxplot C_5/C_{95} des distributions des coûts externes marginaux par segment. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

5

Coût social marginal et aide à la décision : la décongestion routière dans l'évaluation d'un projet de transport

5.1. Introduction

Comme le rappellent Maurice et Crozet (2007), l'évaluation, et donc le calcul économique sont essentiels dans les projets de transport, et ce pour deux raisons principales : d'une part, les investissements que ce secteur suppose sont généralement massifs, et principalement financés sur fonds publics ; d'autre part, les infrastructures réalisées marquent durablement le territoire et ont donc un caractère irréversible. L'utilisation du calcul économique relève donc d'une démarche démocratique de transparence des choix collectifs et de rationalité. Cela ne l'empêche pas de se heurter néanmoins à de nombreuses réticences de la part du public ou des politiques (Baumstark, 2007), qui lui reprochent son caractère technocratique, approximatif, et soulignent les difficultés qu'il rencontre pour prendre en compte les questions d'équité et de redistribution.

En France, le calcul économique public, utilisé justement pour l'évaluation des projets de transport, en particulier lorsqu'ils impliquent une construction d'infrastructure, repose sur deux bases principales. Il fait tout d'abord appel à la théorie

économique standard, et en particulier la maximisation de la valeur actualisée nette telle que présentée par Quinet (1998, chap. 7) ou Gaudry (2007). Ensuite, il emploie un ensemble de valeurs *tutélaires*, servant de références communes à l'ensemble des évaluations. Ces valeurs sont employées pour valoriser c'est-à-dire *monétariser*, les effets, positifs comme négatifs, d'un projet de transport. Ainsi, les gains de temps liés à une infrastructure de transport peuvent être évalués par l'intermédiaire de valeurs tutélaires des gains de temps, différentes selon le motif du déplacement (Boiteux et Baumstark, 2001).

En raison de l'importance de ces valeurs tutélaires, tant par la sensibilité des résultats des évaluations à ces dernières (Chevasson, 2007 ; Grangeon, 2010) que par leur utilisation pour l'ensemble des évaluations économiques, il est essentiel de disposer de valeurs tutélaires cohérentes entre elles et suffisamment proches de la réalité, tout en maintenant un niveau de complexité des méthodologies compatible avec leur utilisation courante dans les évaluations.

Parmi elles, les valeurs tutélaires se rapportant aux gains de temps, qu'il s'agisse de l'évaluation de ces gains de temps ou de leur monétarisation, sont particulièrement importantes, puisque les gains de temps représentent une part très significative des avantages attendus d'un projet de transport (Hautreux, 1969 ; Boiteux et Baumstark, 2001 ; Chevasson, 2007). Ainsi, Chevasson (2007) trouve-t-il, sur un projet d'infrastructure fictif stylisé, une élasticité entre valeur du temps et bénéfice actualisé de l'ordre de 1. De même, Quinet (1998) précise que si une incertitude sur les valeurs du temps peut n'avoir qu'une influence limitée sur les résultats des modèles de trafic, elle risque d'avoir des répercussions importantes sur le surplus. L'accent est donc très souvent mis sur les questions d'évaluation des valeurs du temps, par des méthodes de préférences déclarées ou révélées, et sur la prise en compte de la distribution des valeurs du temps dans les modèles de trafic.

Pourtant, tout aussi importante est la question de l'évaluation des gains de temps en termes physiques, avant leur monétarisation. Si en l'absence, ou quasi-absence, de congestion, il est relativement aisé d'évaluer ces gains de temps, même sans un modèle de trafic performant, il n'en va pas de même lorsque le niveau de congestion est important, comme c'est généralement le cas en milieu urbain. En effet, la congestion routière constitue, comme nous l'avons vu au chapitre 2, un coût externe temporel *de club*, c'est-à-dire que chaque usager impose un coût temporel à l'ensemble des autres usagers *de la route*. C'est pour améliorer l'évaluation en milieu urbain que les rapports Hautreux (1969) et Boiteux et Baumstark (2001) insistent sur la nécessité d'évaluer le plus précisément possible les gains de temps liés à la décongestion d'une infrastructure routière. Ces deux rapports soulignent également que cette évaluation s'est faite jusqu'à présent sur des bases peu solides. Ainsi, Hautreux (1969), qui fournit un ensemble de valeurs issues de mesures ponctuelles

sur des voies parisiennes, précise que ces valeurs n'ont pas vocation à être utilisées directement dans le cadre de l'évaluation des projets. Pourtant, comme le rappellent Boiteux et Baumstark (2001) et Leurent *et al.* (2009), un coefficient « *Hautreux* » a été systématiquement employé depuis, modulé sur la base de l'expertise des chargés d'études, pour tenir compte des particularités des projets évalués.

C'est donc dans ce cadre d'amélioration de la connaissance des gains marginaux de décongestion que l'étude présentée dans ce chapitre a été menée. Elle s'inscrit donc dans la tradition du calcul économique comme aide à la décision. Commanditée par le Ministère chargé des Transports, elle a fait intervenir un ensemble d'acteurs, majoritairement franciliens, autour de la question de la mise à jour des valeurs tutélaires des gains de décongestion routière. Les résultats de cette étude (pour le rapport final, voir Leurent *et al.*, 2009) ont depuis été intégrés dans une étude d'évaluation figurant au rapport annuel 2008 de la Commission des comptes des transports de la Nation (CCTN, 2009) et sont couramment employés pour des travaux internes au Ministère.

En dehors de l'objectif premier opérationnel de cette étude, à savoir actualiser les valeurs tutélaires de décongestion routière, trois objectifs de recherche s'en dégagent. Elle vise ainsi à constater la variabilité spatiale et temporelle de la congestion routière en milieu urbain, et de dégager les facteurs à l'origine de cette variabilité. Ensuite, il s'agit de quantifier les valeurs de ce coût externe de congestion. Enfin, un dernier objectif consiste à investiguer les modalités et les conséquences d'une évaluation agrégée à différentes échelles, en vue de simplifier l'application concrète de la méthode d'évaluation dans un contexte opérationnel. Ce dernier objectif interroge également la possibilité d'employer une valeur unique moyenne. Parmi ces trois objectifs, le chapitre précédent a apporté des éléments réponses concernant les deux premiers. Ce chapitre est donc consacré au dernier objectif.

La seconde section de ce chapitre présente le cadre économique dans lequel s'inscrit notre démarche d'évaluation des coûts marginaux de décongestion. Plus spécifiquement, on se placera dans le cas particulier d'une nouvelle infrastructure de transport collectif, en concurrence avec une infrastructure routière. La troisième section propose un état de l'art de l'évaluation des gains de temps en milieu urbain, ainsi qu'un historique de ces méthodes en France. Ensuite, la quatrième section présente, sur la base de nos analyses, plusieurs méthodes d'évaluation des gains de temps liés à la décongestion. La cinquième section étudie un cas réel, et compare, sur ce cas, les différentes méthodes proposées.

5.2. Concurrence entre modes de transport : congestion de la route et qualité de service

Lorsqu'une relation géographique entre deux lieux est desservie par deux modes de transports concurrents, par exemple la voiture particulière et un moyen de transport collectif, toute modification de l'offre de l'un des deux modes (amélioration de la qualité de service en transport collectif, augmentation de la capacité de l'infrastructure routière, etc.) entraîne une réaction des usagers de la liaison qui constituent la demande : ils peuvent adapter leur itinéraire modal ou changer de mode de transport. Cette réaction de la demande modifie la charge d'utilisation des modes, en particulier la charge supportée par le mode routier, et donc sur le coût local de la congestion.

5.2.1. La répartition modale

Puisque l'amélioration de la qualité de service d'un mode de transport liée à un projet d'infrastructure modifie la répartition modale des usagers entre les différents modes, l'évaluation économique d'un tel projet doit s'appuyer sur une modélisation fine du choix du mode par les usagers.

Parmi les caractéristiques attachées à un mode de transport, la durée de déplacement par ce mode figure en bonne place. De même, les durées de déplacement par les autres modes influent également sur la part modale d'un mode. Autrement dit, l'évaluation des temps de parcours par les différents modes, ainsi que la modification de ces temps de parcours, par un projet d'infrastructure, par exemple, ont une influence directe sur les choix des usagers, et partant, sur l'évaluation économique d'un projet. Or, les temps de parcours, et plus généralement les itinéraires suivis et leurs caractéristiques, sont directement liés aux trafics empruntant les différents modes. En termes de modélisation, cela signifie qu'il existe un bouclage entre l'étape de choix modal et l'étape d'affectation. Si l'évaluation du report modal est cruciale pour l'évaluation des projets, car ce report détermine la part modale, c'est-à-dire la part de marché du mode concerné par le projet d'amélioration, c'est par le bouclage entre affectation et choix modal qu'il est possible d'évaluer les reports entre modes de transport.

Ce bouclage, qui est au cœur du mécanisme de modélisation de la concurrence entre modes de transport, est illustré sur la Figure 5.1, dans le cas classique d'une concurrence entre le mode routier (voiture particulière) et le mode collectif (noté TC sur le schéma).

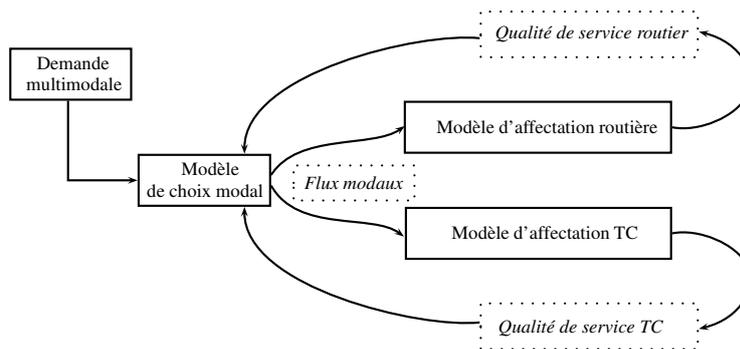


FIGURE 5.1.: Bouclage entre affectation et choix modal. *Source* : Leurent et Breteau (2010)

5.2.2. L'affectation des trafics sur les réseaux

Pour connaître les volumes circulant sur les différents réseaux, il est donc nécessaire de s'intéresser à l'étape d'affectation du trafic. Cette étape peut être considérée de manière indépendante aux deux étapes précédentes (distribution et choix modal), avec la limite que dans ce cas, on considère implicitement que la demande modale de transport (sous la forme d'une matrice O-D) est inélastique. En réalité, le coût de transport influence à la fois le choix du mode et la distribution des flux. C'est pourquoi les modèles procèdent souvent par itérations successives, sans garantie de convergence du fait des risques d'oscillations (Ortúzar et Willumsen, 2001). Des modèles réalisant simultanément les étapes de distribution, choix du mode et affectation ont été proposés (par exemple, Florian et Nguyen, 1978), mais ils sont généralement d'une mise en œuvre pratique difficile. De manière générale, à l'équilibre du système de transport, les flux affectés définissent des caractéristiques de coût de différents modes telles que la demande modale qui émerge est effectivement celle qui a été affectée (Gaudry, 2007, parle d'équilibre demande-performance). La procédure d'affectation sur le réseau routier, et en particulier les questions d'équilibre de l'utilisateur et d'optimum social, a déjà été abordée, du point de vue économique, dans le chapitre 2, et nous n'y reviendrons pas en détail, d'autant que les questions algorithmiques, qui sont au cœur de la procédure d'affectation, ne sont pas l'objet de cette thèse. En revanche, l'évaluation des projets de transport repose sur les résultats d'affectation. Ainsi, si le projet de transport modifie les temps de parcours (ou plus généralement les qualités de service respectives), les répartitions modales des flux d'un certain nombre de liaisons origine-destination vont évoluer, jusqu'à aboutir à un nouvel équilibre des flux sur les réseaux. Le surplus dégagé par le projet, qui constitue une des mesures classiques de la rentabilité d'un projet de transport, se déduit alors de la différence entre les situations d'équilibre hors et avec projet.

Notons à ce propos que plusieurs auteurs (en particulier Jara-Diaz, 1986) ont montré qu'en situation de concurrence parfaite, et en l'absence d'externalités sur les différents marchés pertinents (marché du travail, marché immobilier, etc.) aux origines et destinations de tous les agents, le surplus mesuré *sur le réseau* est égal aux surplus mesurés aux origines et destinations. Plus précisément, en reprenant ses notations, Jara-Diaz (1986) suppose que la demande de transport entre deux zones provient de l'environnement économique, représenté par un marché dans chaque zone. Chacun de ces marchés est caractérisé par une fonction de demande $D_i(p_i)$ et une fonction d'offre $S_i(p_i)$, et il s'établit sur chaque marché un prix d'équilibre, respectivement p_1^0 et p_2^0 tels que $p_1^0 > p_2^0$. Pour toute valeur de $p_1 < p_1^0$, une demande excédentaire $ED_1(p_1) = D_1(p_1) - S_1(p_1)$ apparaît sur le marché 1, tandis que pour toute valeur $p_2 > p_2^0$ apparaît une offre excédentaire $ES_2(p_2) = S_2(p_2) - D_2(p_2)$. Si une quantité Q est produite de manière excédentaire sur le marché 2 et vendue sur le marché 1, on a $Q = ES_2(p_2) = ED_1(p_1)$ et les prix d'équilibre (avec transport entre les deux zones) sont tels que $p_1^* < p_1^0$ et $p_2^* > p_2^0$, comme le montre la partie supérieure de la Figure 5.2(a). La disposition à payer pour consommer sur le marché

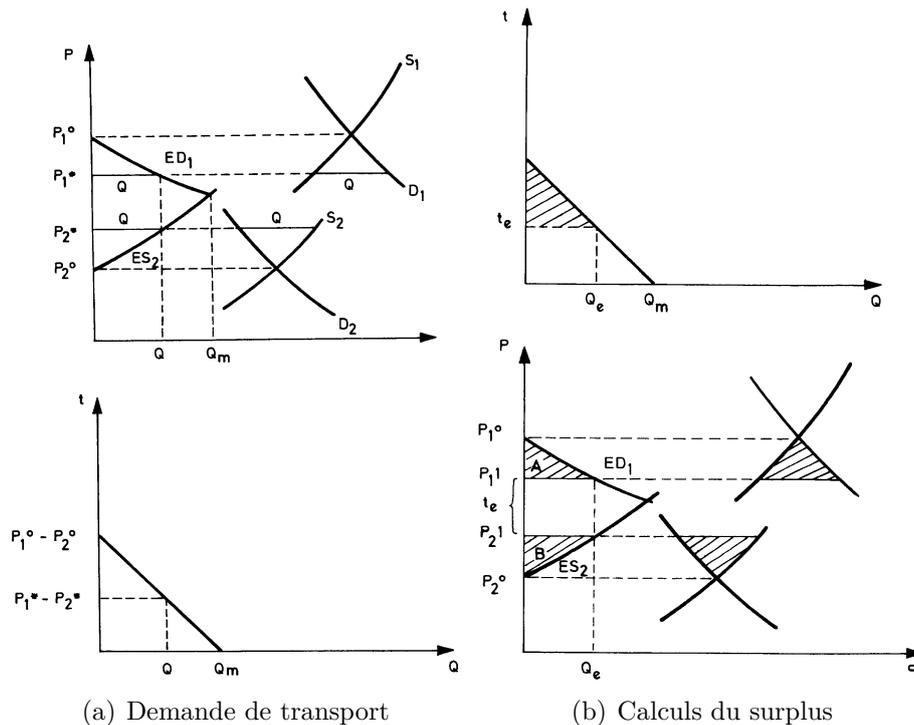


FIGURE 5.2.: Mesure du surplus sur le réseau et sur les marchés locaux. *Source* : Jara-Diaz (1986)

1 une quantité Q produite sur le marché 2 est p_1^* (ou p_2^* plus le coût de transport) ; autrement dit, la disposition à payer pour transporter une quantité Q du marché 2

vers le marché 1 est :

$$t = p_1^* - p_2^* = ED_1^{-1}(Q) - ES_2^{-1}(Q), \quad (5.1)$$

comme le représente la partie inférieure de la Figure 5.2(a). À partir de cette représentation, il découle directement que le surplus calculé sur le réseau correspond exactement à la somme des surplus sur les deux marchés locaux, comme l'illustre la Figure 5.2(b).

Autrement dit, dans ce cas, l'analyse du surplus sur le réseau suffit à mesurer l'ensemble des avantages tirés d'une nouvelle infrastructure. Lorsque les conditions de concurrence sont différentes¹, cette égalité n'est pas vérifiée, mais un certain recouvrement a néanmoins lieu. Par conséquent, le risque de double compte est important si l'on se contente de sommer le surplus mesuré sur le réseau (à travers la valorisation des gains de temps, en particulier) et le surplus mesuré au niveau des autres marchés urbains. Cela souligne l'importance d'une évaluation rigoureuse du surplus généré directement sur les réseaux, prenant en compte l'ensemble des éléments.

Comme nous l'avons déjà signalé, les gains de temps par les usagers représentent une part majeure de l'amélioration du surplus engendrée par une nouvelle infrastructure de transport. Or, trois types d'usagers peuvent être distingués quant à la répartition de ces gains de temps :

- les usagers qui vont emprunter la nouvelle infrastructure ;
- les usagers qui ne se déplaçaient pas (trafic induit) ;
- les usagers qui ne changent pas d'infrastructure.

Parmi ces trois types d'usagers obtenant des gains de temps, ceux des premiers sont à la fois assez intuitifs et relativement aisés à déterminer. Ceux des seconds, en revanche, sont plus délicats à évaluer, car la disposition à payer de ces usagers pour le service de transport n'est pas clairement déterminée. Enfin, ceux des troisièmes sont moins intuitifs : ils correspondent aux gains de temps liés à la décongestion de l'infrastructure existante. En pratique, compte tenu de la modélisation actuelle des transports collectifs, ces gains de temps ne concernent que les usagers de la route, et représentent généralement une part importante des avantages attendus d'un projet de transport. En d'autres mots, dans le cas d'un projet de mise en place d'une nouvelle infrastructure de transport collectif, une part importante des gains de temps est due non pas aux usagers qui empruntent le nouveau mode collectif mis en place, mais au temps gagné par les usagers qui continuent à emprunter la voiture particulière.

1. Typiquement, si les marchés aux origines et destinations sont de nature monopolistique ou oligopolistique.

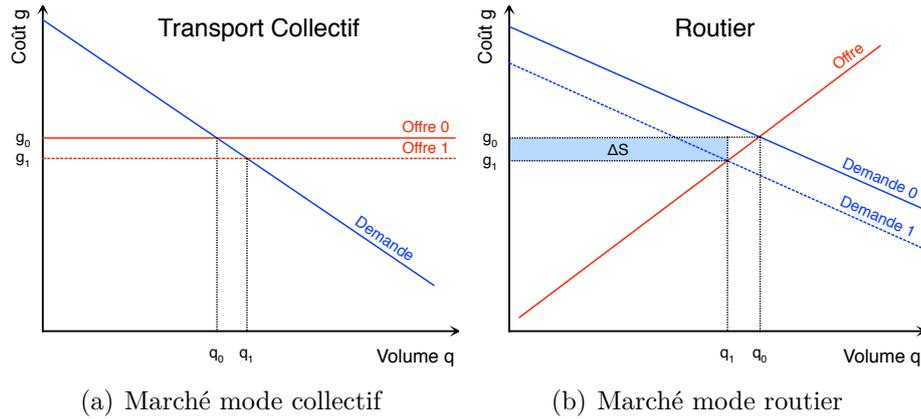


FIGURE 5.3.: Deux marchés de transport en concurrence. *Source* : Leurent et Breteau (2010)

La Figure 5.3, tirée de Leurent et Breteau (2010), illustre ce mécanisme faisant apparaître deux sources de gains de temps, dans le cas d'une concurrence entre le mode routier et un mode collectif qui bénéficie d'une amélioration. Le comportement des usagers y est modélisé par une courbe de demande qui relie le nombre de déplacements utilisant l'un ou l'autre mode, au coût généralisé d'usage de ce mode. La fonction de demande est exogène au marché modal, tandis que les fonctions d'offre diffèrent selon le mode considéré. On suppose que la fonction d'offre en transport collectif est horizontale, autrement dit que le coût du déplacement ne dépend pas du volume de déplacements; cela signifie qu'il n'y a pas de congestion pour le mode collectif. En revanche, pour le mode routier, on suppose que la congestion entraîne une fonction d'offre croissante.

En conséquence, sur le marché du mode collectif, l'amélioration de la qualité de service entraîne une baisse du coût du déplacement, poussant à la hausse le nombre de déplacements supportés par le mode collectif, qui passent donc de q_0^c à q_1^c . En supposant que la demande totale de déplacements reste inchangée (pas d'induction), cette augmentation de demande pour le mode collectif entraîne une chute de demande pour le mode routier jusqu'à ce que le coût de déplacement s'équilibre entre les deux modes. Cela se traduit par une chute du nombre de déplacements supportés par le mode routier, qui passe de q_0^r à q_1^r .

Ces usagers, qui restent sur la route une fois réalisé le projet, effectuent un nombre de déplacements donné par $q_1^r = q_0^r - \Delta q^r$. Chacun des déplacements routiers réalisés s'effectue au coût $g_1^r = g_0^r - \Delta g^r$. Leur gain net total en terme de surplus s'écrit donc :

$$\Delta S^r = q_1^r \Delta g^r = q_0^r \Delta g^r - \Delta q^r \Delta g^r \quad (5.2)$$

En supposant que l'estimation du volume de déplacements reportés vers le mode

collectif soit satisfaisante, l'évaluation du surplus repose sur une détermination précise de Δg^r , autrement dit de la modification du coût unitaire moyen d'un déplacement par le mode routier.

On saisit donc l'importance de l'étape d'affectation du trafic dans la démarche d'évaluation économique d'un projet de transport, puisque c'est principalement à travers elle que les gains ou pertes de temps vont être mis en évidence, directement en association avec un modèle de choix modal, et indirectement du fait de la décongestion. Pourtant, de manière à disposer d'une évaluation rapide des gains de temps liés à la décongestion routière, ou en raison de la difficulté, pour un certain nombre d'acteurs de l'évaluation des projets, de disposer de modèles d'affectation du trafic routier finement calé, plusieurs méthodes approchées d'évaluation des gains de temps de décongestion ont été développées et employées. La section suivante en présente une revue internationale.

5.3. L'évaluation des gains de décongestion routière

Nous venons de le voir, une partie importante des gains de temps apportés par un projet d'infrastructure de transport, collectif par exemple, est liée à la décongestion de la voirie routière, dont bénéficient les usagers qui continuent à utiliser le mode routier. Mesurer ces gains de temps est donc essentiel dans le cadre de l'évaluation économique d'un projet de transport. Pourtant, la méthodologie proposée par l'organisme américain chargé des transports collectifs en matière d'évaluation des projets ne tient apparemment pas compte de ces gains de temps liés à la décongestion (FTA, 2000, 2009), alors même qu'un certain nombre de modèles appropriés ont été développés (DeCorla-Souza *et al.*, 1998).

Il apparaît nécessaire de distinguer plusieurs cas de figures, ainsi que différentes *postures techniques*. Plus précisément, deux cas de figures peuvent se présenter. Si les changements en termes de demande liés à la présence du projet sont petits en comparaison de la demande totale, autrement dit, avec les notations précédentes, si $\Delta q^r \ll q_0^r$, on parle alors de changements *marginaux*, et l'évaluation prend place dans le cadre des gains de temps marginaux. Dans ce cas, en supposant que la fonction d'offre $G(q^r)$ du mode routier est inchangée², on a $\Delta g^r = G(q_1^r) - G(q_0^r)$, ou encore :

$$\Delta g^r = G'(q_0^r)\delta q^r + o(1) \quad (5.3)$$

où δq^r désigne un changement marginal de demande. L'équation (5.2) se réécrit

2. C'est le cas des fonctions de temps de parcours étudiées au chapitre précédent, pourvu que le projet de transport ne modifie pas la capacité des voies.

alors :

$$\Delta S^r = \underbrace{[q_0^r G'(q_0^r)]}_{\chi_0^r} \delta q^r + o(\delta q^r) \quad (5.4)$$

où χ_0^r est le coût externe social marginal de congestion routière.

En revanche, lorsque les changements ne sont pas marginaux, les méthodes d'évaluation à employer diffèrent. Plus précisément, la linéarisation effectuée à l'équation (5.3) n'est plus valable, et une différence entre les résultats d'affectation à l'équilibre de l'utilisateur en situation avec et hors projet devient nécessaire.

Par ailleurs, deux postures techniques peuvent être envisagées (Leurent *et al.*, 2009 ; Leurent et Breteau, 2010) : celle de l'*ingénieur-conseil*, qui réalise une évaluation économique complète sur la base d'une étude de trafic ; celle de l'*inspecteur*, qui doit porter rapidement une appréciation critique sur une étude d'évaluation, sur la base d'un nombre limité d'indicateurs.

5.3.1. La mesure des gains de temps pour l'ingénieur-conseil

L'ingénieur-conseil se doit, lorsqu'il réalise une étude d'évaluation des gains de temps liés à la décongestion routière, d'appliquer les méthodologies correspondant à l'état de l'art de ce type d'études. Or, comme précisé par Leurent *et al.* (2009), la méthode représentative de la pratique courante au niveau international est celle de la différence entre deux résultats d'affectation à l'équilibre de l'utilisateur. Comme nous l'avons mentionné, lorsque les changements sont non marginaux, c'est d'ailleurs la seule méthode acceptable, car elle permet de rendre compte de la non-linéarité des phénomènes liés à la congestion.

Quoi qu'il en soit, dans tous les cas, les différentes études portant sur le sujet recommandent de désagréger l'étude selon l'espace et la période temporelle (Heatco, 2006 ; Lindsay *et al.*, 2007). C'est également la conclusion qui peut être tirée de nos propres travaux, tels que présentés dans le chapitre précédent ainsi que dans Leurent *et al.* (2009) et Leurent et Breteau (2009). En effet, la variabilité tant spatiale que temporelle de la congestion rend nécessaire une telle désagrégation.

Si la méthode non marginaliste par différence d'affectation semble la mieux adaptée aux besoins de l'ingénieur-conseil, une méthodologie marginaliste est envisageable, à condition de respecter un certain niveau de désagrégation spatiale et temporelle du réseau. Ainsi, lorsque le modèle d'affectation routière ne permet pas de tenir explicitement compte des contraintes liées à la capacité, autrement dit de la congestion, une méthode marginaliste forfaitaire peut être utile. Il en va de même si le réseau n'est pas modélisé de manière suffisamment fine. Quoi qu'il en soit, la posture de l'ingénieur-conseil est clairement celle à privilégier dans tous les cas où les modifications de trafic induites par un projet de transport ont une forte probabilité d'être non marginales.

5.3.2. L'estimation des gains de temps pour l'inspecteur

La posture de l'inspecteur exige un niveau de détail fortement réduit par rapport à celle de l'ingénieur-conseil. En effet, l'objectif est pour lui de vérifier le caractère suffisamment réaliste de la modélisation et des résultats, ainsi que de s'assurer de l'adéquation de la méthode employée et de l'évaluation à mener. Son rôle consiste donc principalement à effectuer un audit externe du travail réalisé par l'ingénieur-conseil, en employant des méthodes simplifiées et rapides d'évaluation des gains de temps liés à la décongestion.

Les valeurs de référence fournies dans Maibach *et al.* (2008) se placent dans ce cadre de la posture de l'inspecteur. Ainsi, les auteurs de ce document fournissent-ils une table de valeurs de référence pour l'évaluation des gains de temps liés à la décongestion. La Table 5.1 présente un extrait des résultats de ce rapport. Les valeurs concernent les agglomérations de plus de 2 millions d'habitants et sont exprimées en euros par véhicule-kilomètre.

| Type routier | Min | Médian | Max |
|-----------------------------|------|--------|------|
| Voies rapides urbaines | 0,30 | 0,50 | 0,90 |
| Voies collectrices urbaines | 0,20 | 0,50 | 1,20 |
| Voies de desserte centrales | 1,50 | 2,00 | 3,00 |

TABLE 5.1.: Coût externe marginal de congestion dans le rapport Maibach *et al.* (2008), €/véh.km.

Néanmoins, il nous semble que fournir ces valeurs en termes économiques est trompeur, et rend les résultats non comparables directement, car ils dépendent d'une valeur monétaire du temps. Or, les auteurs eux-mêmes précisent que cette valeur tutélaire du temps doit être choisie par la collectivité, et qu'elle ne correspond pas nécessairement à une valeur du temps comportementale.

C'est pourquoi nous avons, pour notre part, préféré mettre au cœur de notre méthodologie, que nous présenterons dans la section suivante, des évaluations de ces gains de temps en termes physiques, c'est-à-dire en unités de temps. La collectivité est ensuite libre de valoriser ces gains de temps en unités monétaires par fixation d'une ou plusieurs valeurs du temps.

Newbery (1990) propose également une table de valeurs, en termes temporels et monétaires, du coût marginal de la congestion, et ce pour différentes catégories de voiries ou d'environnements géographiques. La Table 5.2 en fournit un extrait, converti en minutes par véhicules-kilomètres.

La posture de l'inspecteur est également l'approche souvent retenue en France depuis le rapport Hautreux (1969), et confirmée par les rapports Boiteux *et al.*

| Type routier | Pointe | Hors Pointe |
|---------------|--------|-------------|
| Voies rapides | 0,03 | 0,03 |
| Centre urbain | 3,25 | 2,61 |
| Périphérie | 1,42 | 0,78 |
| Petite ville | 0,62 | 0,38 |

TABLE 5.2.: Coût externe marginal de congestion dans Newbery (1990), min/véh.km.

(1994) et Boiteux et Baumstark (2001), pour évaluer les gains de temps liés à la décongestion routière dans le cadre des projets urbains de transport collectif. Dans le rapport de 1969, c'est l'ingénieur Frébault qui avait été en charge de la partie relative au traitement de la congestion routière en section courante, dans une annexe datée de 1968. Il se plaçait d'emblée dans le cadre marginaliste, et employait donc la formule classique $\chi = x \frac{dt}{dx}$ pour le coût externe marginal de congestion. L'estimation se faisait donc en termes temporels. Par ailleurs, la forme fonctionnelle de la fonction (*cf.* Chapitre 2) de temps de parcours employée était hyperbolique :

$$t(x) = \frac{1}{A - Bx}$$

où A et B sont deux paramètres à estimer. Cette forme fonctionnelle est assez différente de celle, en puissance, de la fonction BPR que nous avons employée, mais est en revanche à rapprocher de la fonction Davidson.

Les valeurs fournies, en heures, dans le rapport, sont présentées dans la Table 5.3, converties en minutes, où V_{max} est la vitesse de référence (limite) sur la route considérée et V la vitesse de circulation effective des véhicules.

| $V \backslash V_{max}$ | 40 km/h | 50 km/h |
|------------------------|---------|---------|
| 15 km/h | 6,7 | 9,3 |
| 20 km/h | 3,0 | 4,5 |
| 30 km/h | 0,7 | 1,3 |
| 40 km/h | 0,0 | 0,4 |

TABLE 5.3.: Coût externe marginal de congestion dans le rapport Hautreux (1969), min/véh.km.

Ces valeurs ont pour origine des mesures du trafic routier urbain dans Paris intra-muros durant les années 1960. Les conditions de trafic sont aujourd'hui bien différentes, ce qui met en doute la réalité des valeurs ci-dessus, et justifie en outre

leur mise à jour pour une utilisation dans le cadre de la posture de l'inspecteur.

5.3.3. La pratique actuelle de la mesure des gains de temps de décongestion

Bien que la posture de l'ingénieur-conseil soit la mieux adaptée dans le cadre des évaluations de projets de transport collectif urbain, c'est une méthode marginaliste relativement grossière qui a été retenue. Ainsi, pour l'évaluation du poste des gains de temps de décongestion dans les études socio-économiques des infrastructures de transport, la postérité a retenu, en France, une valeur de 0,125 h/véh.km (7,5 min/véh.km). Cette valeur est proche des deux plus grandes valeurs de la Table 5.3. De manière plus précise, nous avons pu établir que les chargés d'études modulent généralement cette valeur unique par un coefficient multiplicatif, variant typiquement de 0,2 à 0,5, afin de tenir compte du contexte local spécifique de l'étude, en particulier la densité du trafic et le type de voirie (Leurent *et al.*, 2009). Néanmoins, cette modulation se fait sur la base de l'expérience et de l'intuition du chargé d'études.

De plus, elle est généralement comprise, par les autorités chargées des évaluations, comme une mesure de la *proportion* de voirie congestionnée, celle-ci étant supposée engendrer des coûts externes de congestion à hauteur de 7,5 min/véh.km (Stif, 2007). Or, cette vision est erronée : en effet, en dehors d'une plage, relativement restreinte en milieu urbain ou périurbain, où le trafic est suffisamment faible pour que les véhicules se déplacent à la vitesse à vide, la congestion est présente. Autrement dit, le coût externe de congestion varie continûment, selon l'intensité d'usage de la voirie, et non sous la forme d'un seuil, qui séparerait une voirie congestionnée, caractérisée par un coût externe de 7,5 min/véh.km et une voirie non congestionnée, où le coût externe serait nul.

Les raisons qui peuvent expliquer que l'emploi d'une telle méthode simplifiée s'est maintenu au cours du temps, malgré l'évolution conjointe des conditions de trafic et des capacités de modélisation, sont les suivantes :

- Dans le contexte d'un milieu urbain complexe, il est à la fois difficile et onéreux de disposer, de calibrer et de maintenir un modèle routier. De nombreux opérateurs de transport collectif ont préféré investir dans un modèle simple, sans prise en compte réelle de la congestion.
- Dans le cadre de la posture de l'inspecteur, une méthode simplifiée a l'avantage de pouvoir être facilement expertisée.
- Une évaluation, même grossière, des gains de temps liés à la décongestion routière serait suffisante pour capturer l'ordre de grandeur de ces gains, et permettre une comparaison de plusieurs projets en concurrence.

5.4. Méthodologies retenues pour l'évaluation des gains de temps de décongestion

Nous venons de le voir, la méthodologie employée actuellement par les opérateurs de transport collectif pour l'évaluation des gains de temps de décongestion présente de sérieuses limitations, au vu de l'évolution des conditions de trafic, et du fait de la large initiative laissée au chargé d'études dans la détermination d'un coefficient de modulation.

Nous allons donc, dans cette section, présenter en détail les différentes méthodologies que nous nous proposons de comparer, dans la section suivante, dans le cadre d'un cas d'étude particulier. Ces méthodes sont au nombre de trois, dont deux sont de type marginaliste et se différencient par le niveau d'agrégation qu'elles supposent, tandis qu'une dernière, servant de référence, est non-marginaliste.

Les évaluations marginalistes reposent toutes les deux sur un calcul du coût externe marginal de la congestion, basé sur la matrice OD des véhicules reportés du mode routier vers le mode collectif. Ainsi, en appelant $[\delta q_{ij}]_{ij}$ cette matrice des reports, les gains de temps de décongestion induits par les flux δq_{ij} dépendent des itinéraires r suivis par ces flux et de l'influence locale que les flux marginaux le long de chaque itinéraire δq_{ij}^r exercent sur les autres usagers de ces itinéraires, à savoir $\delta T_a / \delta q_{ij}^r$ pour un arc a appartenant à l'itinéraire r . Compte tenu de la fonction de temps de parcours de l'arc modélisé de longueur L_a , $T_a = L_a t_a(x_a)$, un flux marginal le long de l'itinéraire r induit le coût externe marginal suivant :

$$\chi_r = \sum_{a \in r} L_a x_a \frac{dt_a}{dx_a} \quad (5.5)$$

En sommant sur l'ensemble des itinéraires suivis entre i et j , avec $\delta q_{ij} = \sum_{r \in (i,j)} \delta q_{ij}^r$:

$$\begin{aligned} \delta \chi_{ij} &= \sum_{r \in (i,j)} \chi_r \delta q_{ij}^r \\ &= \sum_{r \in (i,j)} \left[\sum_{a \in r} L_a x_a \frac{dt_a}{dx_a} \right] \delta q_{ij}^r \\ &= \sum_a L_a x_a \frac{dt_a}{dx_a} \delta^{ij} x_a \end{aligned} \quad (5.6)$$

où $\delta^{ij} x_a = \sum_{r \in (i,j) \cap a} \delta q_{ij}^r$ est le flux reporté de l'arc a provenant de δq_{ij} . Enfin, en sommant sur l'ensemble des paires OD, on obtient le gain de temps total lié à la

décongestion routière induite par le projet, dans une perspective marginaliste :

$$\begin{aligned}\delta\chi &= \sum_{i,j} \delta\chi_{ij} \\ &= \sum_a L_a x_a \frac{dt_a}{dx_a} \delta x_a\end{aligned}\quad (5.7)$$

où $\delta x_a = \sum_{i,j} \delta^{ij} x_a$.

Les trois méthodes marginalistes reposent sur cette même base : selon le degré d'agrégation retenu, un ou plusieurs coefficients remplaceront le terme $x_a \frac{dt_a}{dx_a}$.

Ces coefficients seront tirés des résultats obtenus dans le chapitre précédent. Ainsi, la Table 5.4 fournit les valeurs des coûts externes marginaux moyens par segment de trafic³ en Île-de-France. Néanmoins, pour tenir compte du fait qu'un modèle d'affectation statique produit, dans des conditions de congestion extrême des coûts externes exagérément élevés⁴, nous avons également évalué les coût externes en plafonnant, le cas échéant, dans le calcul par arc, le taux de charge à 1,2. Les valeurs ainsi obtenues figurent dans la Table 5.5.

| HPM | Rapides | Artérielles | Desserte | Total |
|-------------------|---------|-------------|----------|-------|
| Zone centrale | 5,25 | 6,07 | 9,01 | 7,08 |
| Proche périphérie | 3,79 | 8,83 | 8,91 | 6,54 |
| Grande périphérie | 1,64 | 2,51 | 4,15 | 2,65 |
| Total | 2,58 | 4,08 | 5,95 | 4,04 |
| HC | | | | |
| Zone centrale | 0,71 | 0,79 | 0,77 | 0,75 |
| Proche périphérie | 0,13 | 0,57 | 0,67 | 0,39 |
| Grande périphérie | 0,04 | 0,15 | 0,54 | 0,21 |
| Total | 0,14 | 0,28 | 0,61 | 0,31 |
| HPS | | | | |
| Zone centrale | 4,85 | 4,93 | 7,20 | 5,95 |
| Proche périphérie | 1,42 | 4,09 | 5,48 | 3,29 |
| Grande périphérie | 0,64 | 1,49 | 3,25 | 1,66 |
| Total | 1,26 | 2,26 | 4,39 | 2,51 |

TABLE 5.4.: Coût externe marginal moyen par segment en min/véh.km. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

3. Pour la définition précise de ces segments, voir le chapitre précédent.

4. Si l'on peut attendre, dans les cas de congestion sévère, des coûts externes très élevés, la modélisation statique rend difficilement contrôlable la croissance de ces coûts. Une modélisation dynamique, du type modèle de goulot, permet de bien mieux contrôler la croissance des coûts externes lorsque la congestion devient extrême.

| HPM | Rapides | Artérielles | Desserte | Total |
|-------------------|---------|-------------|----------|-------|
| Zone centrale | 4,86 | 5,98 | 8,90 | 6,86 |
| Proche périphérie | 3,07 | 8,17 | 7,84 | 5,73 |
| Grande périphérie | 1,26 | 2,37 | 3,93 | 2,39 |
| Total | 2,10 | 3,82 | 5,54 | 3,64 |
| HC | | | | |
| Zone centrale | 0,71 | 0,79 | 0,77 | 0,75 |
| Proche périphérie | 0,13 | 0,57 | 0,67 | 0,39 |
| Grande périphérie | 0,04 | 0,15 | 0,54 | 0,21 |
| Total | 0,14 | 0,28 | 0,61 | 0,31 |
| HPS | | | | |
| Zone centrale | 4,35 | 4,92 | 7,14 | 5,73 |
| Proche périphérie | 1,39 | 4,08 | 5,33 | 3,23 |
| Grande périphérie | 0,60 | 1,46 | 3,14 | 1,61 |
| Total | 1,18 | 2,24 | 4,27 | 2,43 |

TABLE 5.5.: Coût externe marginal moyen par segment en min/véh.km, avec plafonnement du taux de charge. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

On constate que si, pour l'heure creuse, le plafonnement ne modifie pas les résultats (car très peu d'arcs dépassent un taux de charge de 1,2), les résultats de l'heure de pointe du matin sont sensiblement modifiés, de même que ceux de l'heure de pointe du soir.

Dans les deux cas, remarquons que le coût externe marginal est plus faible sur les voies rapides que sur les voies artérielles et que ce dernier est plus faible que sur les voies de desserte. Il s'agit de l'*effet de capacité* déjà évoqué au chapitre 3 : du fait de leur plus grande capacité, les infrastructures du type voie rapide résistent mieux à la montée en charge, en termes de niveau de service, mais également en termes de coût externe marginal.

Par ailleurs, la Table 5.4 fournit une valeur moyenne du coût externe marginal, au niveau de la région dans son ensemble à l'heure de pointe du matin, d'environ 4 min/véh.km. En supposant un taux d'occupation moyen de 1,2 passagers par véhicule et une valeur monétaire du temps de transport de 15 €₂₀₀₆/h, valeurs raisonnables pour cette période de pointe caractérisée par une forte prédominance des trajets domicile-travail, on aboutit à un coût externe marginal moyen en Île-de-France à l'heure de pointe du matin de 1,2 €₂₀₀₆ par véhicule-kilomètre. Rappelons, à titre de comparaison, que le coût privé moyen pour la même période temporelle s'élève à 0,85 min/km ; autrement dit, à l'heure du pointe du matin en Île-de-France,

le coût externe est plus de quatre fois supérieur au coût privé.

Les valeurs issues de ces deux Tables seront utilisées dans la suite comme bornes d'encadrement des coefficients de décongestion.

5.4.1. La méthode marginaliste traditionnelle

Il s'agit de la variante la plus agrégée, et correspond à l'état actuel de la pratique des opérateurs de transport collectif en Île-de-France.

5.4.1.1. Principe de la méthode

Dans le cadre de cette méthode, le terme de coût marginal $x_a \frac{dt_a}{dx_a}$ est remplacé par un unique coefficient χ_0 . En conséquence, en réécrivant l'équation (5.7), le gain de temps total lié à la décongestion induite par le projet devient :

$$\begin{aligned} \delta\chi^{(1)} &= \chi_0 \sum_a L_a \delta x_a \\ &\approx \chi_0 \sum_{i,j} D_{ij} \delta q_{ij} \end{aligned} \quad (5.8)$$

où D_{ij} est la distance moyenne des itinéraires empruntés entre i et j .

5.4.1.2. Choix du coefficient

La méthode repose, on le voit, sur la détermination de χ_0 pour le projet⁵. Or, cette tâche est complexe, et sujette à caution, puisqu'elle laisse de côté l'extrême variabilité spatiale de la congestion, et tout particulièrement celle de son coût externe marginal, étudiée dans le précédent chapitre de ce mémoire et dans Leurent et Breteau (2009). De plus, elle néglige bien souvent la variabilité temporelle de ce même coût externe marginal.

La Table 5.6 compare les valeurs traditionnellement utilisées, issues du rapport Hautreux, sans abattement (H0), abattue de 50 % (H50), de 60 % (H60) et de 80 % (H80) comme le pratiquent souvent les chargés d'études, avec les valeurs moyennes que nous avons obtenues sur l'ensemble du réseau de l'Île-de-France à l'aide du modèle Quasi-Modus pour différentes périodes temporelles (HPM, HPS, HC), et présentées dans le chapitre précédent. Nous avons considéré (sauf pour l'heure creuse) les valeurs plafonnées comme des bornes inférieures du coût externe marginal moyen, tandis que les valeurs non plafonnées représentent des bornes supérieures.

5. On peut éventuellement supposer que $\chi_0 = A\overline{\chi_0}$, où A est un coefficient d'abattement, spécifique de la zone, et choisi par le chargé d'étude sur la base de son expérience.

| Traditionnels | | | | Renouvelés | | |
|---------------|------|------|------|-------------|-------------|------|
| H0 | H50 | H60 | H80 | HPM | HPS | HC |
| 7,50 | 3,75 | 3,00 | 1,50 | 3,64 – 4,04 | 2,43 – 2,51 | 0,31 |

TABLE 5.6.: Comparaison des coefficients Hautreux avec nos évaluations des coûts externes moyens en Île-de-France, min/véh.km.

Il apparaît que si la valeur traditionnelle sans abattement est plus élevée que nos évaluations, la valeur abattue à 50 % correspond assez bien à la valeur moyenne du coût externe marginal de congestion à l'heure de pointe du matin.

Cependant, l'abattement est généralement employé par les chargés d'étude pour rendre compte de l'environnement du terrain d'étude. Ainsi, la valeur abattue à 50 % est couramment utilisée pour les zones urbaines centrales, la valeur abattue à 60 % est employée comme valeur par défaut correspondant à des zones urbaines moins centrales, tandis que l'abattement à 80 % est réservé pour les zones périurbaines. La Table 5.7 compare donc les valeurs issues du rapport Hautreux (H50, H60, H80), avec les valeurs que nous avons obtenues à l'heure de pointe du matin pour les trois types géographiques retenus dans notre étude, la zone centrale (ZC), la proche périphérie (PP) et la grande périphérie (GP).

| Traditionnels | | | Renouvelés | | | |
|---------------|------|------|-------------|-------------|-------------|-------------|
| H50 | H60 | H80 | ZC | PP | GP | Ensemble |
| 3,75 | 3,00 | 1,50 | 6,86 – 7,08 | 5,73 – 6,54 | 2,39 – 2,65 | 3,64 – 4,04 |

TABLE 5.7.: Comparaison des coefficients Hautreux avec nos évaluations des coûts externes moyens à l'HPM en Île-de-France, min/véh.km.

On constate que les résultats issus de nos simulations sont, pour chaque type géographique, plus élevés que les valeurs généralement retenues par les chargés d'étude, du moins pour l'heure de pointe du matin. Néanmoins, ces moyennes cachent la très grande variabilité des valeurs, en particulier entre les types routiers. Il ressort de ces différentes comparaisons que les coefficients généralement retenus par les chargés d'étude ne correspondent que partiellement à la réalité actuelle de la congestion en Île-de-France. Nous avons donc choisi, dans l'application de la méthode marginaliste traditionnelle, de fournir les résultats sous forme d'intervalles⁶.

6. Il ne s'agit pas à proprement parler d'intervalles de confiance. Il serait possible d'obtenir de tels intervalles à partir des valeurs des quantiles à 5 % et 95 % des coûts externes marginaux de décongestion.

5.4.1.3. Mise en œuvre pratique

La mise en œuvre de la méthode marginaliste traditionnelle est particulièrement simple (c'est d'ailleurs là son intérêt principal) puisqu'elle ne repose, en théorie, que sur la connaissance de la quantité de trafic reportée du mode routier vers le mode collectif (ou inversement). Celle-ci peut être évaluée, à l'étape de choix modal, par le trafic prévu sur l'infrastructure de transport collectif.

En effet, le gain de décongestion est obtenu par multiplication de cette quantité de trafic, exprimée en véhicules-kilomètres, par un coefficient de décongestion, dépendant éventuellement du milieu géographique ou du type de voies concernées par le projet, exprimé en minutes par véhicule-kilomètre. Nous verrons plus en détail l'application de cette méthode dans le cas d'étude que nous avons retenu.

5.4.2. La méthode marginaliste segmentée

Nous l'avons vu au chapitre précédent, la très forte variabilité des coûts externes de congestion sur le réseau francilien impose une segmentation de ce réseau, selon des critères croisant type routier et milieu géographique. Comme nous allons le voir, cette segmentation peut être utilisée pour déterminer des coefficients de décongestion par segment.

5.4.2.1. Principe de la méthode

En supposant que les arcs a sont regroupés en segments z caractérisés par des coefficients de décongestion χ_z , nous pouvons réécrire l'équation (5.7) de la manière suivante :

$$\begin{aligned}\delta\chi^{(2)} &= \sum_a \chi_{z(a)} L_a \delta x_a \\ &= \sum_{z \in Z} \chi_z \left[\sum_{a \in z} L_a \delta x_a \right]\end{aligned}\tag{5.9}$$

La méthode repose donc d'une part sur la détermination des coefficients par segment, et d'autre part sur l'évaluation des termes $\sum_{a \in z} L_a \delta x_a$, autrement dit des quantités de trafic reportées, sur chaque segment du réseau. C'est dans la nécessité de cette évaluation que réside la principale différence avec la méthode traditionnelle. Si cela représente une contrainte, cela permet également d'atteindre un niveau de précision supplémentaire, puisqu'une partie de la variabilité des coûts externes de congestion est capturée par l'utilisation de plusieurs coefficients de décongestion.

Comme pour la méthode traditionnelle, nous emploierons, dans l'étude de cas, les valeurs présentées dans les Tables 5.4 et 5.5, respectivement comme valeurs hautes et basses du coefficient de décongestion par segment. La question du choix d'un

coefficient unique, ou d'un pourcentage d'abattement, disparaît donc, au profit de l'évaluation correcte des quantités de trafic routier économisées. Cette évaluation est tout aussi cruciale, mais peut reposer sur des modèles, éventuellement très simples, de trafic, et non pas uniquement sur l'expérience des chargés d'étude.

5.4.2.2. Mise en œuvre pratique

La mise en œuvre de la méthode repose d'une part sur la donnée d'une table de coefficients correspondant aux différents segments retenus, et d'autre part sur la détermination des quantités de trafic économisées sur chaque segment. Leur évaluation est effectuée par décomposition des distances parcourues moyennes entre zones d'origine et de destination, selon les segments de réseau utilisés. En pratique, c'est le plus court chemin (en temps, à l'heure de pointe du matin) entre chaque paire OD qui est utilisé comme base de calcul pour déterminer la proportion d'utilisation de chacun des segments.

Le gain total de décongestion est ensuite obtenu par simple multiplication de distances parcourues sur les différents segments par le coefficient de décongestion moyen correspondant à ce segment.

5.4.3. La méthode non marginaliste

Dans cette approche, les variations δx_a des débits locaux x_a entre deux états (hors et avec projet) résultent d'un mécanisme plus complexe, qui incorpore non seulement la sensibilité au trafic du temps de parcours, mais aussi le choix d'itinéraire par les usagers. De plus, cette méthode permet de tenir compte de la non linéarité du phénomène de congestion.

Plus précisément, une variation du débit local Δx_a entraîne une variation ΔT_a du temps local (moyen par usager), et celle-ci peut entraîner une réaffectation des déplacements aux itinéraires et ce pour l'ensemble des déplacements, et pas uniquement pour les volumes O-D qui font la différence entre les deux états. Cet élargissement à l'ensemble des déplacements est un effet de réseau qui s'ajoute à celui du report d'itinéraire.

5.4.3.1. Principe de la méthode

Entre un état *hors projet*, caractérisé par des flux x_a^0 et des temps de franchissement T_a^0 sur les arcs a , et un état *avec projet*, caractérisé par des flux x_a^1 et des temps de franchissement T_a^1 , le gain social de décongestion s'écrit :

$$\Delta C^{(0)} = \sum_a x_a^0 T_a^0 - \sum_a x_a^1 T_a^1 \quad (5.10)$$

Néanmoins, ce gain social est composé d'un gain *privé* correspondant au temps que les usagers reportés sur le mode collectif passaient sur la route (et qu'ils ne passent plus) et d'un gain *externe*, correspondant au temps que ces usagers reportés faisaient subir aux autres usagers, qui eux, restent sur la route. Autrement dit, la grandeur à comparer avec les autres méthodes est fournie par l'expression suivante :

$$\begin{aligned}\Delta\chi^{(0)} &\approx \sum_a (x_a^0 T_a^0 - x_a^1 T_a^1) - \sum_a T_a^0 (x_a^0 - x_a^1) \\ &\approx \sum_a x_a^1 (T_a^0 - T_a^1)\end{aligned}\quad (5.11)$$

Cette formulation convient pour des valeurs assez faibles de $x_a^0 - x_a^1$, sans qu'elles soient pour autant marginales. Autrement dit, cette formule convient pour des variations assez faibles, telles que T_a^0 représente au premier ordre le coût qui aurait été subi par un usager reporté. Si on approche ce coût individuel plus justement par $(T_a^0 + T_a^1)/2$, alors, le gain externe de décongestion est fourni par :

$$\Delta\chi^{(0)'} \approx \sum_a \frac{x_a^0 + x_a^1}{2} (T_a^0 - T_a^1)\quad (5.12)$$

Nous verrons que compte tenu des très faibles valeurs des termes $x_a^0 - x_a^1$ dans le cas d'étude qui nous concerne, nous avons choisi de retenir la forme (5.11). Par ailleurs, à la différence des deux méthodes marginalistes présentées précédemment, la méthode non-marginaliste fournit une valeur unique pour le gain de temps de décongestion, et non un intervalle.

5.4.3.2. Mise en œuvre pratique

La mise en œuvre pratique de cette méthode non marginale passe par la réalisation de deux affectations à l'équilibre de l'utilisateur sur le réseau routier, l'une à partir de la matrice OD hors projet, l'autre à partir de la matrice avec projet, c'est-à-dire ne comprenant pas les déplacements reportés vers le mode collectif.

La différence, en termes de temps total passé sur le réseau, entre les résultats de ces deux affectations fournit directement la valeur $\Delta C^{(0)}$. L'obtention de $\Delta\chi^{(0)}$ passe par l'évaluation du temps privé économisé, $\sum_a T_a^0 (x_a^0 - x_a^1)$. Or, cette expression peut se réécrire, en distinguant les différentes paires OD, et en notant Δx_p est la

variation du flux total de déplacement pour la paire OD p :

$$\begin{aligned}
 \sum_a T_a^0(x_a^0 - x_a^1) &= \sum_{p \in OD} \sum_{a \in p} T_a^0(x_a^0 - x_a^1) \\
 &= \sum_{p \in OD} T_p^0 \sum_{a \in p} x_a^0 - x_a^1 \\
 &= \sum_{p \in OD} T_p^0 \Delta x_p
 \end{aligned} \tag{5.13}$$

Le passage de la première à la deuxième ligne est rendu possible par application de la définition de l'équilibre de Wardrop (1952). Autrement dit, le coût privé total économisé peut être obtenu à partir d'une part de la matrice des temps de parcours en situation de projet, et d'autre part de la matrice des allègements de trafic liés aux reports vers le mode collectif.

5.5. Comparaison des méthodes sur un cas d'application

Nous avons établi, à la section précédente, trois méthodes d'évaluation des gains de décongestion routière lors de la mise en place d'un projet de transport collectif. De manière à les comparer, et mieux saisir leurs avantages et limites, nous allons les appliquer successivement sur un cas d'étude particulier, choisi pour son caractère relativement typique des opérations de mise en place d'un nouveau mode de transport collectif en Île-de-France.

5.5.1. Présentation du cas d'étude : le projet de tram-train Massy-Évry

5.5.1.1. Contexte et description du projet

Selon le Dossier d'Objectifs et de Caractéristiques Principales (DOCP) établi par la SNCF pour le compte du Syndicat des transports d'Île-de-France (Stif) et de Réseau ferré de France (RFF) (SNCF, 2008), le projet consiste en une liaison entre Massy-Palaiseau et Évry-Courcouronnes d'une longueur totale de 20,7 km, dont l'exploitation et le matériel roulant sont de type tram-train (21 rames dans le projet étudié). Sur ces 20,7 km, une partie, plus précisément 9,1 km entre Massy-Palaiseau et Épinay-sur-Orge, reprendrait les emprises et l'infrastructure ferroviaire de la ligne de Grande Ceinture (aujourd'hui ligne RER C – Réseau Express Régional) ; tandis que 11,6 km seraient des voies nouvelles traitées en insertion paysagère urbaine et constitueraient la partie « tramway ». La Figure 5.4 présente le tracé du projet,

avec les arrêts ainsi que le type d'exploitation prévu sur les différentes portions.

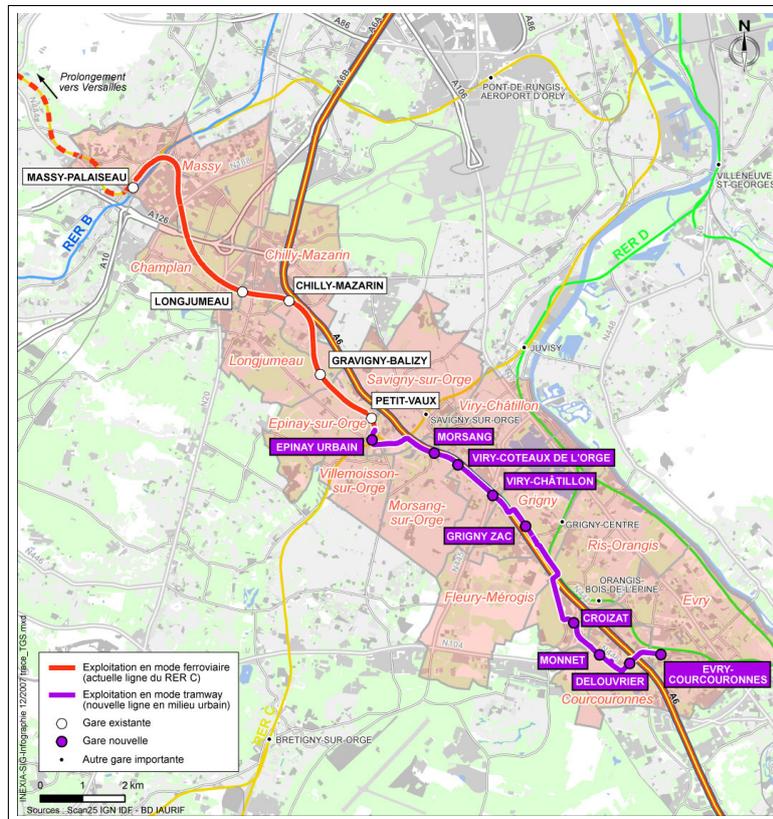


FIGURE 5.4.: Plan synthétique du projet de tram-train Massy-Évry. *Source* : SNCF (2008)

De façon plus détaillée, entre Massy-Palaiseau et Épinay-sur-Orge, les dessertes se substitueraient aux dessertes du RER C. Les dessertes seraient organisées en missions omnibus, assurées à la fréquence d'un service par sens toutes les 10 minutes en heure de pointe, et d'un service par sens toutes les 15 minutes en heure creuse. Le temps de parcours de l'ensemble de la ligne est estimé à 32 minutes pour chaque sens. La vitesse commerciale sur la ligne serait donc de 39 km/h en moyenne — vitesse élevée pour du transport collectif urbain mais faible pour du transport ferroviaire interurbain.

Les atouts du projet sur le secteur reposent sur la desserte cadencée et avec une bonne fréquence sur l'axe concerné, entre Massy-Palaiseau et Évry-Courcouronnes, et sur des correspondances directes avec les composantes radiales des lignes de RER B, C et D.

5.5.1.2. Prévisions de trafic et reports potentiels

La fréquentation attendue, obtenue avec le modèle de trafic Antonin du Stif est de 4 600 voyageurs à l'heure de pointe du matin, soit 27 600 voyageurs par jour à l'horizon 2008, choisie comme année de référence. La longueur moyenne des trajets (voyages TC de gare de montée à gare de descente) sur la ligne de tram-train serait de 6,5 km, avec une faible proportion de déplacements utilisant le tram-train de pôle à pôle entre Massy et Évry.

Pour appréhender l'influence potentielle sur le réseau routier, la Figure 5.5 indique la position du secteur d'insertion relativement au réseau routier magistral. Les axes routiers magistraux potentiellement impactés par le projet sont donc : la RN7 et l'A6, en premier lieu, qui empruntent entièrement le corridor d'insertion, la RN445 en partie, et les voies desservant Massy : A10, A6, RN20 et RN188, ainsi que la Francilienne (A104) par effet de réseau.

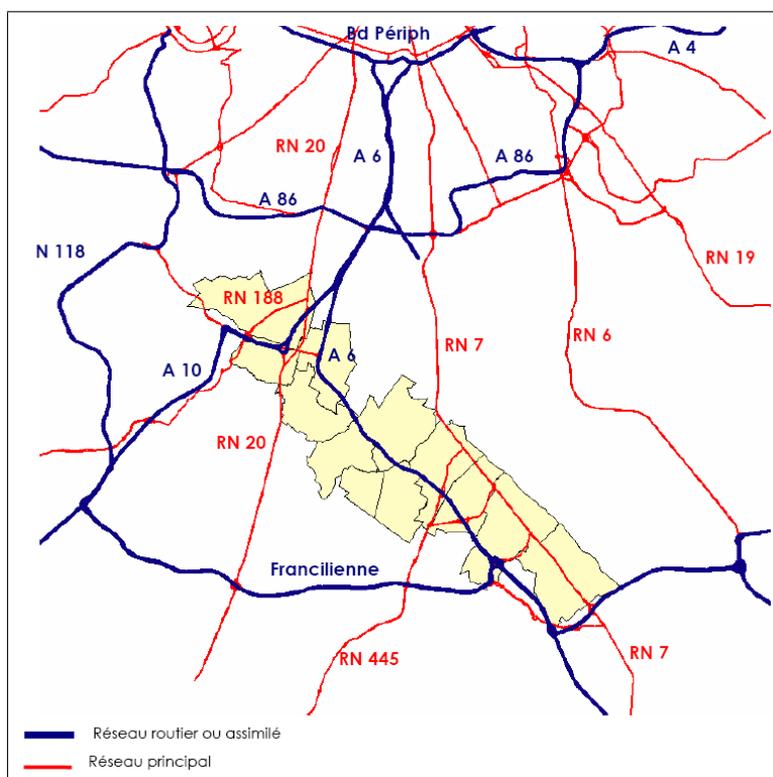


FIGURE 5.5.: Plan du réseau routier magistral autour du projet de tram-train Massy-Évry. *Source* : SNCF (2008)

La simulation avec le modèle Antonin du Stif a montré que la part des anciens utilisateurs de la voiture particulière parmi les utilisateurs du tram-train est de 20 %. La distance qu'ils parcouraient en moyenne en voiture particulière était de 12,5 km.

5.5.1.3. Bilan socio-économique

Le DOCP fournit également un bilan socio-économique mettant en balance les différents coûts et avantages attendus du projet, parmi lesquels figurent les gains de décongestion du réseau routier.

Ainsi, le bilan socio-économique, évalué aux conditions économiques de 2006 a été établi à partir :

- des coûts d'investissement, qui s'élèvent à 391 M€₂₀₀₆ HT (CE 2006), et se répartissent en 307 M€ HT pour les installations fixes et 84 M€ HT pour le matériel roulant ;
- des coûts d'exploitation, estimés à 2,46 M€ annuels en plus des 8,84 M€ annuels pour le RER C en situation de référence ;
- des gains de temps, valorisés par une valeur du temps de 15,63 €₂₀₀₆/h par voyageur ;
- concernant le mode automobile, des gains des usagers (décongestion de la voirie, entretien de la voirie, économies d'utilisation et de stationnement), obtenus en raison des reports modaux vers les transports collectifs ;
- des gains sur les externalités, valorisés sur la base des valeurs tutélaires.

De façon plus précise, sur le réseau de TC les gains de temps nets du projet correspondent au solde :

- des gains des utilisateurs du tram-train ;
- des pertes liées à la disparition de certaines liaisons directes.

Le gain de temps moyen pour chaque ancien utilisateur des transports collectifs est de l'ordre de 5 minutes. Par ailleurs, le gain de temps moyen par nouvel usager TC, qu'il soit reporté du mode automobile ou qu'il s'agisse d'un déplacement induit, est fixé conventionnellement à la moitié du gain moyen d'un ancien usager TC. En ce qui concerne les gains sur la voiture particulière, et sur les externalités, des valeurs conventionnelles ou tutélaires ont été utilisées. En particulier la valeur Hautreux traditionnelle de 0,125 h/véh.km pour la part supposée congestionnée du trafic a été employée pour estimer les gains de décongestion. La part congestionnée a été fixée à 20 % de l'ensemble des véhicules-kilomètres économisés⁷.

Le résultat de ce bilan est un taux de rentabilité interne de 8,2 %. Il convient de noter que ce résultat repose sur un certain nombre d'hypothèses relativement critiquables. En particulier, on l'a vu, le coefficient d'abattement retenu pour le calcul des gains de décongestion routière n'est pas très réaliste, au vu de nos résultats d'analyse du coût externe de congestion en Île-de-France.

7. On pourra se reporter au 5.3.3 pour une critique de cette interprétation du coefficient d'abattement.

5.5.2. Comparaison des différentes méthodes

L'étude d'évaluation des gains de décongestion routière repose sur une simulation du trafic reporté depuis le mode automobile vers le mode TC. Pour l'accomplir il faut disposer au minimum :

- d'un modèle d'affectation du trafic automobile aux itinéraires sur le réseau routier ;
- d'un modèle d'affectation du trafic aux itinéraires sur le réseau de transports collectifs ;
- d'un modèle de choix modal entre automobile et TC.

Le Stif disposant de tels modèles de simulations pour le projet considéré, il nous a fourni directement le résultat attendu, autrement dit la matrice OD des allègements routiers, avec les informations de géocodage associées. Néanmoins, les différences de zonage entre le modèle employé par le Stif et le modèle Quasi-Modus que nous avons utilisé, ont rendu délicate l'utilisation, dans Quasi-Modus, de la matrice des allègements routiers. Une présentation détaillée des procédures suivies pour chacune des méthodes est fournie en Annexe 5.A.

5.5.2.1. La référence : la méthode non marginaliste

Nous considérerons la méthode non marginaliste, présentée au 5.4.3, comme la méthode de référence, servant de base aux comparaisons avec les autres méthodes. Elle se base en effet sur une différence entre deux situations d'équilibre, et permet donc de prendre en compte l'ensemble des effets liés à la décongestion : non linéarité et effets de réseau.

L'application de la méthode non marginaliste au cas du projet tram-train Massy-Évry a fourni, pour l'heure de pointe du matin⁸, les résultats présentés dans la Table 5.8.

Ces résultats amènent un certain nombre de commentaires. Ainsi, le gain de temps total lié au projet, pendant l'heure de pointe du matin s'élève à un peu plus de 700 heures, ce qui est relativement faible, mais néanmoins non négligeable. En ce qui concerne la répartition des gains entre les différents segments, on constate que la majorité des gains de décongestion, soit près de 70 %, ont lieu en grande périphérie, et même plus précisément, sur les voies rapides de cette zone. Or, le gain marginal d'un véhicule-kilomètre économisé sur une voie rapide est généralement plus faible que le gain marginal du même véhicule-kilomètre sur une voie de desserte. Par conséquent, le fait que le gain total sur les voies rapides soit sensiblement identique

8. Le choix de cette période temporelle est lié au fait que la pointe du matin est généralement considérée comme plus « dimensionnante » que celle du soir, en particulier pour les projets de transport collectif. Par conséquent, le Stif nous a fourni une matrice des allègements routiers pendant la période de pointe du matin.

| Gains (h) | Rapides | Artérielles | Desserte | Total |
|-------------------|---------|-------------|----------|-------|
| Zone centrale | 7 | 8 | 12 | 27 |
| Proche périphérie | 56 | 68 | 76 | 200 |
| Grande périphérie | 174 | 145 | 173 | 492 |
| Total | 237 | 221 | 261 | 719 |

TABLE 5.8.: Résultats de l'application de la méthode non marginaliste au cas du projet de tram-train. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

au gain total sur les voies de desserte implique que les volumes reportés à partir des voies rapides sont très supérieurs aux volumes reportés à partir des voies de desserte.

La Figure 5.6 permet de mieux comprendre l'origine de ces chiffres. Elle met en évidence les arcs sur lesquels se situent les principaux reports liés au projet. On constate qu'en dehors de quelques artefacts anecdotiques, liés à des problèmes de critères de convergence (Bar-Gera *et al.*, 2010), la majorité des allègements ont lieu à proximité du projet de tram-train, en particulier le long de l'autoroute A6, de la Francilienne, de l'A10, de la RN20 et de la RN445 (*cf.* Figure 5.5). On notera également, à des fins de comparaison avec les autres méthodes, que l'économie, en termes de distance totale parcourue, trouvée avec cette méthode, s'élève à 10 867 véh.km.

5.5.2.2. Le choix de la simplicité : la méthode traditionnelle

De manière à appliquer la méthode traditionnelle d'évaluation des gains de temps de décongestion du réseau routier, nous avons du choisir les coefficients à employer (le détail de la procédure est présenté en Annexe 5.A.2). Ainsi, nous avons utilisé, pour déterminer un intervalle « *ancien* », le coefficient Hautreux abattu à 80 % (c'est l'abattement qui a été retenu par le STIF pour ce projet) comme borne inférieure et le coefficient abattu à 50 % comme borne supérieure. Pour déterminer les intervalles « *renouvelés* », nous avons choisi les bornes fournies à la Table 5.7 pour la grande périphérie, puisque c'est dans cette zone, que le projet a des impacts, majoritairement. Nous avons également fait figurer un intervalle obtenu en utilisant les bornes correspondant à la proche périphérie, ainsi qu'un intervalle obtenu à partir des bornes correspondant à l'Île-de-France dans son ensemble, dont les coefficients figurent également dans la Table 5.7.

Les résultats de la méthode marginaliste traditionnelle dans le cas du projet tram-train Massy-Évry figurent dans la Table 5.9, pour l'heure de pointe du matin. Ils reposent sur une évaluation de l'économie, en termes de distance totale parcourue,

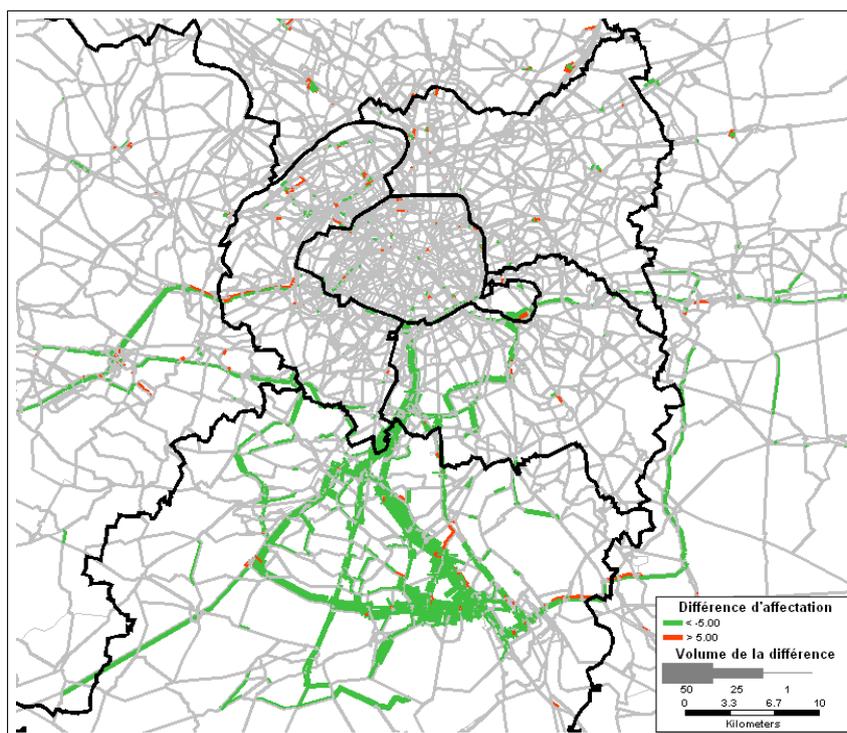


FIGURE 5.6.: Reports de trafics à l'HPM liés au projet de tram-train. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

qui s'élève à 11 344 véh.km, soit moins de 500 véh.km de plus que l'estimation fournie par la méthode non marginaliste.

| Coefficients | Gains (h) |
|--------------------------------|---------------|
| Anciens | 284 – 709 |
| Renouvelés - Grande Périphérie | 452 – 501 |
| Renouvelés - Proche Périphérie | 1 083 – 1 237 |
| Renouvelés - Île-de-France | 688 – 764 |

TABLE 5.9.: Résultats de l'application de la méthode traditionnelle au cas du projet de tram-train. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

On constate en premier lieu que la valeur de référence, présentée précédemment, de 719 heures économisées, n'est comprise que dans le dernier intervalle, obtenu à partir des valeurs moyennes renouvelées de l'Île-de-France. Cet intervalle présente également l'avantage d'être relativement réduit, puisqu'il correspond à une valeur de 726 h, $\pm 5\%$. Cela signifie que les conditions de circulation autour du projet sont assez représentatives des conditions en Île-de-France. Cela implique également que pour un projet différent, inséré dans un milieu géographique différent, il est

probable qu'utiliser des coefficients moyens se rapportant à l'Île-de-France dans son ensemble ne fournirait pas de résultats satisfaisants.

L'intervalle *ancien* est quant à lui très large. On notera également que la valeur basse, 284 heures, correspond au coefficient d'abattement de 80 % retenu par le Stif pour cette étude, et qu'il sous-estime fortement les gains de décongestion.

Contrairement à ce qu'on aurait pu attendre, l'intervalle basé sur les coefficients de la grande périphérie sous-estime également ces gains. Cela est clairement dû au fait que la grande périphérie parisienne est vaste et regroupe des situations urbaines extrêmement variées en termes de trafic. Plus précisément, le projet de tram-train est situé en grande périphérie proche, où les conditions de circulation sont souvent difficiles, entraînant des coûts externes marginaux de congestion élevés. Utiliser, pour évaluer les gains de décongestion, des valeurs moyennes de grande périphérie, où les conditions de circulation sont généralement meilleures qu'autour du projet, conduit à une sous-estimation de ces gains. Néanmoins, on constate qu'utiliser les coefficients correspondant à la proche périphérie aboutit à l'inverse à un intervalle grandement surestimé.

En dépit de sa grande simplicité de mise en œuvre, la méthode marginaliste traditionnelle n'offre pas la précision nécessaire à une évaluation solide des gains de décongestion routière.

5.5.2.3. Précision et simplicité : la méthode marginaliste segmentée

La méthode marginaliste segmentée repose sur l'usage de l'ensemble des coefficients contenus dans les Tables 5.4 et 5.5. Contrairement à la méthode précédente, la méthode marginaliste segmentée permet d'avoir accès au détail des gains de décongestion par segment de trafic. Ainsi, la Table 5.10 présente-t-elle les résultats obtenus par cette méthode, dans le cas du projet de tram-train Massy-Évry. Ceux-ci reposent sur une estimation de l'économie en termes de distance totale parcourue qui s'élève à 10 599 véh.km⁹, soit une valeur très proche de celle fournie par la méthode non marginaliste.

On constate en premier lieu que cette méthode sous-estime les gains totaux de décongestion de près de 25 %. Toutefois, l'examen du tableau permet de constater que cette sous-estimation est principalement due à une mauvaise évaluation des gains de décongestion pour les segments de grande périphérie. L'explication que nous avons fournie précédemment, liée à la sous-estimation des coefficients de grande périphérie dans le cas du projet de tram-train est donc également valable ici.

Pour les autres segments en effet, les estimations fournies par la méthode segmentée sont relativement proches des résultats de la méthode non marginaliste de

9. La distance parcourue sur les connecteurs n'est pas prise en compte par cette méthode.

| Gains (h) | Rapides | Artérielles | Desserte | Total |
|-------------------|---------|-------------|----------|---------|
| Zone centrale | 22–23 | 5 | 15 | 42–43 |
| Proche périphérie | 74–91 | 63–68 | 40–45 | 177–204 |
| Grande périphérie | 82–106 | 88–94 | 121–127 | 291–327 |
| Total | 178–220 | 156–167 | 176–187 | 510–574 |

TABLE 5.10.: Résultats de l'application de la méthode marginaliste segmentée au cas du projet de tram-train. *Source* : DRIEA/SCEP/DPAT et calculs de l'auteur.

référence. On retrouve de plus le fait que les gains ont majoritairement lieu en grande périphérie (ce qui accentue la sous-estimation), et sur les voies rapides, ce que la méthode non marginaliste de référence avec également mis en évidence.

Une segmentation plus adaptée à la grande périphérie, distinguant par exemple les zones situées en deçà de la Francilienne, de celles situées au delà, permettrait donc d'améliorer fortement la qualité des évaluations fournies par cette méthode marginaliste, qui présente l'avantage d'être très simple à mettre en œuvre (*cf.* Annexe 5.A.3 pour le détail de la procédure). Autrement dit, l'utilisation de coefficients segmentés permet de concilier précision et interprétabilité des résultats avec la simplicité de mise en œuvre qui était un de nos critères de comparaison.

5.5.2.4. Bilan de la comparaison des méthodes

Les trois méthodes que nous avons présentées et comparées se distinguent principalement par la robustesse des résultats, leur niveau de détail, et la simplicité de mise en œuvre. Le choix d'une méthode parmi elles dépend donc de l'importance donnée à chacun de ces critères lors de l'évaluation des gains de décongestion liés à un projet d'infrastructure de transport.

Plus précisément, la méthode non marginaliste fournit des résultats à la fois robustes et détaillés. Il est en effet possible de les analyser aussi finement que voulu, jusqu'au niveau de l'arc routier. Cependant, l'utilisation de cette méthode implique d'une part de disposer d'un modèle d'affectation routière prenant en compte la congestion par l'intermédiaire de fonction de temps de parcours et d'un algorithme d'obtention de l'équilibre, et d'autre part, que les sorties des modèles routier et TC soient cohérents entre eux, et puissent être fusionnés. Les différents modèles utilisés devront également avoir été finement calibrés. Autrement dit, il s'agit de disposer d'un modèle multimodal intégré moderne, sous peine de devoir mettre en cohérence les sorties des deux modèles, tâche longue et approximative. Si ces conditions sont remplies, cette méthode non marginaliste est à la fois simple d'utilisation et fiable. En conséquence, elle s'inscrit clairement dans la posture de l'ingénieur conseil, et

beaucoup moins dans celle de l'inspecteur.

La méthode marginaliste segmentée, pour sa part, fournit des résultats relativement robustes (à condition que la variabilité intra-classe des segments en termes coût externe de congestion soit assez faible), et un niveau de détail important. La mise en œuvre de la méthode est relativement simple, une fois les coefficients par segments obtenus. Ces derniers peuvent ensuite être directement appliqués à des allègements routiers par segment fournis en véhicules-kilomètres, obtenus par affectation au plus court chemin sur un réseau chargé. Autrement dit, cette méthode, si elle n'offre pas la robustesse de la précédente, permet néanmoins d'atteindre un niveau de précision satisfaisant tout en nécessitant une mise en œuvre aisée. Elle représente un compromis valable entre la posture de l'ingénieur conseil et celle de l'inspecteur.

La méthode marginaliste traditionnelle fournit un résultat dont la robustesse est difficile à estimer : lorsque le projet s'insère dans un contexte assez représentatif d'un des segments que nous avons définis, voire de l'Île-de-France dans son ensemble, les résultats sont assez solides ; en revanche, lorsque le contexte d'insertion est mixte, ou très particulier, cette méthode n'est pas adaptée. Elle présente en revanche l'avantage d'être d'une mise en œuvre très aisée, plus simple encore que la méthode segmentée. Elle convient donc parfaitement à la posture de l'inspecteur, mais ne devrait pas être employée par l'ingénieur conseil.

Notons tout de même qu'en dehors de la précision des méthodes proposées pour évaluer les gains de décongestion se pose la question de l'expansion temporelle des résultats obtenus pour une (ou éventuellement plusieurs) périodes temporelles bien spécifiques (HPM, HPS, etc.). En effet, la très forte non-linéarité du coût externe marginal rend délicate l'expansion temporelle à partir de coefficients de passages qui, par définition, supposent une linéarité du phénomène.

La Figure 5.7 propose, en guise de synthèse de cette comparaison des différentes méthodes, un arbre de décision pour l'évaluation des gains de décongestion (ou des coûts de congestion) dans le cas d'un projet de transport.

L'utilisation d'une simulation dynamique du trafic routier permet de mesurer les gains de décongestion dans les cas de congestion sévère avec formation de files d'attente. Ainsi, Aguiléra et Leurent (2010) proposent une méthode s'appuyant sur un modèle dynamique opérationnel.

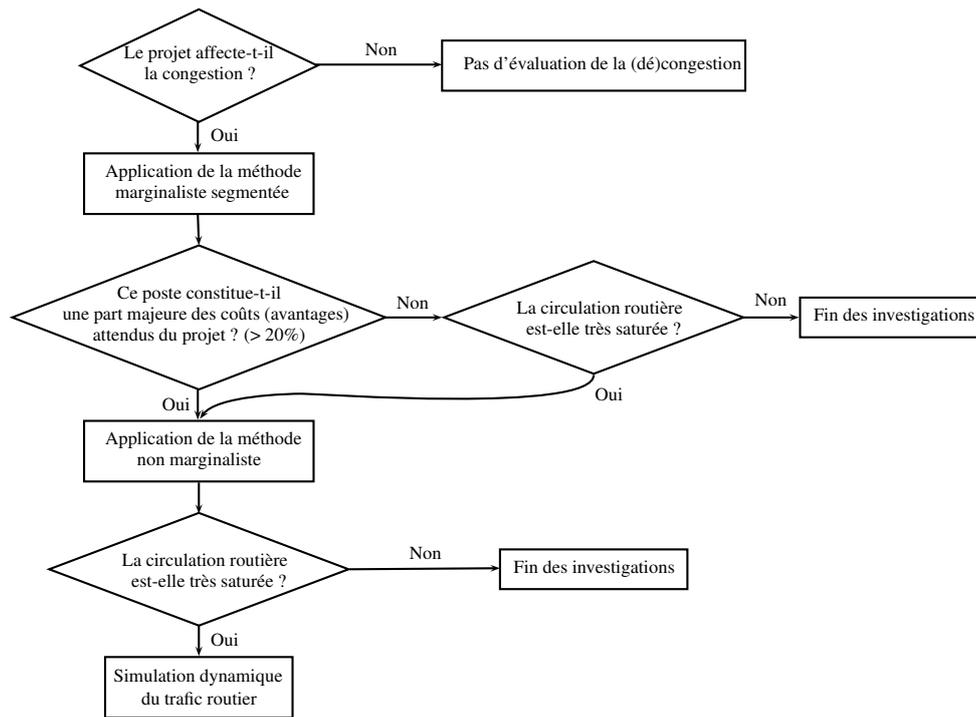


FIGURE 5.7.: Arbre de décision pour l'évaluation des gains de décongestion. *Source* : Leurent *et al.* (2009)

5.6. Conclusion

Intérêt des méthodes marginalistes

Évaluer les gains dus à la décongestion de la voirie liée à un projet de transport collectif à l'aide d'une méthode à la fois robuste, fiable et de mise en œuvre aisée s'est avéré faisable. Néanmoins, un certain niveau de segmentation du réseau routier, suivant des critères comme le type de voie et le milieu géographique, s'est révélé nécessaire afin d'assurer la fiabilité souhaitée.

Autrement dit, la méthode marginaliste traditionnelle, consistant en l'utilisation d'un unique coefficient (éventuellement modulé) applicable à l'ensemble des véhicules-kilomètres économisés, ne permet pas d'atteindre un niveau de robustesse suffisant selon nous. En revanche, elle reste valable, avec des coefficients toutefois renouvelés et modernisés, pour une utilisation dans la posture de l'inspecteur, c'est-à-dire pour fournir un ordre de grandeur. La méthode marginaliste segmentée s'est avérée pour sa part capable d'être employée dans les deux types de posture. Néanmoins, la méthode non marginaliste par différence d'affectations reste la méthode de référence à déployer en priorité. Mais elle nécessite la mise en œuvre d'un modèle multimodal performant et bien calibré, ce dont peuvent se passer les méthodes

marginalistes.

En conclusion, la méthode non marginaliste de référence devrait être choisie par les ingénieurs conseils pour les projets importants où les gains de décongestion seront vraisemblablement élevés. En revanche, pour les inspecteurs, ou dans le cas de projets où les gains de décongestion seront *a priori* plus faibles, la méthode marginaliste segmentée présente l'intérêt d'être très simple à mettre en œuvre, tout en offrant un bon niveau de précision et de robustesse.

L'évaluation de la congestion : vers la prise en compte de la congestion de l'espace

Nous l'avons vu dans la première section de ce chapitre, les évaluations marginalistes sont utilisées depuis plusieurs décennies dans le cas des projets de transport. En effet, elles sont bien souvent adaptées, puisqu'un projet n'induit généralement que des modifications marginales dans les choix d'itinéraires des usagers.

L'évaluation des projets urbains, plus directement liés à l'usage du sol, ne prend bien souvent pas en compte les coûts ou gains de congestion ou de décongestion de l'espace, c'est-à-dire les coûts ou les gains sociaux directement liés à l'usage du sol, et plus particulièrement à l'intensité de cet usage. Les travaux de Jara-Diaz (1986) montrent la difficulté d'éviter les doubles comptes dans une telle évaluation.

Nous pensons néanmoins que l'utilisation de méthodes marginalistes, inspirées de celles que nous avons déployées dans ce chapitre, permettrait de faciliter la prise en compte de ce type particulier d'externalités liées à l'usage du sol. Autrement dit, si les modèles d'interaction urbanisme et transport (modèles *LUTI*) peuvent apporter des réponses à ce type de questionnement, ils sont d'une mise en œuvre extrêmement lourde, les destinant bien souvent à une utilisation restreinte à la recherche, tandis qu'ils seraient utiles comme outils d'aide à la décision. Une méthode de mise en œuvre simple qui fournirait un ordre de grandeur des coûts ou gains à attendre d'un projet urbain en termes de congestion de l'espace pourrait remplir ce rôle.

Annexe 5.A Détail des procédures d'évaluation des gains de décongestion

5.A.1 Méthode non marginaliste

La méthode non marginaliste repose sur une différence d'affectation, entre un état hors projet et un état avec projet. Par conséquent, il est nécessaire de disposer d'une matrice des reports se rapportant au même zonage que la matrice originale. C'est le cas lorsque les modèles routiers et TC sont conçus sur le même découpage. Néanmoins, ce n'est pas toujours le cas, même au sein d'un même modèle multimodal. Cela l'est encore moins lorsque comme nous, il s'agit d'utiliser des matrices issues de modèles totalement différents. Comme on peut le constater sur la Figure 5.8, les zonages employés dans Quasi-Modus et dans le modèle du Stif adapté au cas du tram-train sont très différents.

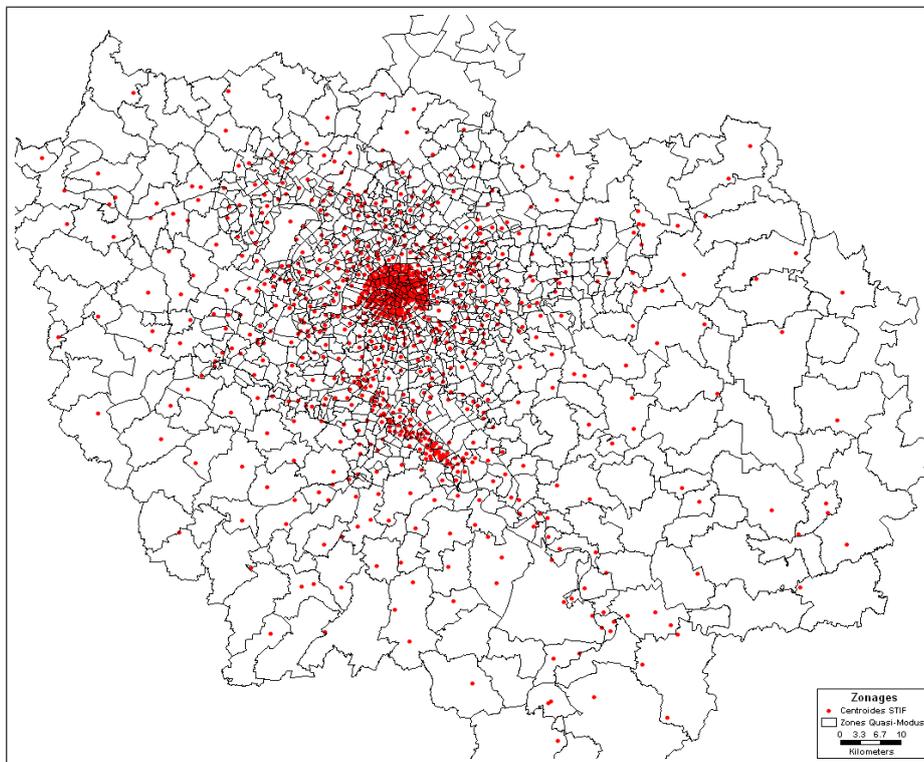


FIGURE 5.8.: Zonages du modèle Quasi-Modus et du modèle du Stif employé dans l'étude du projet de tram-train. *Source* : DRIEA/SCEP/DPAT et STIF.

Un rattachement de chaque centroïde du modèle du Stif au centroïde de Quasi-Modus le plus proche a donc été réalisé, de manière à rendre les matrices issues des deux modèles compatibles. Compte tenu du niveau de détail supérieur du modèle du Stif autour du projet de tram-train, une partie des flux reportés se sont retrouvés sur

la diagonale de la matrice au format Quasi-Modus. Ainsi, sur les 840 déplacements de la matrice de reports, la diagonale contenait 80 déplacements. Ces derniers ne sont pas pris en compte dans l'affectation avec projet. Il s'agit d'une première limitation à l'usage de la méthode non marginaliste lorsque les matrices sont issues de modèles différents.

Par ailleurs, lorsque les flux sur certaines OD sont faibles, et du même ordre que les flux reportés, la matrice avec projet peut comporter des flux négatifs, ce que les modèles d'affectation ne prennent pas bien en compte. Il faut donc adapter, dans la matrice avec projet, dans la matrice des reports ou dans la matrice hors projet, ces flux, de manière qu'aucun flux ne soit négatif. Nous avons choisi, dans le cadre du cas d'étude, de compenser les flux au niveau de la matrice hors projet.

Enfin, les critères de convergence de la procédure d'affectation doivent être choisis de manière qu'en effectuant la différence entre les affectations avec et hors projet, les flux reportés (c'est-à-dire le *signal*) soient plus importants que les artefacts d'oscillation directement liés à l'algorithme employé pour l'affectation (c'est-à-dire le *bruit*). Autrement dit, ces critères doivent être suffisamment sévères. Cela implique un temps de calcul généralement élevé, ce qui constitue une autre limite de la méthode non marginaliste, à moins de disposer d'un algorithme d'affectation à la fois rapidement convergent et fiable (Bar-Gera *et al.*, 2010).

5.A.2 Méthode marginaliste traditionnelle

La méthode marginaliste traditionnelle repose d'une part sur la donnée d'un (ou plusieurs dans le cas d'un intervalle) coefficient, fournissant le coût externe marginal par unité de trafic, et d'autre part sur une évaluation de la distance totale parcourue économisée (ou supplémentaire).

Dans le cas du projet de tram-train, nous avons estimé cette économie en termes de véhicules-kilomètres, à partir d'une recherche de plus court chemin sur le réseau Quasi-Modus chargé à l'heure de pointe du matin, et de la matrice des allègements routiers. Néanmoins, contrairement à la méthode précédente, nous n'avons pas eu à rattacher chaque centroïde du modèle du Stif à un centroïde de Quasi-Modus : il a suffi de connecter les centroïdes du Stif au réseau de Quasi-Modus.

Une fois obtenue la matrice des distances des plus courts chemins à l'HPM, une simple multiplication de chaque élément de cette matrice par l'élément correspondant de la matrice des allègements routiers fournit, après sommation, la distance totale économisée.

Il suffit alors de multiplier cette valeur par le (ou les) coefficients de décongestion retenus.

5.A.3 Méthode marginaliste segmentée

La mise en œuvre de la méthode marginaliste segmentée est identique à la méthode précédente, avec une nuance cependant. Lors de la constitution de la matrice des plus courts chemins, il est nécessaire de distinguer, pour chaque paire OD, les distances à parcourir sur chacun des segments de réseau retenus. Plus précisément, à la place d'une unique matrice de plus court chemin, la méthode marginaliste segmentée exige de disposer d'une matrice par segment, permettant d'obtenir la part du plus court chemin parcourue sur chacun des segments.

Ensuite, une simple multiplication des éléments de ces matrices par les éléments de la matrice des allègements routiers, puis une multiplication des distances totales économisées sur chaque segment par le (ou les) coefficients correspondant à ces segments, fournit les résultats souhaités.

Troisième partie .

**Localisation des ménages et des
emplois et usage des transports : un
modèle d'équilibre urbain**

6

Un modèle d'équilibre urbain à distribution exogène des lieux d'emplois

6.1. Introduction

L'objectif de ce chapitre est de disposer d'un cadre d'analyse permettant d'étudier l'influence de la distribution des emplois sur la localisation résidentielle des ménages et les distances domicile-travail. Le développement de ce modèle nécessitera également — ce sera d'ailleurs l'un des objets principaux de ce chapitre — de prouver l'existence d'un équilibre urbain dans ce cadre théorique, et de mettre en évidence certaines propriétés générales du modèle. À ce titre, ce chapitre reprend en grande partie les travaux présentés dans Breteau (2010) et Breteau et Leurent (2010).

La compréhension de l'organisation interne des villes constitue un enjeu majeur de recherche sur l'urbain. Plusieurs disciplines, en particulier la géographie, la sociologie, ou les sciences régionales dans leur ensemble, s'attachent, avec leurs méthodologies propres, à améliorer la connaissance de la structure urbaine actuelle. L'approche économique, que nous avons choisi, et plus particulièrement la modélisation économique, fournissent des outils de description des phénomènes, de prévision et d'analyse des résultats de politiques publiques. Si de tels modèles ne

permettent pas d'expliquer l'ensemble des phénomènes urbains, ils apportent toutefois un éclairage souvent essentiel sur les mécanismes à l'œuvre, et fournissent certaines clés d'explication.

Parmi les modèles développés en économie urbaine théorique, le modèle monocentrique (Alonso, 1964 ; Muth, 1969 ; Mills, 1972), que nous avons déjà présenté dans le chapitre 1, tient une place centrale, du fait de la simplicité de ses hypothèses d'une part, d'autre part de ses capacités à représenter de manière relativement convaincante les arbitrages des ménages et les localisations résidentielles qui en résultent et, enfin, par les nombreuses possibilités d'extension qu'il autorise.

Ainsi, il a rendu possibles de nombreuses avancées dans la compréhension de la structure urbaine, en particulier sur l'étalement urbain, et il constitue encore aujourd'hui la pierre angulaire de l'économie urbaine théorique. Cependant, le modèle monocentrique standard présente également un certain nombre de limites. Nous en avons évoqué deux principales au chapitre 1. Nous souhaitons ici insister sur la question de la localisation *exogène* et *exclusive* des emplois en un centre *ponctuel*, le CBD. Ainsi, la localisation réelle des emplois ne peut pas être représentée et l'étude des distances domicile-travail parcourues par les ménages à l'aide de ce modèle est par conséquent fortement biaisée. De plus, les phénomènes liés à la déconcentration et à la décentralisation de l'emploi, empiriquement constatées par de nombreux travaux (voir par exemple Cooke, 1983 ; Small et Song, 1994 ; Glaeser et Kahn, 2001) ne peuvent pas y être étudiés. Enfin, les interactions complexes entre localisation résidentielle des ménages et localisation des emplois, en particulier le caractère partiellement endogène de la localisation des emplois par rapport à celle des ménages, ne peuvent pas être traitées dans le cadre du modèle monocentrique.

Des extensions au modèle monocentrique ont été développées pour le rendre capable de prendre en compte la dispersion spatiale des emplois. La grande majorité de ces modèles étudient un équilibre spatial de long terme dans lequel firmes et ménages choisissent simultanément leurs localisations urbaines par l'intermédiaire de la courbe de rente foncière et de la courbe de salaires. Nous avons choisi, pour notre part, de nous placer dans un équilibre de moyen terme, dans lequel les ménages choisissent leur localisation résidentielle en prenant leur lieu d'emploi comme fixé.

Plus précisément, nous considérons une ville où les emplois sont répartis de manière exogène dans une zone étendue où le coût de transport est non nul. À ces hypothèses déjà présentes dans les articles de Sullivan (1983b,a), nous avons ajouté l'homogénéité des ménages en termes de préférences ainsi qu'en revenu brut, et nous supposons que chaque ménage a un lieu d'emploi fixé qui influence sa localisation résidentielle. Nous modélisons ainsi un équilibre de moyen terme, où les ménages participent au marché immobilier tandis que les entreprises sont fixes. Nous avons montré que la position résidentielle, l'utilité du ménage et la surface du logement

augmentent quand le lieu d'emploi s'éloigne du centre, tandis que la rente foncière diminue.

La deuxième section du chapitre propose une revue de littérature succincte consacrée aux différentes extensions du modèle monocentrique visant à décrire la décentralisation de l'emploi. Nous chercherons en particulier à préciser le positionnement de notre modèle par rapport à ces extensions. La troisième section présente le modèle proprement dit, avec ses hypothèses particulières. La quatrième section est consacrée à la résolution de l'équilibre urbain, tandis que la cinquième section présente les propriétés générales de l'équilibre.

6.2. La décentralisation de l'emploi dans le modèle monocentrique

La grande majorité des modèles de nature monocentrique suppose que l'ensemble des emplois est concentré en un point unique au centre de la ville (Alonso, 1964, en particulier), ou que le CBD occupe une surface non nulle fixe de la ville, et que le transport s'y fait à coût nul (voir par exemple Solow, 1972, 1973a ; Dixit, 1973). Dans les deux cas, la distribution des emplois au sein du CBD ne joue en fait aucun rôle, ou bien parce qu'elle n'est pas possible dans un CBD réduit à un point, ou bien parce qu'elle ne modifie aucunement les coûts de transport et donc les localisations résidentielles.

Pourtant, l'hypothèse d'une concentration quasi ponctuelle de l'emploi urbain est fortement réductrice de la dispersion observée des emplois (voir par exemple Giuliano et Small, 1993 ; McMillen et McDonald, 1998 ; Glaeser et Kahn, 2001). En fait, ces modèles n'explorent que l'influence du coût de transport *jusqu'*au CBD, ce dernier étant vu comme un *trou noir* sur lequel on n'a aucune information. Il s'agit donc de modèles purement résidentiels, dans lesquels les caractéristiques des emplois sont laissées de côté.

Pour dépasser ces limites, deux types de modèles ont été développés : des modèles monocentriques à emploi dispersé d'une part, et des modèles polycentriques d'autre part. En reprenant les termes employés par Goffette-Nagot (2000), tandis que les premiers visent à représenter et à mieux comprendre les causes et conséquences de la *décentralisation*, c'est-à-dire le mouvement par lequel une partie de la population et des activités se déplace du centre vers la périphérie des villes, les seconds s'intéressent pour leur part à la *déconcentration*, qui se traduit par une croissance plus rapide des petites agglomérations par rapport aux grandes. Concernant ces derniers, on pourra consulter, par exemple, Anas et Kim (1996). Notre modèle se plaçant dans la lignée de la première catégorie de modèles qui visent à analyser

le phénomène de décentralisation de l'emploi, nous nous concentrons, dans cette section sur ces modèles généralement conçus comme des extensions du modèle monocentrique standard, en autorisant toutes ou une partie seulement des firmes à se localiser en dehors du CBD.

6.2.1. Distribution endogène des emplois

De nombreux auteurs ont cherché à améliorer la description de la localisation urbaine des emplois, en la rendant endogène à leurs modèles. Leur objectif était généralement d'expliquer l'apparition d'un CBD et de rendre endogène la structure monocentrique et centripète des déplacements urbains. À titre d'exemple, Ogawa et Fujita (1980), Imai (1982) et Sullivan (1983b) mettent en œuvre des forces classiques d'agglomération au niveau productif pour expliquer l'apparition d'un CBD, et détailler la localisation des emplois ainsi que le niveau des salaires. Ces modèles mettent en évidence l'apparition d'un gradient de salaires, ceux-ci diminuant avec la distance au centre de la ville. Ce résultat découle d'un arbitrage des firmes entre proximité du centre (ou des autres firmes) et salaire versé aux travailleurs. Ce gradient est effectivement observé (voir par exemple Eberts, 1981 ; McMillen et Singell, 1992 ; Timothy et Wheaton, 2001), même s'il a tendance à diminuer lorsque la dispersion devient importante (Glaeser et Kahn, 2001). Toutefois, compte tenu des difficultés à caractériser précisément les emplois en termes de qualifications, il demeure possible qu'une partie du gradient soit lié aux différences de spécialisation des emplois dans l'espace.

Néanmoins, comme le remarquent Anas *et al.* (1998), ces modèles débouchent fréquemment sur une structure purement monocentrique, c'est-à-dire sans zone d'emplois secondaire et avec un usage exclusif du sol par les entreprises dans la zone d'emplois. En effet, les entreprises ont généralement une préférence plus prononcée que les ménages pour les localisations centrales. Plus précisément, leurs enchères foncières sont généralement plus pentues que celles des ménages, si bien que l'ensemble des emplois se trouve ségrégué au centre de l'agglomération. De même, lorsqu'ils aboutissent à un usage mixte du sol, cela implique qu'entreprises et ménages proposent les mêmes enchères foncières, résultat dont le réalisme n'a pas été empiriquement montré. Autrement dit, les modèles à localisation endogène des entreprises proposent l'alternative suivante : absence totale de mixité entre emplois et résidences d'un côté, mixité impliquant des enchères identiques de la part des ménages et des entreprises.

Wheaton (2004) tente de réconcilier ces alternatives. Il propose un modèle caractérisé par une mixité fonctionnelle en tout point de la ville, obtenue par une hypothèse particulière : le niveau relatif des enchères foncières des ménages et des entreprises détermine la part de surface foncière attribuée à chacune de ces catégo-

ries d'agents. Ce modèle ne souffre pas des limites que nous venons d'évoquer, mais repose sur un mécanisme *ad hoc* de répartition du sol entre entreprises et ménages. De plus, le niveau de dispersion des emplois provient de la force et de la forme des économies d'agglomération en présence. Cette hypothèse, par ailleurs très riche, limite les possibilités d'action du modélisateur sur le modèle : il n'est pas possible d'agir *directement* sur la distribution des emplois.

6.2.2. Distribution exogène des emplois

Certains auteurs, à commencer par Solow (1973b) et White (1976) ont par ailleurs développé des modèles dans lesquels la distribution des emplois est exogène. En effet, la difficulté à résoudre les modèles à localisation endogène de l'emploi les limite souvent à l'étude des arbitrages entre économies d'agglomération et coûts de transport. Au contraire, les modèles à localisation exogène de l'emploi peuvent s'intéresser aux cas de plusieurs types de ménages, aux effets des taxes ou du *zoning* (White, 1999).

Ainsi, le modèle développé par Sullivan dans une série d'articles suppose une séparation fonctionnelle prédéfinie entre zone résidentielle et zone d'emplois (Sullivan, 1983b,a,c). Il se place néanmoins dans une perspective de long terme, où les ménages, les emplois, et les appariements entre ménages et emplois sont en situation d'équilibre. Autrement dit, l'ajustement se fait par la courbe des salaires. De même, dans une contribution majeure, White (1988a) s'attache à analyser les choix de localisation dans un modèle où l'emploi est dispersé, et où les travailleurs sont éventuellement distingués par niveaux de qualification. Son cadre de modélisation est proche de celui de Sullivan (1983b) et là encore, c'est la situation d'équilibre de long terme qui est considérée, dans laquelle les salaires s'adaptent à la localisation des entreprises.

Une critique majeure de ces modèles peut, selon nous, être tirée de Mills et Hamilton (1994). Les auteurs y expliquent que du fait de l'ajustement de la courbe des salaires, « *la forme de la fonction de rente n'est pas affectée par le fait que l'emploi est décentralisé.* »¹ Ce point est à souligner, car il implique également que dans le cadre de ces modèles, la décentralisation de l'emploi n'explique en rien le phénomène d'étalement urbain observé depuis les années 1940 aux États-Unis. Or, il est acquis que la décentralisation de l'emploi a accompagné, voire en partie causé, l'étalement urbain (Huriot et Bourdeau-Lepage, 2009).

Une autre conséquence de cette hypothèse d'ajustement des salaires est que l'ensemble des ménages atteint le même niveau d'utilité à l'équilibre, comme dans le modèle monocentrique classique². Autrement dit, ces modèles, comme tous ceux

1. Traduction de l'auteur.

2. C'est d'ailleurs l'un des critères d'atteinte de l'équilibre retenus.

qui s'appuient sur cette hypothèse d'équilibre de long terme, ne permettent pas simplement de traiter les questions d'équité spatiale, en particulier concernant des politiques de transport ou d'aménagement. En particulier, les politiques d'aménagement liées à la distribution des emplois (taille de la zone d'emplois, densité des emplois, mixité avec la zone résidentielle), qui sont pourtant des outils souvent mis en avant dans le cadre des politiques publiques en milieu urbain, ne peuvent être testées dans le cadre de ces modèles. Ils pourront fournir des résultats sur l'évolution globale du niveau d'utilité des ménages suite à telle ou telle politique, mais ne pourront pas renseigner le modélisateur sur la répartition, *entre* les ménages des gains ou pertes en termes d'utilité.

Enfin, notons également que les échelles de temps des différents acteurs économiques ne sont pas toutes identiques. En particulier, le revenu d'un ménage évolue à un rythme généralement plus important que la localisation de son emploi. Les modèles de long terme repoussent l'ensemble des évolutions, sans distinction, à un horizon infiniment lointain.

6.3. Un modèle pour étudier la structure urbaine

L'ensemble des limites constatées tant du côté du modèle monocentrique standard que du côté de ses extensions à emplois distribués nous encourage à proposer un cadre de modélisation permettant de répondre à la question de la structure urbaine interne avec des hypothèses différentes de celles du modèle monocentrique classique ou de ses extensions habituelles à emplois distribués. En particulier, nous souhaitons, en tant que modélisateur, pouvoir agir de manière simple sur la distribution des emplois, de manière à disposer d'un outil de test de certaines politiques d'aménagement.

Plus précisément, le cadre que nous allons détailler dans les paragraphes suivants permet, en se plaçant dans une perspective d'équilibre de moyen terme, d'aboutir à un traitement différencié de ménages apparemment semblables, et de mettre à l'épreuve les politiques de transport et d'aménagement en termes d'équité spatiale. Le modèle, tel qu'il est présenté dans la suite, ne traite que d'une différenciation continue spatiale des ménages (qui se traduit en terme de revenu), mais nous verrons à la fin de ce chapitre que le modèle est adaptable et reste valable pour d'autres types de différenciation, plus généraux, des ménages (salaire, valeur du temps, etc.)

6.3.1. Des emplois distribués

On considère une ville dans laquelle les entreprises, et donc les emplois, sont réparties autour d'un centre prédéfini³ en un disque de rayon ρ_f . Nous l'appelons CBD ou disque des firmes. Par hypothèse, les emplois sont distribués selon une densité radiale $f(\rho)$, où ρ désigne la distance au centre pour les emplois, supposée non nulle sur l'intervalle $[0, \rho_f]$, et nulle au-delà de ρ_f , engendrant une fonction de répartition $F(\rho) = \int_0^\rho f(u)du$. La fonction F est donc strictement croissante sur $[0, \rho_f]$. Le nombre d'emplois est fixé à N , autrement dit $F(\rho_f) = N$. Fixer le nombre d'emplois revient à se placer dans le cadre théorique d'une *ville fermée*.

Les hypothèses formulées pour la distribution des emplois amènent donc à une localisation de l'emploi urbain entièrement déterminée de manière exogène. Il s'agit là d'un choix fort, dont la raison première est la simplicité de traitement que nous recherchions en vue d'obtenir une résolution analytique. La seconde raison vise à faire de cette distribution des emplois un outil des politiques d'aménagement. Plus précisément, en fixant de manière exogène la zone où les entreprises peuvent s'installer, il est possible d'observer par exemple l'effet d'une augmentation de la taille de cette zone, ou encore de la mise en place d'une politique de réduction de la vitesse dans la zone centrale. Ces questions seront l'objet du chapitre 8.

En ce qui concerne le marché foncier des entreprises, nous avons choisi de ne pas le représenter. Ce choix découle de notre volonté de nous focaliser sur la question de la localisation des ménages, et de leurs déplacements domicile-travail. Par ailleurs, nous ne souhaitons ni opter pour un marché entièrement commun, qui a pour conséquence d'entraîner ségrégation fonctionnelle ou égalité des fonctions d'enchère, ni pour un mécanisme *ad hoc* de mixité fonctionnelle. Tout se passe donc, dans notre modèle, comme si les marchés fonciers des ménages et des entreprises étaient entièrement distincts. Compte tenu des réglementations en vigueur en France notamment, qui rendent relativement difficile le changement de statut d'un local (commercial ou résidentiel), une telle hypothèse ne semble pas totalement irréaliste.

6.3.2. Les ménages

En ce qui concerne les ménages, nous supposons :

- (i) que les ménages sont homogènes en termes de préférences ;

3. Le modèle ne cherche pas à expliquer l'apparition de ce centre. Le chapitre 1 de ce mémoire présente une revue des modèles du phénomène d'agglomération. Fujita *et al.* (2001) proposent également des modèles expliquant l'apparition de centres urbains à l'échelle régionale.

- (ii) que chaque ménage a un seul membre employé par une entreprise du CBD⁴ ;
- (iii) que chaque ménage retire de cet emploi un salaire Y , indépendant de sa localisation résidentielle et professionnelle⁵ ;
- (iv) les ménages sont différenciés par leur lieu d'emploi supposé fixe.

Le dernier point est central dans notre modèle. Exprimé différemment, il signifie que les ménages ont un lieu d'emploi fixe, et qu'ils choisissent leur lieu de résidence en fonction de ce lieu d'emploi. À la question soulevée par de nombreux travaux (Steinnes, 1982 ; Boarnet, 1994, par exemple) : *do "people follow jobs" or "jobs follow people" ?*, nous avons opté, dans notre modèle, pour la première alternative. Nous nous inscrivons, de ce point de vue, dans la tradition monocentrique, qui, selon Boarnet (1994) se caractérise par l'hypothèse suivante : la localisation des ménages est principalement déterminée par le coût de transport vers un lieu d'emploi localisé de manière exogène. Nous avons, dans le chapitre 1, indiqué quelques éléments permettant d'appuyer cette hypothèse.

Le point (iii) implique que le marché du travail n'est pas explicitement représenté dans le modèle. Cela se manifeste par le fait que les ménages touchent le même revenu brut. On peut l'analyser en considérant un marché du travail fonctionnant de la manière suivante⁶ : du côté de l'offre, les ménages sont tous représentés par un syndicat qui négocie pour eux les salaires, en exigeant des employeurs qu'ils paient tous les travailleurs, indifférenciables par leurs préférences ou leur qualification, de la même manière ; du côté de la demande, les entreprises ne peuvent que s'aligner sur les exigences du syndicat.

Par ailleurs, comme l'indiquent Van Ommeren et Gutiérrez-i-Puigarnau (2010), les employeurs n'ont pas toujours la capacité d'observer les coûts de transport subis par les travailleurs. Les auteurs donnent deux raisons à cela : d'une part s'ils peuvent éventuellement observer les caractéristiques physiques des déplacements, ils ne peuvent pas connaître la manière dont les travailleurs valorisent ces déplacements ; d'autre part, les employeurs n'ont pas nécessairement les moyens légaux

4. Par conséquent, nous emploierons, dans la suite, indifféremment les termes *ménage* et *travailleur*.

5. Les entreprises sont donc supposées indifférentes à leur localisation et à celle de leurs employés. Cette hypothèse est fortement réductrice de la distribution statistique des salaires puisque, comme nous l'avons déjà évoqué, des travaux empiriques ont montré l'existence d'un gradient intra-urbain des salaires (Eberts, 1981 ; McMillen et Singell, 1992 ; Timothy et Wheaton, 2001, par exemple), même si Glaeser et Kahn (2001) ont montré que la déconcentration de l'emploi semble propice à une uniformisation des salaires. Toutefois, il est possible d'adapter notre modèle pour rendre compte de l'existence d'un gradient de salaire, fixé de manière exogène : l'essentiel des développements reste valable dans ce cadre.

6. Nous sommes ici redevable envers Jacques-François Thisse et Masahisa Fujita qui, par leurs remarques lors d'une conférence à l'Université de Lille 3 en juin 2010, nous ont suggéré cette possibilité d'interprétation.

d'adapter le salaire des travailleurs en fonction de leur coût de transport une fois qu'ils ont été recrutés.

Comme nous l'avons déjà évoqué, il s'agit surtout de permettre la discussion des questions d'équité spatiale, que nous verrons dans le chapitre 8. En d'autres termes, de manière similaire à Thisse (2007, p. 373), ce choix est guidé par la volonté de se concentrer sur la dimension spatiale de l'effet de la décentralisation des emplois.

Le choix d'une localisation résidentielle en r par un ménage dont le lieu d'emploi est ρ lui laisse un revenu net $I = Y - T(\rho, r)$ où $T(\cdot, \cdot)$ est le coût de transport, qui dépend donc à la fois du lieu de résidence et du lieu d'emploi du ménage.

L'utilité U d'un ménage dépend de la surface de son logement s , et de la quantité z d'un bien de consommation composite pris comme numéraire. La fonction U est supposée croissante et de classe \mathcal{C}^2 en chacune de ses variables. D'une façon générale, chaque ménage est considéré comme un décideur rationnel, qui cherche à maximiser son niveau d'utilité sous contrainte de budget.

6.3.3. Structuration spatiale de la zone résidentielle

Nous supposons que la zone résidentielle dans laquelle les ménages se localisent est un anneau autour du centre de la ville, s'étendant sur l'intervalle $[r_0, r_f]$. Cet anneau peut donc être un disque si $r_0 = 0$. Dans ce dernier cas, il y a mixité fonctionnelle sur l'ensemble de la zone d'emplois. Dans les autres cas ($r_0 > 0$), une zone autour du centre de la ville est exclusivement réservée aux emplois, entourée d'une zone mixte, elle-même entourée d'une zone exclusivement résidentielle. La zone mixte peut éventuellement ne pas exister. Dans ce cas, il y a parfaite séparation fonctionnelle. Dans tous les cas, la zone résidentielle s'étend plus loin du centre que la zone d'emplois.

En effet, la localisation centrale des entreprises peut s'expliquer à l'aide d'un argument simple : si les entreprises ont un intérêt clair à se rapprocher du centre (elles se rapprochent du lieu où les forces d'agglomération sont les plus fortes), les travailleurs n'ont eux d'intérêt que pour la proximité de l'entreprise pour laquelle ils travaillent. Un schéma où les ménages seraient au centre de la ville, et les entreprises à l'extérieur n'est donc pas une situation d'équilibre à l'échelle urbaine, puisque la situation des firmes y est moins favorable, tandis que celle des ménages est globalement inchangée.

Comme nous l'avons indiqué dans le paragraphe consacré aux emplois, seul le marché foncier des ménages nous intéresse ici, et le prix unitaire du sol en r , ou rente foncière, dans la zone résidentielle, est noté $R(r)$. Par ailleurs, le coût d'opportunité du sol, correspondant par exemple à un usage agricole, est noté R_A .

De manière à tenir compte de la capacité foncière dans la zone résidentielle, traduisant à la fois la disponibilité des terrains pour un usage résidentiel, et l'existence d'immeubles à plusieurs niveaux qui augmentent artificiellement la surface disponible pour les ménages, nous noterons $L(r)$ cette capacité foncière en r . Autrement dit, entre r et $r + dr$, la surface offerte pour un usage résidentiel est fournie par $L(r)dr$.

Enfin, le coût de transport $T(\rho, r)$, supposé différentiable en chacune de ses variables, est décroissant en ρ et croissant en r . Ces hypothèses de croissance supposent qu'il n'y a pas de déplacements centrifuges (« trajets inverses »). Elles sont donc équivalentes à l'hypothèse de Wheaton (2004), imposant aux déplacements d'être tous centripètes. Toutefois, il s'avère que ces hypothèses sont, dans notre cadre de modélisation, très naturelles, puisqu'elles rendent simplement compte du fait que, comme nous l'avons signalé plus haut, les ménages n'ont aucun intérêt à se concentrer plus que les emplois. Autrement dit, dans le cas extrême, les ménages résident à l'endroit même où ils travaillent. Dans la réalité, bien sûr, les déplacements centrifuges existent⁷. Les ménages peuvent en effet avoir une préférence pour le centre, du fait de la présence d'aménités particulières. De plus, dans un contexte polycentrique dominé par un centre plus important, on observe de très nombreux déplacements centrifuges (par rapport au centre prédominant) qui sont centripètes par rapport à un centre de moindre importance. Enfin, étant donné que les ménages ne changent pas nécessairement de localisation résidentielle lorsque leur emploi se délocalise, du fait d'une préférence pour le centre et de l'existence de coûts de mutation, certains auteurs ont observé que le nombre de déplacements centrifuges a tendance à augmenter avec la décentralisation des emplois (Aguiléra *et al.*, 2009, pour l'Île-de-France par exemple).

6.4. Localisation d'équilibre des ménages et équilibre urbain

Les paragraphes précédents ont permis de poser les hypothèses essentielles du modèle. D'autres seront nécessaires, mais nous ne les détaillerons que lorsqu'elles le deviendront, de manière à conserver le plus longtemps possible le plus haut niveau de généralité.

7. Boiteux et Huriot (2000) notent néanmoins qu'en France en 1990, 4,5 millions d'actifs effectuaient chaque jour un trajet en direction du centre-ville. De même, 1,5 millions d'actifs sub-urbains, soit 52 % des résidents des couronnes périphériques, effectuaient un trajet journalier en direction des pôles urbains, dont 34 % se rendaient dans le centre-ville, alors que les trajets inverses ne concernaient que 300 000 personnes.

6.4.1. Localisation d'équilibre des ménages

6.4.1.1. Programme des ménages et enchère foncière

Le programme économique du ménage employé en ρ s'exprime sous la forme suivante :

$$\max_{r,s,z} U(s, z) \quad \text{s. c.} \quad z + R(r)s \leq Y - T(\rho, r) \quad (6.1)$$

Pour analyser l'occupation du sol par les ménages, on note $h(r)$ la densité des ménages en r . Il s'agit de la principale variable endogène du modèle. Par hypothèse, $h(r) = 0$ pour $r \in [0, r_0[$. r_0 marque donc le début de la zone résidentielle (éventuellement mixte). On peut montrer, de la même manière que Fujita (1989, ch. 2), que si la capacité foncière $L(r) > 0$ sur l'intervalle $[r_0, r_f]$ (ou plus généralement pour $r > r_0$), alors $h(r) > 0$ sur ce même intervalle, limité par r_f , rayon de la ville, en lequel la rente foncière égale la rente agricole. La fonction de répartition des ménages, définie par $H(r) = \int_{r_0}^r h(u)du$, est donc strictement croissante sous cette hypothèse de disponibilité foncière.

Pour analyser le choix de localisation des ménages, on définit la fonction d'enchère foncière (Von Thünen, 1826 ; Alonso, 1964) d'un ménage, $\Psi_\rho(r, u)$ comme la rente unitaire maximale qu'un ménage travaillant en ρ accepte de payer pour résider en r tout en atteignant le niveau d'utilité u . De manière formelle, cette fonction d'enchère foncière s'écrit donc :

$$\Psi_\rho(r, u) = \max_{s,z} \left\{ \frac{Y - T(\rho, r) - z}{s} \mid U(s, z) = u \right\} \quad (6.2)$$

Cette formulation, à ρ fixé, possède les mêmes propriétés de continuité que dans le modèle classique. De même, Ψ_ρ est décroissante en r et en u . La surface de logement $S_\rho(r, u)$ à l'enchère maximale, solution pour s du problème d'optimisation précédent, est croissante en r et en u .

Dépendant de la rente foncière R et du revenu disponible $I_\rho = Y - T(\rho, r)$, nous définissons également la fonction d'utilité indirecte d'un ménage (Solow, 1972), qui joue un rôle fondamental dans la suite, car elle ne dépend pas directement de ρ :

$$V(R, I_\rho) = \max_{s,z} \left\{ U(s, z) \mid z + Rs \leq I_\rho \right\} \quad (6.3)$$

$V(R, I_\rho)$ est continue en R et I_ρ , croissante en I_ρ et décroissante en R .

6.4.1.2. Choix de localisation des ménages et distribution des emplois

Les principales variables et fonctions permettant de déterminer les choix de localisation des ménages ayant été fixées, nous pouvons établir les propriétés de la

localisation d'équilibre des ménages. Ainsi, la dernière propriété du paragraphe précédent induit la suivante :

Proposition 6.1 (Utilité indirecte et rente foncière).

$$\forall \rho, \forall r, \quad V(R(r), I_\rho) \geq V(\Psi_\rho(r, u), I_\rho) \quad \text{comme} \quad R(r) \leq \Psi_\rho(r, u) \quad (6.4)$$

Étant donné que $V(\Psi_\rho(r, u), I_\rho) = u$, en utilisant la propriété précédente, on obtient une règle inspirée de Fujita (1989, ch. 2), permettant d'établir la localisation d'équilibre d'un ménage employé en ρ :

Règle 6.2 (Sur la localisation d'équilibre). La fonction de rente foncière $R(r)$ étant donnée, un ménage employé en ρ optimise son niveau d'utilité \hat{u}_ρ en un lieu de résidence \hat{r}_ρ si et seulement si :

$$R(\hat{r}_\rho) = \Psi_\rho(\hat{r}_\rho, \hat{u}_\rho) \quad \text{et} \quad R(r) \geq \Psi_\rho(r, \hat{u}_\rho), \quad \forall r$$

Si R et Ψ_ρ sont dérivables en \hat{r}_ρ , ce que nous supposons dans toute la suite, alors la Règle 6.2 impose la tangence des deux courbes en ce point. En appliquant le théorème de l'enveloppe à la fonction d'enchère foncière (6.2), on en déduit la relation de Muth :

$$R'(\hat{r}_\rho) = -\frac{1}{S_\rho(\hat{r}_\rho, \hat{u}_\rho)} \frac{\partial T(\rho, \hat{r}_\rho)}{\partial r} \quad (6.5)$$

Cette analyse, menée pour un ménage particulier, permet en réalité de comparer deux ménages indexés par $i \in \{1, 2\}$ ayant des lieux d'emploi différents, comme l'illustre la Figure 6.1, où nous avons également représenté la partie symétrique de chaque courbe d'enchère.

En Annexe 6.A.1 où figure la démonstration de la Proposition 6.3, nous commençons par montrer que si $\rho_1 < \rho_2$, alors les courbes d'enchère Ψ_{ρ_1} du ménage 1 sont plus pentues que celles, Ψ_{ρ_2} , du ménage 2. Donc, la Règle 2.3 de Fujita (1989, p. 28)⁸ s'applique et produit le résultat suivant :

Proposition 6.3 (Sur l'ordre des ménages dans la zone résidentielle). *Si deux ménages indexés 1 et 2 sont tels que $\rho_1 < \rho_2$ alors $\hat{r}_{\rho_1} < \hat{r}_{\rho_2}$: les ménages sont ordonnés, dans la zone résidentielle, comme leurs emplois respectifs dans la zone d'emplois.*

On notera que la démonstration de la Proposition 6.3 nécessite une hypothèse sur le coût de transport : $\partial T / \partial r$ doit être constant ou décroissant par rapport à

8. La règle est la suivante : Si la fonction d'enchère d'un ménage i est plus pentue que celle d'un ménage j , la localisation d'équilibre du ménage i est plus proche du centre de la ville que celle du ménage j .

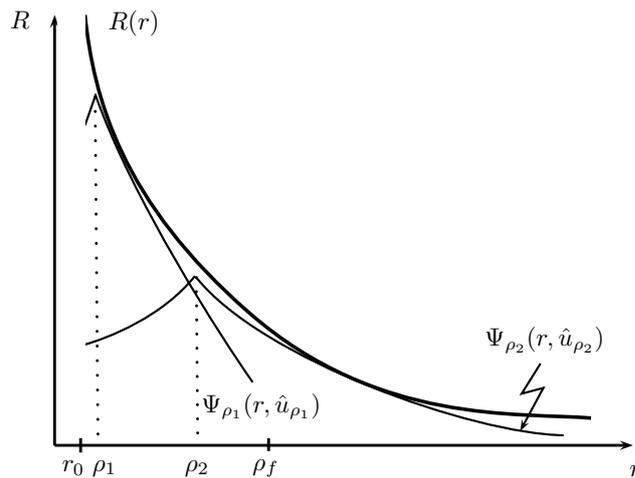


FIGURE 6.1.: Courbes d'enchère de deux ménages employés dans des lieux différents.

ρ . La localisation de l'emploi d'un travailleur n'ayant pas, *a priori*, de raison de modifier le coût marginal de transport en r , cette hypothèse est raisonnable.

Signalons toutefois qu'un coût marginal décroissant avec ρ peut également être envisageable, en se plaçant dans un cadre multimodal. On peut en effet supposer, par exemple, que les ménages ayant leur emploi au centre de la zone d'emplois ont tendance à utiliser un mode de transport lent ou coûteux (marche, bus, etc.), du fait de l'indisponibilité d'infrastructures de transport lourdes et rapides au centre de la zone. Au contraire, les ménages travaillant en périphérie de la zone d'emplois peuvent utiliser des modes plus rapides (voiture, train, etc.) Dans ce cas, le coût marginal de transport, en son lieu de résidence, d'un ménage employé à proximité du centre sera plus élevé que celui d'un ménage employé plus loin du centre.

Cette propriété sur l'ordre des ménages est semblable à la propriété d'absence de *cross-commuting* établie pour le modèle Ogawa et Fujita (1980). Autrement dit, aucun travailleur ne peut avoir à la fois un lieu d'emploi plus proche du centre et un lieu de résidence plus éloigné du centre qu'un autre ménage. L'implication contraposée de la Proposition 6.3 en découle directement :

Corollaire 6.4. *Tous les ménages résidant en r travaillent au même endroit.*

Ces résultats impliquent également qu'à l'équilibre, le *numéro* d'un ménage dans l'ordre des emplois, autrement dit $F(\rho)$, coïncide de manière univoque avec son numéro dans l'ordre des résidences, $H(r)$. Nous pouvons donc définir la fonction r_ω , qui à un lieu d'emploi ρ associe le lieu de résidence r du ménage correspondant.

Formellement, r_ω est définie de la manière suivante :

$$\begin{aligned} [0, \rho_f] &\rightarrow [r_0, r_f] \\ \rho &\mapsto r_\omega(\rho) \quad \text{tel que} \quad H(r_\omega(\rho)) = F(\rho) \end{aligned}$$

Autrement dit, les fonctions H , F et r_ω sont liées par la relation suivante :

$$H = F \circ r_\omega^{-1} \quad (6.6)$$

Nous appelons r_ω la *fonction de transition travail-domicile*, et sa réciproque $\rho_\omega = r_\omega^{-1}$ la *fonction de transition domicile-travail*. Par composition des fonctions croissantes, ces deux fonctions sont croissantes.

6.4.2. Équilibre urbain

Ayant posé les principales caractéristiques du modèle et établi certaines propriétés de la localisation des ménages, intéressons-nous à l'équilibre urbain, c'est-à-dire à l'équilibre du système de localisation des ménages. Il s'agit de déterminer conjointement, du côté de la demande, la localisation des ménages exprimée par la fonction de densité $h(r)$ ou sa primitive $H(r)$, et du côté de l'offre la rente foncière $R(r)$.

Notre objectif est de définir l'équilibre urbain sous une forme mathématique convenable afin de démontrer l'existence et l'unicité de l'équilibre. Comme nous allons le voir, la forme adéquate est un système différentiel qui relie la densité résidentielle et la rente foncière à la localisation résidentielle.

6.4.2.1. Le marché de la localisation

Demande de localisation Du côté de la demande, à fonction de rente foncière $R(r)$ fixée, un ménage employé en ρ choisit la localisation résidentielle r qui maximise son utilité. On réécrit la fonction d'utilité indirecte sous la forme d'une fonction W définie de la manière suivante :

$$W(r, \rho) = V(R(r), Y - T(\rho, r)) \quad (6.7)$$

Le programme micro-économique du demandeur de logement est donc :

$$\max_r W(r, \rho) \quad (6.8)$$

En notant \hat{r}_ρ la localisation résultant de l'optimisation du ménage, \hat{u}_ρ le niveau d'utilité à l'optimum du ménage, et \hat{s}_ρ la surface du logement obtenue dans ces conditions par le ménage, si la distribution de l'ensemble des demandeurs est opti-

male, alors nécessairement, $H(\hat{r}_\rho) = F(\rho)$, et par conséquent :

$$\hat{r}_\rho = H^{-1} \circ F(\rho) = r_\omega(\rho) \quad (6.9)$$

Offre de localisation Du côté de l'offre, à distribution des demandeurs $H(r)$ fixée, et donc à densité radiale $h(r) \geq 0$ donnée, le prix unitaire local atteint en tout point la valeur maximale qu'un demandeur de logement accepte de payer ou, à défaut, le prix agricole R_A . La condition micro-économique associée s'écrit :

$$R(r) = \max \{R_A, \max_\rho \Psi_\rho(r, \hat{u}_\rho)\} \quad (6.10)$$

De plus, au-delà du rayon limite r_f tel que $H(r_f) = N$, la rente foncière est égale à la rente agricole, autrement dit :

$$\forall r \geq r_f, R(r) = R_A = \Psi_{\rho_f}(r_f, \hat{u}_{\rho_f}) \quad (6.11)$$

Enfin, en toute position $r \geq r_0$ telle que $R(r) > R_A$, la capacité foncière radiale $L(r)$ est nécessairement saturée, si bien que :

$$h(r)\hat{s}_\rho = L(r) \quad \text{si} \quad r = H^{-1} \circ F(\rho) = r_\omega(\rho) \quad (6.12)$$

Définitions de l'équilibre urbain Les conditions micro-économiques déterminant la demande et l'offre de localisation, ainsi que les conditions aux limites et de saturation peuvent être rassemblées en une définition formelle de l'équilibre urbain :

Définition 6.5 (Équilibre urbain). L'équilibre urbain est un couple de fonctions (H, R) sur $Z_H = [r_0, +\infty[$ tel que H croît de 0 à N jusqu'en $r_f \geq r_0$ puis reste constante, et vérifie les conditions (6.8), (6.9), (6.10), (6.11) et (6.12).

Il est utile de déterminer une définition alternative de l'équilibre urbain ne faisant référence, pour la condition de demande, qu'à l'ensemble des localisations résidentielles Z_H . En effet, en notant $Z_F = [0, \rho_f[$, la condition (6.8) revient à :

$$W(r_\omega(\rho), \rho) \geq W(r', \rho), \quad \forall (r', \rho) \in Z_H \times Z_F \quad (6.13)$$

En situation d'équilibre, la fonction $r_\omega = H^{-1} \circ F$ est croissante, comme composée de deux fonctions croissantes. L'application réciproque ρ_ω l'est également. Fixons $\rho \in Z_F$ et $r' = r_\omega(\rho)$, donc $\rho = \rho_\omega(r')$. La condition (6.13) implique :

$$W(r', \rho_\omega(r')) \geq W(r, \rho_\omega(r')), \quad \forall (r, r') \in Z_H \times r_\omega(Z_F) \quad (6.14)$$

On aboutit donc à une définition alternative :

Définition 6.6 (Équilibre offre-demande du système de localisation). L'équilibre urbain est un couple de fonctions (H, R) sur Z_H tel que H croît de 0 à N jusqu'en $r_f \geq r_0$ puis reste constante, et vérifie les conditions (6.14), (6.9), (6.10), (6.11) et (6.12).

Proposition 6.7 (Équivalence des deux définitions de l'équilibre). *Un équilibre urbain est un équilibre offre-demande du système de localisations, et réciproquement.*

Démonstration. Nous avons déduit (6.14) de (6.8), donc le système définissant un équilibre offre-demande est vérifié par un équilibre urbain. Donc un équilibre urbain est un équilibre offre-demande. Réciproquement, pour $\rho \in Z_F$, la condition (6.14) pour $r' = r_\omega(\rho)$ assure (6.13), donc (6.8) : un équilibre offre-demande est un équilibre urbain. \square

6.4.2.2. Système différentiel caractéristique

Soit un équilibre offre-demande tel que les fonctions R et H soient *suffisamment régulières*, c'est-à-dire en pratique de classe \mathcal{C}^1 . Alors, la condition nécessaire d'optimalité au premier ordre associée à (6.14) est :

$$\partial W(r, \rho) / \partial r = 0 \quad \text{au point } \rho = \rho_\omega(r) \quad (6.15)$$

Autrement dit :

$$\dot{R}(r) \frac{\partial V}{\partial R} = \frac{\partial T(\rho_\omega(r), r)}{\partial r} \frac{\partial V}{\partial I} \quad (6.16)$$

en notant $\dot{g}(r)$ la dérivée, par rapport à r , d'une fonction $g(r)$. Cette notation sera conservée tout au long du chapitre.

L'identité de Roy fournit une expression de la demande Marshallienne de surface : $\hat{s}(R, I) = -\frac{\partial V}{\partial R} / \frac{\partial V}{\partial I}$, si bien que la relation précédente se reformule :

$$\dot{R}(r) = -\frac{\partial T(\rho_\omega(r), r)}{\partial r} / \hat{s}(R, I_{\rho, r}) \quad (6.17)$$

Nous retrouvons donc la relation de Muth, nécessairement valable en tout point à l'équilibre urbain.

Du côté de l'offre, (6.12) entraîne que $\hat{s}(R, I) = L(r)/h(r)$ pour $h = \dot{H}$, et donc :

$$\dot{H}(r) = L(r) / \hat{s}(R, I_{\rho, r}) \quad (6.18)$$

Finalement, comme $\rho_\omega = F^{-1} \circ H$, les relations (6.17) et (6.18) constituent un système différentiel en (H, R) :

$$\begin{cases} \dot{R}(r) &= -\frac{\partial T(\rho_\omega(r), r)}{\partial r} / \hat{s}(R, I_{\rho, r}) \\ \dot{H}(r) &= L(r) / \hat{s}(R, I_{\rho, r}) \end{cases} \quad (6.19)$$

avec :

$$\begin{cases} \rho_\omega(r) &= F^{-1} \circ H(r) \\ I_{\rho,r} &= Y - T(\rho_\omega(r), r) = Y - T(F^{-1} \circ H(r), r) \end{cases}$$

Ce système traduisant des conditions nécessaires sur l'équilibre urbain, on en déduit directement la proposition suivante :

Proposition 6.8 (Conditions nécessaires sur l'équilibre urbain). *À l'équilibre urbain, le couple (H, R) vérifie le système différentiel (6.19). De plus, les conditions aux limites sont une valeur initiale $H(r_0) = 0$ pour H et une valeur terminale $R(r_f) = R_A$ en r_f tel que $H(r_f) = N = F(\rho_f)$.*

La proposition suivante fournit, de la même manière, les conditions suffisantes de l'équilibre urbain :

Proposition 6.9 (Conditions suffisantes pour l'équilibre urbain).

- (i) *Si la surface est un bien normal, la condition nécessaire (6.17) sur la demande suffit pour assurer (6.14), c'est-à-dire l'optimalité micro-économique pour le demandeur employé en ρ .*
- (ii) *En supposant de plus (6.11), les conditions (6.19) impliquent la condition suffisante (6.10) d'optimalité micro-économique pour l'offre.*
- (iii) *Sous ces conditions microéconomiques, en toute localisation r telle que $R(r) > R_A$, la capacité foncière $L(r)$ est égale à la demande locale des ménages, autrement dit, la capacité est saturée.*

La preuve de cette proposition figure en Annexe 6.A.2.

6.4.2.3. Existence et unicité de l'équilibre

Le système différentiel (6.17), associé à la condition à la limite (6.11) représente donc les conditions nécessaires et suffisantes pour l'équilibre urbain. Les propriétés de ce système différentiel permettent en outre de prouver l'existence et l'unicité de cet équilibre urbain. La proposition suivante détaille, sous forme de lemme, certaines de ces propriétés utiles par la suite :

Lemme 6.10 (Propriétés du système différentiel).

- (i) *À $R_0 = R(r_0)$ fixé, le système différentiel détermine deux fonctions H et R respectivement croissante et décroissante.*
- (ii) *Pour deux valeurs initiales R_{0_1} et $R_{0_2} > R_{0_1}$, et toutes choses égales par ailleurs, les solutions respectives (H_1, R_1) et (H_2, R_2) du système différentiel satisfont, en tout point r : $R_1 \leq R_2$, $H_1 \leq H_2$. De même, $S_1 \geq S_2$, $\rho_{\omega_1} \leq \rho_{\omega_2}$ et $I_{\rho_{\omega_1}(r),r} \leq I_{\rho_{\omega_2}(r),r}$.*

(iii) On arrête l'intégration du système (6.19) dès que l'état suivant est atteint en un point \bar{r}_f : $R(\bar{r}_f) = R_A$ ou $H(\bar{r}_f) = N$. S'il existe $A > 0$ et $\hat{r} > r_0$ tel que :

$$\forall r \leq \hat{r}, \hat{s}(r) \leq A \quad \text{et} \quad \int_{r_0}^{\hat{r}} L(r) dr \geq AN$$

Alors, nécessairement, l'intégration du système mène à un tel état.

L'hypothèse du point (iii) signifie en fait que les surfaces sont bornées et que la capacité foncière est suffisante. La preuve de ce lemme figure en Annexe 6.A.3.

La proposition suivante établit l'existence et l'unicité de l'équilibre urbain :

Proposition 6.11 (Existence et unicité de l'équilibre).

(i) Il existe une valeur \hat{R}_0 initiale telle que l'intégration du système différentiel (6.19) induit un équilibre urbain.

(ii) Cette valeur est unique, ainsi que l'équilibre urbain résultant.

La preuve figure en Annexe 6.A.4.

En résumé, dans le cadre de modélisation que nous avons défini, l'équilibre urbain existe, il est par ailleurs unique, et caractérisé par le système différentiel (6.19) et la condition à la limite (6.11).

6.4.2.4. Fonction de rente terminale

La preuve, fournie en annexe, de la proposition précédente, fait appel aux flots $\bar{R}(R_0, r)$ et $\bar{H}(R_0, r)$. Il est possible, à partir de ces flots, et en s'appuyant sur les propriétés citées dans la démonstration, de définir la fonction suivante :

$$R_0 \mapsto \hat{R}(R_0) = \bar{R}(R_0, \bar{H}^{-1}(R_0, N))$$

où \bar{H}^{-1} est la réciproque, par rapport à r , de \bar{H} . Autrement dit, cette fonction \hat{R} associe à une valeur initiale de la rente foncière la valeur de la rente foncière au rayon \bar{r}_f tel que $\bar{H}(R_0, \bar{r}_f) = N$. Le Lemme 6.10(i) assure que la fonction \hat{R} est croissante.

Comme la démonstration le montre par des arguments de continuité, l'équilibre urbain est caractérisé par $\hat{R}(\hat{R}_0) = R_A$. Autrement dit, la fonction \hat{R} joue, dans notre modèle un rôle comparable à celui joué par la fonction de rente frontière de Fujita (1989) dans le cadre du modèle monocentrique classique. Nous l'appelons *fonction de rente terminale*.

6.5. Propriétés générales de l'équilibre et conséquences sur la structure urbaine

L'étude de la structure urbaine à l'équilibre passe par l'analyse du système différentiel caractéristique. Compte tenu de son caractère très général, seules quelques propriétés peuvent en être tirées. Nous allons voir, toutefois, que celles-ci apportent un éclairage non négligeable sur la structure urbaine issue du cadre de modélisation que nous avons élaboré ici.

Nous commençons par établir, pour toute valeur initiale R_0 de la rente foncière, que l'utilité maximale (ou indirecte) et la surface du logement sont des fonctions croissantes avec la position résidentielle. Puis, nous analysons la sensibilité de l'équilibre urbain relativement à la population de la ville d'une part, et à la rente alternative (agricole) d'autre part.

6.5.1. Niveaux d'utilité et surface des logements

La proposition suivante établit les variations du niveau d'utilité des ménages et de la surface qu'ils occupent, en fonction de la localisation résidentielle :

Proposition 6.12. *Le système différentiel permet d'établir que :*

- (i) *le niveau d'utilité des ménages augmente avec r ; autrement dit, la fonction $r \mapsto V(R(r), I_{\rho_{\omega}(r),r})$ est croissante ;*
- (ii) *la surface des logements augmente avec r , si la surface est un bien normal ; autrement dit, la fonction $r \mapsto \hat{s}(R(r), I_{\rho_{\omega}(r),r})$ est croissante.*

La preuve de cette proposition figure en Annexe 6.A.5.

Si le point (ii) est compatible avec les résultats du modèle monocentrique standard, le point (i) en revanche, amène à quelques commentaires. En effet, l'obtention d'un niveau d'utilité identique pour l'ensemble des ménages homogènes à l'équilibre est un résultat classique de l'économie urbaine théorique. Et c'est même une condition souvent imposée lors de la recherche de l'optimum (par exemple De Palma *et al.*, 2008), et ce, pour éviter le « paradoxe » de Mirrlees (1972).

La raison pour laquelle notre modèle ne fournit pas un tel résultat est simple : les ménages sont certes homogènes en termes de préférences et de revenu brut, mais ils sont différenciés *par leur lieu d'emploi*. Autrement dit, il s'agit d'un équilibre urbain pour un *continuum* de types de ménages. La distribution des lieux d'emplois est en effet analogue à une distribution des revenus bruts parmi les ménages. Fujita (1985, 1989), par exemple, traitent le cas de plusieurs classes de ménages différenciées par leur salaire, et obtient bien que les deux classes atteignent, à l'équilibre, des niveaux différents d'utilité.

On pourra noter à ce propos que la méthode d'obtention de l'équilibre urbain proposé par Fujita (1985), est fondée sur les fonctions d'enchère frontières⁹ de chaque classe de ménages, et n'est par conséquent utilisable que pour un nombre fini et discret de types de ménages. L'adaptation de la méthode de la rente d'enchère frontière n'est donc pas possible. Au contraire, transformer, dans notre modèle, la distribution des localisations des ménages en distribution des revenus est une tâche relativement aisée. Nous ne le ferons pas ici, de manière à maintenir la cohérence de l'ensemble du mémoire, mais il s'agit clairement d'une piste à explorer. La fonction de rente terminale, que nous avons définie, et qui s'apparente à la fonction d'enchère frontière, peut s'avérer utile pour y parvenir.

Ce résultat met en évidence le fait que l'existence même de l'espace, et donc la nécessité, pour les agents économiques, de se localiser, peut introduire, lorsqu'on ne la prend pas en compte, une forme d'inégalité entre des agents par ailleurs semblables. Plus précisément, dans notre modèle, le syndicat, lorsqu'il choisit de ne considérer que les préférences et la qualification des travailleurs pour exiger un salaire identique pour tous, néglige le fait que les travailleurs ont des lieux d'emploi distincts, et qu'en choisissant leur localisation résidentielle par rapport à leur lieu d'emploi, ils obtiennent des niveaux d'utilité différents.

Un exemple illustrant ce phénomène peut être trouvé dans la méthode de rémunération des fonctionnaires, qui ne dépend que très peu (voire pas du tout) de la localisation de leur emploi, que ce soit à l'échelle de la ville ou entre régions. Or les coûts de transport, et le niveau des prix peuvent être très différents et engendrer des niveaux d'utilité eux-mêmes très différents, alors même que les agents sont *a priori* « indiscernables ».

6.5.2. Influence de la population sur l'équilibre urbain

On peut légitimement s'attendre à ce qu'une croissance de la population de la ville ait, toutes choses égales par ailleurs, une influence sur l'équilibre urbain, et donc sur la structure urbaine résultante. Ainsi, de par la concurrence accrue des ménages pour le sol, on peut s'attendre à ce que la pression foncière augmente dans la ville, et il semble par ailleurs logique que la surface occupée par la ville augmente également. Ces résultats sont résumés par la propriété suivante :

Propriété 6.13. Si le nombre de ménages dans la ville augmente, la rente foncière augmente partout, la densité également, de même que le rayon de la ville. Autrement dit, on obtient une ville plus dense et plus étendue, avec des prix fonciers plus élevés.

9. La fonction d'enchère frontière correspond au lieu des points $(r_f, R_f(r_f))$ définis de la façon suivante : le niveau de la rente foncière tel que l'ensemble de la population de la classe considérée soit localisé dans une zone limitée par r_f est $R_f(r_f)$. On peut montrer que chaque classe de ménages donne lieu à une famille de fonctions d'enchère frontières différente.

Démonstration. On suppose défini l'équilibre urbain pour N_1 ménages par (H_1, R_1) , r_{f_1} et R_{0_1} . On a donc :

$$\begin{aligned} R_1(r_0) &= R_{0_1} \text{ et } R_1(r_{f_1}) = R_A \\ H_1(r_0) &= 0 \text{ et } H(r_{f_1}) = N_1 \end{aligned}$$

Soit $N_2 > N_1$. On cherche (H_2, R_2) , r_{f_2} et R_{0_2} , autrement dit l'équilibre urbain pour N_2 ménages. Avant de pouvoir intégrer le système différentiel, une hypothèse supplémentaire est nécessaire : on suppose que la distribution des ménages dans la situation 2 est identique à celle de la situation 1 jusqu'à ρ_{f_1} , et se prolonge ensuite jusqu'à ρ_{f_2} . Autrement dit, $\forall \rho \in [0, \rho_{f_1}]$, $f_1(\rho) = f_2(\rho)$.

En intégrant le système différentiel à partir de R_{0_1} , mais pour une population N_2 , on obtient :

$$\begin{aligned} R_1(r_{f_1}) &= R_A \quad (\text{inchangé}) \\ H_1(r_{f_1}) &< N_2 \end{aligned}$$

Le Lemme 6.10(ii) entraîne que $R_{0_2} \geq R_{0_1}$, et :

$$\begin{aligned} \forall r, R_2(r) &\geq R_1(r) \\ H_2(r) &\geq H_1(r) \end{aligned}$$

Enfin, on a nécessairement $r_{f_2} \geq r_{f_1}$, car R_A est atteint « plus tard » dans le cas 2, puisque $R_2 \geq R_1$. \square

On remarque que ce résultat est comparable à celui obtenu dans le cadre du modèle monocentrique classique. Autrement dit, ajouter l'hypothèse de distribution des emplois ne modifie pas la manière dont la structure urbaine évolue en cas de croissance démographique. C'est un résultat significatif, car, d'une certaine manière, il valide en partie notre modèle comme extension du modèle monocentrique classique.

6.5.3. Influence de la rente agricole sur l'équilibre urbain

De manière similaire, une augmentation de la rente agricole a de fortes chances, en contraignant plus fortement le marché foncier, d'aboutir à une ville plus compacte, conséquence directe d'une augmentation de la pression foncière. Cela est résumé par la propriété suivante :

Propriété 6.14. Si la rente agricole augmente, la rente foncière augmente partout, la densité également, et le rayon de la ville diminue. Autrement dit, on obtient une ville plus dense et moins étendue, avec des prix fonciers plus élevés.

Démonstration. On fait ici appel à la propriété de croissance de la fonction de rente terminale.

Soient $R_{A_2} > R_{A_1}$. La valeur de la rente initiale qui fournit un équilibre étant donnée par $\hat{R}(R_0) = R_A$ et \hat{R} étant croissante, on a $R_{0_1} \leq R_{0_2}$. Le Lemme 6.10(ii) entraîne que $\forall r, R_1(r) \leq R_2(r)$ et $H_1(r) \leq H_2(r)$. Par conséquent $r_{f_1} \geq r_{f_2}$. \square

Comme dans le cas de la population, on constate que l'influence de la rente agricole sur l'équilibre urbain est similaire à ce qui est tiré du modèle monocentrique classique. Là encore, il s'agit d'une validation partielle de notre modèle comme extension du modèle monocentrique classique.

L'étude de sensibilité, que nous venons de mener, sur les influences respectives de la population et de la rente agricole sur l'équilibre urbain, met en évidence la manière dont les ménages réagissent à une modification des conditions de concurrence pour le sol urbain. En effet, une augmentation de la population correspond à une croissance de la demande, tandis que la hausse de la rente agricole est la traduction d'une raréfaction de l'offre, ou d'une demande accrue des agriculteurs.

En fait, il s'avère que relâcher une des hypothèses majeures du modèle monocentrique classique, à savoir la concentration exclusive des emplois en un point central de la ville, ne modifie pas, de manière qualitative, la sensibilité du modèle à la population urbaine ou au niveau de la rente agricole. En revanche, nous n'avons pas été capables d'étudier, dans le cas général, la sensibilité de l'équilibre urbain à d'autres modifications des paramètres, comme le salaire, le coût de transport et le niveau de dispersion des emplois. Ces analyses de sensibilité seront menées, dans des cas particuliers, dans le chapitre suivant.

6.6. Conclusion

Nous visions, avec ce chapitre, le développement d'un modèle monocentrique étendu, dans lequel la localisation des emplois serait exogène et non simplement limitée à un centre ponctuel. Partant du constat que le modèle monocentrique standard n'offre aucune description de la localisation des emplois, l'objectif était de disposer d'un cadre d'analyse des choix de localisation des ménages dans lequel la description des lieux d'emplois et donc du coût de transport subi par les ménages soit plus proche de la réalité. Plus précisément nous avons voulu rendre compte de leur dispersion, et nous donner la possibilité d'analyser des déplacements domicile-travail non triviaux, pas tous dirigés vers un centre ponctuel. Ainsi la signification même du centre est transformée, puisque dans notre modèle il ne s'agit plus que d'une origine de localisation. Ce modèle nous autorise à poser la question d'une « ville cohérente » (Korsu et Massot, 2006) avec des déplacements domicile-travail

de taille limitée, même si l'agglomération est considérablement étendue. Le choix de ne pas rendre cette localisation endogène au modèle est lié à la volonté de disposer d'un cadre d'étude simple des politiques d'aménagement. Une fois les hypothèses identifiées, il nous a été possible de caractériser l'équilibre urbain et d'en montrer l'existence et l'unicité.

Cet équilibre présente la particularité d'aboutir à des niveaux d'utilité différents selon les ménages, puisque ceux-ci sont différenciés par leur lieu d'emploi. Ce résultat, qui peut paraître surprenant, est en réalité la traduction du fait que notre modèle introduit une différenciation continue des ménages. Si nous avons opté ici pour une différenciation spatiale, le même cadre théorique s'applique à une différenciation en termes de salaire, par exemple. Par ailleurs, nous avons montré que les paramètres tels que la population de la ville et la rente agricole influencent l'équilibre de notre modèle d'une manière semblable au modèle monocentrique standard, ce qui valide en partie ce dernier qualitativement.

Le cadre de modélisation ainsi défini et caractérisé permet d'étudier simplement certaines politiques d'aménagement ou de transport en milieu urbain, d'étudier entre autres, le rôle de l'étalement des emplois sur l'usage des transports ainsi que l'influence de la congestion ou de certaines politiques de transport sur l'étalement urbain. Il constitue donc un outil bien adapté à la suite de nos investigations, présentées dans les chapitres suivants.

Quelques limites et pistes d'amélioration à notre travail sont à signaler. De manière à nous concentrer sur les questions d'aménagement urbain, nous avons choisi de ne pas véritablement représenter le marché du travail. Si ce choix est une limite à notre travail, il met néanmoins en évidence le fait que la localisation spatiale est un élément d'utilité, et qu'elle peut donc introduire une forme d'iniquité parmi des agents autrement indiscernables. De plus, notre modèle permet d'aboutir à un équilibre urbain dans lequel les ménages sont continûment différenciés par la localisation de leur emploi. Il constitue donc, selon nous, un apport théorique par rapport à des modèles à classes de ménages : la différenciation peut se faire, dans notre modèle, de manière continue plutôt que discrète.

Par ailleurs, le modèle ne permet pas, en l'état, de mettre en évidence l'existence des « trajets inverses », c'est-à-dire des déplacements centrifuges, dont certains auteurs (Aguiléra *et al.*, 2009, par exemple) ont observé le développement en lien avec l'étalement des emplois. L'introduction d'une préférence pour des aménités centrales permettrait de rendre compte de ce phénomène dans le cadre de notre modèle.

Annexe 6.A Annexes mathématiques

6.A.1 Preuve de la Proposition 6.3

Démonstration. Pour pouvoir appliquer la règle 2.3 de Fujita (1989), il suffit de montrer que les courbes d'enchère foncière d'un ménage travaillant en ρ_1 sont plus pentues que celles d'un ménage travaillant en ρ_2 , si $\rho_1 < \rho_2$. Soient $\Psi_{\rho_1}(r, u_{\rho_1})$ et $\Psi_{\rho_2}(r, u_{\rho_2})$ deux représentants de la famille des courbes d'enchère d'un ménage travaillant en ρ_1 et en ρ_2 respectivement, avec $\rho_1 < \rho_2$, et dont l'intersection se situe en x , ie :

$$\exists \bar{R} \in \mathbb{R}_+, \quad \Psi_{\rho_1}(x, u_{\rho_1}) = \Psi_{\rho_2}(x, u_{\rho_2}) = \bar{R}$$

Du fait de la décroissance de $T(r, \rho)$ par rapport à ρ , on a :

$$Y - T(x, \rho_1) < Y - T(x, \rho_2)$$

En définissant $\hat{s}(R, I_\rho)$ la demande de sol non compensée, correspondant à la solution du problème des ménages (6.1), on a la relation suivante :

$$S_{\rho_1}(x, u_{\rho_1}) = \hat{s}(\bar{R}, Y - T(x, \rho_1)) < \hat{s}(\bar{R}, Y - T(x, \rho_2)) = S_{\rho_2}(x, u_{\rho_2})$$

car en supposant que le sol est un bien normal, l'effet du revenu sur la demande non compensée est positif.

Enfin, l'application du théorème de l'enveloppe à la fonction d'enchère foncière (6.2) fournit l'égalité :

$$\frac{\partial \Psi_\rho(r, u)}{\partial r} = - \frac{\partial T / \partial r(\rho, r)}{S_\rho(r, u)}$$

En combinant ces deux dernières relations, on obtient, à condition que $\partial T / \partial r$ soit constant ou décroissant par rapport à ρ , en un point r donné :

$$- \frac{\partial \Psi_{\rho_1}(x, u_{\rho_1})}{\partial r} = \frac{\partial T / \partial x(\rho_1, x)}{S_{\rho_1}(x, u_{\rho_1})} > \frac{\partial T / \partial x(\rho_2, x)}{S_{\rho_2}(x, u_{\rho_2})} = - \frac{\partial \Psi_{\rho_2}(x, u_{\rho_2})}{\partial r}$$

Ce qui prouve bien que la pente des courbes d'enchère foncière du ménage travaillant en ρ_1 est supérieure à celle du ménage travaillant en $\rho_2 > \rho_1$. L'utilisation de la règle 2.3 de Fujita (1989) permet de conclure. \square

6.A.2 Preuve de la Proposition 6.9

Démonstration. Commençons par le point (i). Dans l'équation (6.17), comme $\partial T / \partial r \geq 0$ et $\hat{s} \geq 0$, nécessairement $\dot{R} \leq 0$. Par ailleurs, l'égalité (6.17) équivaut à (6.15). À ρ fixé, il nous faut montrer que la fonction $r \mapsto W(r, \rho)$ atteint son maximum en

$r = r_\omega(\rho)$. Par dérivation de $W(r, \rho) = V(R(r), Y - T(\rho, r))$, il vient :

$$\frac{\partial W(r, \rho)}{\partial r} = \dot{R} \frac{\partial V}{\partial R} - \frac{\partial V}{\partial I} \frac{\partial T}{\partial r}$$

On définit la fonction $g^{(\rho)} : r \mapsto \frac{\partial W(r, \rho)}{\partial r} / \frac{\partial V}{\partial I}$. Elle a le même signe que $\frac{\partial W(r, \rho)}{\partial r}$ car $\frac{\partial V}{\partial I} \geq 0$. Or, $g^{(\rho)}(r) = -\dot{R} \hat{s}(R, I) - \frac{\partial T}{\partial r}$ puisque $\hat{s}(R, I) = -\frac{\partial V}{\partial R} / \frac{\partial V}{\partial I}$ par la relation de Roy.

Étant donné que $\rho \mapsto T(\rho, r)$ est une fonction décroissante, alors $\rho \mapsto Y - T(\rho, r) = I_{\rho, r}$ est croissante, de même que $\rho \mapsto \hat{s}(R, I_{\rho, r})$ puisque \hat{s} croît avec le revenu si la surface est un bien normal. De plus, $\dot{R} \leq 0$ ne dépend pas avec ρ . En supposant, comme nous avons déjà été amené à le faire au 6.A.1, que $\rho \mapsto \partial T / \partial r$ est constant ou décroissant, on obtient que la fonction $\rho \mapsto \varphi^{(r)}(\rho) = g^{(\rho)}(r)$ est une fonction croissante de ρ . Or, cette fonction s'annule en $\rho = \rho_\omega(r)$ si (6.17) est vérifiée. On en déduit donc progressivement :

$$\begin{aligned} \varphi^{(r)}(\rho) &\leq \varphi^{(r)}(\rho_\omega(r)) \quad \text{si } \rho \leq \rho_\omega(r) \\ \varphi^{(r)}(\rho) &\leq 0 \quad \text{si } \rho \leq \rho_\omega(r) \\ g^{(\rho)}(r) &\leq 0 \quad \text{si } \rho \leq \rho_\omega(r) \text{ i.e. } r_\omega(\rho) \leq r \\ \frac{\partial W(r, \rho)}{\partial r} &\leq 0 = \frac{\partial W(r, \rho_\omega(r))}{\partial r} \quad \text{si } r_\omega(\rho) \leq r \end{aligned}$$

Ainsi, à ρ fixé, la fonction $r \mapsto \frac{\partial W(r, \rho)}{\partial r}$ est positive si $r \leq r_\omega(\rho)$, puis négative si $r \geq r_\omega(\rho)$. $r \mapsto W(r, \rho)$ croît donc jusqu'en $r_\omega(\rho)$ puis décroît. Cela signifie que $r_\omega(\rho)$ est l'unique argument maximal de W . Il est donc possible de définir sans ambiguïté la fonction d'utilité maximale $\rho \mapsto \hat{u}_\rho = W(r_\omega(\rho), \rho)$ sur Z_F . Autrement dit, la condition nécessaire (6.17) suffit à assurer l'optimalité micro-économique pour le travailleur en ρ .

Passons au (ii). Il reste à montrer que la fonction $r \mapsto R(r)$, obtenue par le système différentiel, coïncide effectivement avec les conditions de marché que les demandeurs s'imposent mutuellement, compte tenu de la capacité foncière offerte. Plus précisément, nous devons montrer que conditionnellement aux niveaux d'utilité $(\hat{u}_\rho)_{\rho \in Z_F}$, $R(r)$ est en tout point égal à une enchère d'un demandeur qui a intérêt à enchérir en r car c'est là qu'il retire une utilité maximale, et que cette enchère est plus élevée que celles des autres demandeurs au même point.

Montrons d'abord que $r \mapsto R(r)$ coïncide avec les enchères des demandeurs travaillant en ρ en leur localisation résidentielle optimale $r_\omega(\rho)$. La condition (6.11) assure l'égalité pour ρ_f , en $r_\omega(\rho_f) = r_f$:

$$R_A = R(r_f) = \Psi_{\rho_f}(r_f, \hat{u}_{\rho_f})$$

Montrons que les dérivées $\dot{R}(r)$ et $d\Psi/dr$ de $r \mapsto \Psi_{\rho_\omega(r)}(r, \hat{u}_{\rho_\omega(r)})$ coïncident en tout point, ce qui assurera l'égalité des fonctions en tout point. Notons $\phi(r, \rho) = \Psi_\rho(r, \hat{u}_\rho)$, $\psi(I, u)$ la fonction d'enchère de Solow et $s(I, u)$ la fonction de Solow pour la demande de surface à l'enchère maximale, pour un revenu net I et un niveau d'utilité u . Comme les demandeurs ont des préférences homogènes, les fonctions ψ et s ne dépendent pas de ρ . Au point d'enchère maximale, on a :

$$\Psi_\rho(r, \hat{u}_\rho) = \frac{Y - T(\rho, r) - Z(s, u)}{s} \quad \text{pour } s = s(I, u)$$

On en déduit $\partial\phi/\partial\rho$ par le théorème de l'enveloppe, à s fixé :

$$\frac{\partial\phi}{\partial\rho} = \frac{1}{s} \left(-\frac{\partial T}{\partial\rho} - \frac{\partial Z}{\partial u} \frac{\partial\hat{u}_\rho}{\partial\rho} \right)$$

Comme $V(R, I) = \max U(s, z)$ sous la contrainte budgétaire $sR + z = I$, on a $V(R, I) = U(s, I - sR)$ en $s = \hat{s}(R, I)$ et par le théorème de l'enveloppe, $\partial V(R, I)/\partial I = \partial U/\partial z$. Par ailleurs, puisque Z est la fonction réciproque de U à s fixée, $\partial Z/\partial u = (\partial U/\partial z)^{-1}$. Par conséquent, $\partial Z/\partial u = (\partial V/\partial I)^{-1}$. Enfin, si la condition (6.17) est satisfaite, et donc également (6.16), alors, par la définition de \hat{u}_ρ il vient que :

$$\frac{\partial\hat{u}_\rho}{\partial\rho} = \frac{\partial V}{\partial I} \Big|_{r_\omega(\rho)} \cdot \frac{\partial I}{\partial\rho} \Big|_{r_\omega(\rho)}$$

En rassemblant les dernières expressions, on obtient :

$$\frac{\partial\phi}{\partial\rho} = -\frac{V_I|_r}{V_R|_r} \left[-\frac{\partial T(\rho, r)}{\partial\rho} - \frac{V_I|_{r_\omega(\rho)}}{V_I|_r} \frac{\partial I}{\partial\rho} \Big|_{r_\omega(\rho)} \right]$$

Or, $\partial I/\partial\rho = -\partial T(\rho, r)/\partial\rho$, donc en $r = r_\omega(\rho)$, $\partial\phi/\partial\rho = 0$.

On considère la fonction $r \mapsto \hat{\phi}(r) = \phi(r, r_\omega(r))$. La dérivée de $\hat{\phi}$ s'écrit, à l'aide de l'expression de ϕ :

$$\begin{aligned} \frac{d\hat{\phi}}{dr} &= \frac{\partial\phi}{\partial r} + \frac{\partial\phi}{\partial\rho} \frac{d\rho_\omega(r)}{dr} \\ &= \frac{\partial\phi}{\partial r} \\ &= -\frac{1}{s} \frac{\partial T}{\partial r} \end{aligned}$$

Cette expression coïncide avec celle de $\dot{R}(r)$ telle que définie par (6.17), ce qui, associée à la condition à la limite (6.11) assure l'égalité des deux fonctions primitives. Autrement dit, $R(r) = \Psi_{\rho_\omega(r)}(r, \hat{u}_{\rho_\omega(r)})$. Ainsi, $R(r)$ fournit, en tout point, une enchère proposée par un demandeur.

Il s'agit de montrer maintenant que ces enchères dépassent les enchères des autres demandeurs. Étant donné que la fonction d'utilité indirecte V décroît en R , un ménage employé en ρ satisfait la Propriété 6.1. Or, $V(\Psi_\rho(r, u), I_{\rho, r}) = u$, donc en particulier $V(\Psi_\rho(r, \hat{u}_\rho), I_{\rho, r}) = \hat{u}_\rho$. Et comme \hat{u}_ρ est le niveau d'utilité maximal pour un travailleur en ρ , R étant donné, on déduit de la Propriété 6.1 que :

$$\forall \rho, \forall r, \quad R(r) \geq \Psi_\rho(r, \hat{u}_\rho)$$

Associée à l'égalité $R(r) = \Psi_{\rho_{\omega(r)}}(r, \hat{u}_{\rho_{\omega(r)}})$, on obtient que :

$$\forall r \in r_\omega(Z_F), \quad R(r) = \max_{\rho \in Z_F} \Psi_\rho(r, \hat{u}_\rho)$$

En résumé, la courbe de rente R obtenue en intégrant le système différentiel (6.19) sous la condition (6.11) définit effectivement les conditions de marché.

Il ne reste alors plus qu'à vérifier le point (iii), autrement dit que tous les ménages trouvent un logement, c'est-à-dire que les conditions micro-économiques sont compatibles avec la capacité foncière, $L(r)$, tant au niveau local qu'au niveau global. En rapportant entre eux les deux termes de (6.19) au point r , on trouve $h(r)s = L(r)$, ce qui assure la compatibilité locale (6.12) aux conditions d'offre R et de demande \hat{u}_ρ . Quant à la compatibilité globale, la condition (6.11) assure que les ménages qui se logent entre r_0 et r_f aux conditions R sont au nombre de N .

Au total, les conditions (6.11) et (6.19), qui incluent la condition (6.9), assurent les conditions (6.8), (6.10) et (6.12), donc elles suffisent pour caractériser (H, R) comme équilibre offre-demande, et donc comme équilibre urbain. \square

6.A.3 Preuve du Lemme 6.10

Démonstration. (i) Connaissant les signes de $\partial T / \partial r \geq 0$, $\hat{s} \geq 0$, $L \geq 0$ et $f \geq 0$, (6.17) fournit directement $\dot{R} \leq 0$, et donc que R décroît avec r , tandis que (6.18) fournit $\dot{H} \geq 0$, et donc H est croissante.

(ii) À l'origine r_0 de la localisation résidentielle, $R_{0_2} \geq R_{0_1}$ avec $I_{0_1} = I_{0_2} = Y - T(0, r_0)$. Donc $S_1(R_{0_1}, I_{0_1}) > S_1(R_{0_1}, I_{0_1}) \geq 0$, induisant $h_2(r_0) > h_1(r_0)$ et $\dot{R}_1(r_0) \geq \dot{R}_2(r_0)$.

On suppose que l'hypothèse relative à l'état courant dans l'énoncé de (ii) est vérifiée jusqu'en r . On considère $r' = r + \varepsilon$, où ε est un incrément marginal. Puisqu'on suppose $S_1|_r \geq S_2|_r$, alors $h_2(r) \geq h_1(r)$, et par intégration, l'écart $H_2 - H_1$ augmente. Par conséquent $\rho_{\omega_1}(r') \leq \rho_{\omega_2}(r')$ puisque $\rho_{\omega_i} = F^{-1} \circ H_i$ avec F^{-1} croissante. Comme $T(\rho, r)$ décroît avec ρ , on en déduit que $T(\rho_{\omega_1}(r'), r') \geq T(\rho_{\omega_2}(r'), r')$, et donc que $I_{\rho_{\omega_1}(r'), r'} \leq I_{\rho_{\omega_2}(r'), r'}$.

Concernant la rente foncière, si $R_1(r) \leq R_2(r)$, montrons, par l'absurde, que $R_1(r') \leq R_2(r')$ même si $\dot{R}_1(r) \geq \dot{R}_2(r)$:

Supposons que $R_1(r') > R_2(r')$. Soit $\hat{r} \in [r, r']$ tel que $R_1(\hat{r}) = R_2(\hat{r})$ ¹⁰. Comme $I_1(\hat{r}) \leq I_2(\hat{r})$, alors $S_1(\hat{r}) \leq S_2(\hat{r})$, si bien qu'entre r et \hat{r} , il existe un \tilde{r} tel que $S_1(R_1(\tilde{r}), I_1(\tilde{r})) = S_2(R_2(\tilde{r}), I_2(\tilde{r}))$. Or, étant donné que $\tilde{r} \leq \hat{r}$, on a $R_1(\tilde{r}) \leq R_2(\tilde{r})$. Compte tenu de l'expression de \dot{R} , cela signifie qu'à partir de \tilde{r} , $\dot{R}_1(r) \leq \dot{R}_2(r)$. R_1 décroît donc plus vite que R_2 , impliquant $R_1(r') \leq R_2(r')$, et donc une contradiction avec l'hypothèse $R_1(r') > R_2(r')$.

Au total, on a donc bien conservation des inégalités le long du chemin d'intégration pour les différentes fonctions R , H , S , ρ_ω et $I_{\rho_\omega, r}$.

(iii) À partir de $R_0 \geq R_A$, puisque R décroît, elle atteint R_A ou reste supérieure à R_A . De son côté, H croît à partir de 0 en r_0 . Il y a donc deux cas :

- (a) R atteint R_A . L'état d'arrêt de l'intégration est toujours atteint : ou bien parce que $R(\bar{r}_f) = R_A$ avant que H n'atteigne N ou bien parce que $H(\bar{r}_f) = N$ avant que R n'atteigne R_A .
- (b) R n'atteint jamais R_A . La condition supplémentaire assure qu'il existe une distance \bar{r}_f où H atteint N .

□

6.A.4 Preuve de la Proposition 6.11

Démonstration. (i) Les équations du système différentiel (6.19) caractérisant H et R , associées aux conditions initiales $H(r_0) = 0$ et $R(r_0) = R_0 \in [R_A, +\infty]$ vérifient les hypothèses du théorème de Cauchy-Lipschitz, puisque $\partial T / \partial r$, \hat{s} et L sont strictement positifs sur un intervalle $[0, r_{max}[$, où r_{max} est tel que $I_{\rho_f, r_{max}} = 0$ (ce qui assure que \hat{s} soit strictement positif), par hypothèse ou par construction.

Par conséquent, en définissant les *flots* $\bar{H}(R_0, r)$ et $\bar{R}(R_0, r)$, le théorème assure que \bar{H} et \bar{R} sont des fonctions continues de leurs deux variables. On définit ensuite $\bar{H}^{-1}(R_0, n)$ la réciproque, par rapport à r , de \bar{H} . \bar{H}^{-1} est continue en n comme réciproque d'une fonction continue d'un intervalle sur un intervalle. La continuité par rapport à R_0 est également assurée. Autrement dit, la fonction $R_0 \mapsto \bar{H}^{-1}(R_0, N)$ est continue. Par continuité de la composée de fonctions continues, $R_0 \mapsto \bar{R}(R_0, \bar{H}^{-1}(R_0, N)) - R_A = \Delta \hat{R}(R_0)$ est continue.

Cette fonction $\Delta \hat{R}$ associe à une valeur initiale R_0 la rente différentielle au rayon limite de la ville. Or, le Lemme 6.10(i) assure la décroissance de R , si bien que :

$$\Delta \hat{R}(R_A) \leq \bar{R}(R_A, r_0) - R_A = 0$$

10. le théorème des valeurs intermédiaires assure qu'il en existe nécessairement un, car la rente foncière est une fonction continue définie sur un intervalle de \mathbb{R} .

Par ailleurs, on peut toujours choisir une valeur R_∞ arbitrairement grande, telle que :

$$\Delta\hat{R}(R_\infty) \geq 0$$

c'est-à-dire telle que la valeur de la rente au rayon limite de la ville soit supérieur à la rente agricole, puisque celle-ci est fixée de manière exogène et n'influence pas le système différentiel.

Du fait de la continuité de $\Delta\hat{R}$ sur $[R_A, +\infty]$, le théorème des valeurs intermédiaires impose l'existence d'une valeur \hat{R}_0 telle que $\Delta\hat{R}(\hat{R}_0) = 0$, autrement dit telle qu'au point r_f où $H(r_f) = N$, on ait également $R(r_f) = R_A$.

(ii) L'unicité provient du maintien des inégalités entre les variables de deux états initiaux (Lemme 6.10(ii)) : si $\hat{R}_0 = R_{01}^{(i)} \leq R_{02}^{(i)}$, alors $H_1^{(i)}(r_{f1}^{(i)}) = N \geq H_2^{(i)}(r_{f1}^{(i)})$ tandis que $R_1^{(i)}(r_{f1}^{(i)}) = R_A \leq R_2^{(i)}(r_{f1}^{(i)})$, ce qui interdit que R_{02} induise un équilibre. \square

6.A.5 Preuve de la Proposition 6.12

Démonstration. (i) La dérivée totale de la fonction $r \mapsto W(r, \rho_\omega(r))$ est :

$$\frac{dW}{dr} = \frac{\partial W}{\partial r} + \frac{\partial W}{\partial \rho} \dot{\rho}_\omega(r)$$

Or, $\partial W/\partial r = 0$ lorsque $\rho = \rho_\omega(r)$ d'après (6.8), tandis que $\dot{\rho}_\omega(r) \geq 0$ puisque ρ_ω est croissante, et $\partial W/\partial \rho = (\partial V/\partial I)(-\partial T/\partial \rho)$ est non négatif comme produit de deux termes non négatifs.

(ii) La dérivée totale de la fonction $r \mapsto I(r, \rho)$ est :

$$\dot{I}(r) = \frac{dI}{dr} = \frac{\partial I}{\partial r} + \frac{\partial I}{\partial \rho} \dot{\rho}_\omega(r)$$

Or, la relation de Muth fournit : $\partial I/\partial r = -\partial T/\partial r = \dot{R}\hat{s}$ où \hat{s} est la demande non compensée pour la surface.

Par ailleurs, la dérivée totale de la fonction $r \mapsto S(r)$ peut s'écrire :

$$\frac{dS}{dr} = \frac{\partial \hat{s}}{\partial R} \dot{R} + \frac{\partial \hat{s}}{\partial I} \dot{I}$$

Et en injectant l'expression de \dot{I} :

$$\frac{dS}{dr} = \left(\frac{\partial \hat{s}}{\partial R} + \hat{s} \frac{\partial \hat{s}}{\partial I} \right) \dot{R} + \frac{\partial \hat{s}}{\partial I} \frac{\partial I}{\partial \rho} \dot{\rho}_\omega$$

L'équation de Slutsky fournit :

$$\frac{\partial \hat{s}}{\partial R} + \hat{s} \frac{\partial \hat{s}}{\partial I} = \frac{\partial \tilde{s}}{\partial R} \leq 0$$

où \tilde{s} est la demande compensée de surface. Donc le premier terme dans l'expression de dS/dr est positif en tant que produit de deux termes négatifs. Le second terme est un produit de trois termes non négatifs, $\hat{\rho}_\omega$, $\partial I/\partial \rho = -\partial T/\partial \rho$ et $\partial \hat{s}/\partial I$ qui est non négatif si la surface est un bien normal. Au total, donc, $dS/dr \geq 0$. \square

7

Résolution analytique et sensibilité du modèle à l'équilibre

7.1. Introduction

L'objectif de ce chapitre est d'étudier, dans un cadre d'analyse simplifié, la sensibilité du modèle présenté et défini au chapitre précédent, afin de mieux saisir l'influence des divers paramètres, et en particulier de ceux pouvant faire l'objet d'une politique de transport ou d'aménagement. Ainsi, nous cherchons à mettre en évidence les variables particulièrement sensibles parmi les différentes variables caractéristiques de l'équilibre urbain, et les facteurs d'influence de ces variables.

Le cadre général présenté au chapitre précédent ne permet pas, en effet, d'analyser la sensibilité de l'ensemble des variables aux variations des paramètres. En revanche, le choix d'un certain nombre d'hypothèses spécifiques permet d'aboutir à une résolution analytique complète, autorisant des analyses de sensibilité. Plus précisément, pour des formes simples de la fonction d'utilité, du coût de transport, de la distribution des emplois et de la capacité foncière, nous obtenons des formes closes pour la localisation résidentielle, la rente foncière, la taille de logement et la densité résidentielle en fonction du lieu d'emploi. Il s'agit d'un premier résultat de ce chapitre

Dans un second temps, nous analysons la sensibilité du modèle à des variations des différents paramètres. Nous montrons ainsi que pour chaque lieu d'emploi, cha-

cune des fonctions évoquées ci-dessus varie de manière monotone en fonction des paramètres caractéristiques de la ville, à savoir le rayon limite d'implantation des firmes, leur densité locale, la capacité immobilière, la rente alternative, le revenu brut, et les arbitrages entre surface de logement et consommation du bien courant.

La première section de ce chapitre présente les hypothèses spécifiques de ce cadre simplifié ainsi que la résolution complète de l'équilibre urbain, tandis que la deuxième section expose les analyses de sensibilité menées à l'équilibre, sur les variables du modèle, en s'efforçant de mettre l'accent sur les implications possibles en termes de politiques de transport et d'aménagement.

7.2. Présentation du cadre d'analyse simplifié et résolution analytique

La définition d'un cadre d'analyse simplifié, à travers la mise en œuvre d'hypothèses spécifiques supplémentaires, nous a permis d'obtenir des expressions analytiques complètes pour les différentes variables du modèle, et par conséquent d'analyser, à l'équilibre urbain, la sensibilité de ces dernières à des variations des paramètres du modèle. Bien que certaines hypothèses soient fortement simplificatrices, le maintien du caractère analytique de nos résultats les justifiait selon nous. Nous précisons, lorsque cela fait sens, les implications de certains de ces choix.

7.2.1. Hypothèses spécifiques

7.2.1.1. Distribution des emplois et capacité foncière

Concernant la localisation des emplois, nous supposons que les N emplois sont uniformément distribués sur un intervalle $[0, \rho_f]$, avec une densité b . Autrement dit :

$$\forall \rho \geq 0, f(\rho) = b \cdot \mathbf{1}_{[0, \rho_f]}(\rho), \quad (7.1)$$

$$F(\rho) = b \min(\rho, \rho_f), \quad (7.2)$$

$$\forall n \geq 0, F^{-1}(n) = \min\left(\frac{n}{b}, \rho_f\right). \quad (7.3)$$

Nous supposons ensuite que la capacité foncière est uniforme à partir d'un rayon r_0 , qui marque donc le début de la zone résidentielle :

$$L(r) = \begin{cases} 0 & \text{pour } r < r_0 \\ \lambda & \text{pour } r \geq r_0 \end{cases}. \quad (7.4)$$

Le choix d'un telle spécification pour la capacité foncière, qui fait de notre ville-modèle une ville « linéaire », peut paraître fortement réducteur de la réalité de la capacité foncière des villes réelles. Notons dans un premier temps que la capacité foncière que nous considérons ici est une capacité qui tiens compte des aménagements effectivement constatés. Ainsi, en ce qui concerne l'Île-de-France, l'aménagement vertical en zone centrale, et la faiblesse des surfaces bâties en périurbain compensent l'élargissement concentrique en $2\pi r$, et font de l'hypothèse de constance de la capacité foncière une option valable et robuste. En réalité, la capacité foncière en Île-de-France peut être relativement bien modélisée par une fonction faiblement croissante, du type $L(r) = \lambda(r/r_0)^k$, avec $\lambda \approx 8$ km, $r_0 \approx 1$ km et $k \ll 1 \approx 0,1$. Nous n'avons pas retenu cette spécification plus précise pour des raisons de solubilité du modèle.

7.2.1.2. Coût de transport

Concernant le coût de transport, nous le supposons fonction affine de r et de ρ . Autrement dit, nous ne prenons pas explicitement en compte la congestion des réseaux de transport de manière *endogène*. Toutefois, une prise en compte *exogène* est possible, *via* des coûts marginaux de transport différents suivant les localisations. Nous avons exploité cette possibilité en choisissant la spécification suivante pour la fonction $T(\rho, r)$:

$$\forall(\rho, r) \in [0, \rho_f] \times [r_0, +\infty], \quad T(\rho, r) = a_0 + ar - a'\rho. \quad (7.5)$$

Cette spécification nous permet de disposer d'un coût marginal de transport différent dans la zone centrale destinée principalement à l'emploi, et dans une couronne occupée par des logements¹. Ainsi, dans le cas extrême où $r_0 = \rho_f$, c'est-à-dire si la zone résidentielle commence là où finit la zone d'emplois, impliquant une séparation fonctionnelle totale, le coût de transport s'écrit pour tous les ménages :

$$T(\rho, r) = \bar{a}_0 + a(r - \rho_f) + a'(\rho_f - \rho) = (\bar{a}_0 + (a' - a)\rho_f) + ar - a'\rho. \quad (7.6)$$

Si $r_0 < \rho_f$, l'équation de coût (7.6) n'est valable, en toute rigueur, que pour les ménages résidant au-delà de ρ_f . Pour les autres, le coût de transport se réduit à $T(\rho, r) = \bar{a}_0 + a'(r - \rho)$. Toutefois, on considérera que le coût de transport fourni par (7.6) reste valable pour tous les ménages, en s'appuyant sur l'argument suivant.

1. Cette spécification peut également rendre compte de la réalité du partage modal, comme indiqué dans SESP (2007), où des valeurs sensiblement distinctes sont avancées pour le coût marginal dans Paris et le coût marginal dans le reste de la région.

Pour un ménage résidant hors de la zone centrale, le coût de transport reste :

$$T(\rho, r) = (\bar{a}_0 + (a' - a)\rho_f) + ar - a'\rho. \quad (7.7)$$

Tandis que pour un ménage résidant à l'intérieur de la zone centrale, il devient :

$$T(\rho, r) = (\bar{a}_0 + (a' - a)r) + ar - a'\rho. \quad (7.8)$$

On peut supposer que les ménages résidant dans le cœur de la ville subissent un coût fixe \bar{a}_0^c plus important que celui subi par ceux qui résident en périphérie, \bar{a}_0^p , en raison de la difficulté de stationnement et de son coût en zone centrale. En admettant de plus que ce coût fixe en zone centrale \bar{a}_0^c croît lorsque r se rapproche de r_0 , on peut supposer² que :

$$\bar{a}_0^c = \bar{a}_0^p + (\rho_f - r)(a' - a). \quad (7.9)$$

Dans ces conditions, l'expression du coût de transport (7.5) est bien valable pour tous les ménages.

L'Annexe 7.A.1 propose une interprétation différente de la forme particulière que nous avons donnée à la fonction de coût de transport, basée sur les comportements de mobilité des ménages.

7.2.1.3. Revenu et fonction d'utilité des ménages

Nous supposons que les N ménages ont chacun un actif, sous la forme d'un revenu brut Y et sont homogènes en préférences, autrement dit qu'ils ont la même fonction d'utilité, de type Cobb-Douglas à deux paramètres $\alpha > 0$ et $\beta > 0$:

$$U(z, s) = U_0 z^\alpha s^\beta. \quad (7.10)$$

Les paramètres d'élasticité réduits $\alpha' = \alpha/(\alpha + \beta)$ et $\beta' = \beta/(\alpha + \beta)$ seront fréquemment employés dans la suite.

Avec la spécification Cobb-Douglas choisie pour la fonction d'utilité, on montre aisément que la fonction d'utilité indirecte, la demande non compensée de surface

2. Cette expression se justifie si l'on considère que dans une agglomération réelle, la différence $a' - a$ va diminuer si, à N constant, ρ_f augmente.

de logement et la fonction d'enchère des ménages à la Solow s'écrivent :

$$V(R, I) = V_0 R^{-\beta} I^{\alpha+\beta} \quad \text{avec } V_0 = U_0 \alpha'^{\alpha} \beta'^{\beta}, \quad (7.11)$$

$$\hat{s}(R, I) = \beta' \frac{I}{R}, \quad (7.12)$$

$$\psi(I, u) = \left(\frac{u}{V_0} \right)^{-1/\beta} I^{1/\beta'}. \quad (7.13)$$

Ces fonctions seront utiles par la suite.

7.2.2. Résolution analytique du système différentiel caractéristique

Dans le cadre des hypothèses que nous venons de présenter, le système différentiel (6.19) s'écrit :

$$\begin{cases} \dot{R}(r) &= -\frac{a}{\lambda} \dot{H}(r) \\ \dot{H}(r) &= \frac{1}{\beta'} \frac{\lambda R(r)}{Y - T(\rho_{\omega}(r), r)} \end{cases} \quad (7.14)$$

La première équation du système s'intègre directement en :

$$R(r) = R_0 - \frac{a}{\lambda} H(r) \quad \text{pour } r \geq r_0. \quad (7.15)$$

La condition à la frontière urbaine est $R_A = R_0 - a/\lambda N$, si bien que $R_0 = R_A + a/\lambda N$. On peut noter à ce stade que la rente à l'origine de la zone résidentielle s'exprime, comme dans le modèle monocentrique standard, de manière très simple en fonction des paramètres du modèle.

La seconde équation du système, pour sa part, peut s'écrire, compte tenu de (7.15) et que $\rho_{\omega}(r) = H(r)/b$ sous la forme d'une équation différentielle autonome sous forme résolue :

$$\dot{H}(r) = \frac{1}{\beta'} \frac{\lambda R_0 - aH(r)}{Y_0 - ar + a'H(r)/b}, \quad (7.16)$$

où $Y_0 = Y - a_0 = Y - \bar{a}_0 - (a' - a)\rho_f = \bar{Y}_0 - (a' - a)\rho_f$.

La résolution complète de cette équation différentielle figure en Annexe 7.B.1. Elle fournit une relation entre la localisation résidentielle r et la rente foncière R :

$$r = A + B \frac{R}{R_0} - C \left(\frac{R}{R_0} \right)^{\beta'}, \quad (7.17)$$

avec :

$$\begin{aligned} A &= \frac{\bar{Y}_0}{a} - \frac{a' - a}{a} \rho_f + \frac{a' \lambda R_0}{a ab} = \frac{Y_0}{a} + \frac{a'}{a} \left(\frac{\lambda R_A}{ab} + \frac{N}{b} \right), \\ B &= \frac{\beta a' \lambda R_0}{\alpha a ab} = \frac{\beta a'}{\alpha a} \left(\frac{\lambda R_A}{ab} + \frac{N}{b} \right), \\ C &= A + B - r_0. \end{aligned}$$

Il s'avère que la forme la plus adaptée pour représenter les différentes variables du modèle à l'équilibre est une forme paramétrique, la localisation des lieux d'emplois ρ jouant le rôle de paramètre central. L'intérêt principal d'une telle formulation est qu'elle permet de s'affranchir, dans un cadre analytique, des hypothèses sur la valeur des élasticités de l'utilité à la consommation de bien et à la surface (coefficients α et β) généralement nécessaires à la résolution analytique du modèle monocentrique, telle qu'elle est menée par Fujita (1989) par exemple. Les fonctions de transition r_ω et ρ_ω prennent ici toute leur importance, en permettant le passage d'une variable à l'autre. De manière à rappeler le caractère paramétrique des relations qui viennent, on ajoutera l'indice ω aux différentes variables. Ainsi, en utilisant l'équation (7.15) et le fait que $H(r_\omega(\rho)) = F(\rho)$, on obtient l'équation de rente foncière en r , paramétrée par ρ :

$$R_\omega(\rho) = R_0 - \frac{ab}{\lambda} \rho \quad (7.18)$$

Avec cette dernière relation, l'équation (7.17) se réécrit de la même manière, en fonction du paramètre ρ :

$$\forall \rho \in [0, \rho_f], \quad r_\omega(\rho) = A + B \left(1 - \frac{\rho}{\tilde{\rho}} \right) - C \left(1 - \frac{\rho}{\tilde{\rho}} \right)^{\beta'} \quad (7.19)$$

où $\tilde{\rho} = \frac{\lambda R_0}{ab}$.

L'analyse de sensibilité que nous allons mener dans la section suivante sera basée sur cette représentation paramétrique en fonction du lieu d'emploi. Néanmoins, l'Annexe 7.B.2 présente un cas particulier pour lequel il est possible de trouver une relation analytique entre R et r , en se passant du paramètre ρ .

7.3. Analyses de sensibilité à l'équilibre urbain

Nous étudierons successivement la sensibilité, aux paramètres du modèle, des variables suivantes : la localisation résidentielle et le rayon limite d'urbanisation, la taille des logements et la densité résidentielle, la rente foncière, les dépenses et

les niveaux d'utilités des ménages. Nous soulignerons, lorsque cela nous a semblé nécessaire, les interprétations des résultats en termes de politiques de transport et d'aménagement, ou de comportement des ménages.

7.3.1. Localisation résidentielle et rayon limite d'urbanisation

La Proposition 7.1 établit les propriétés de sensibilité de r_ω , c'est-à-dire la localisation résidentielle des ménages en fonction d'un lieu d'emploi donné ρ , aux différents paramètres du modèle.

Proposition 7.1 (Sensibilité de la localisation résidentielle). *À lieu d'emploi ρ donné, la localisation résidentielle $r_\omega(\rho)$:*

- (i) *augmente avec le rayon résidentiel initial r_0 ;*
- (ii) *augmente avec le revenu brut Y et diminue avec le coût fixe de transport a_0 ;*
- (iii) *diminue avec a' ;*
- (iv) *décroît avec la capacité foncière λ ;*
- (v) *décroît avec la rente alternative R_A ;*
- (vi) *augmente avec β et décroît avec α ;*

La preuve de cette proposition figure en Annexe 7.B.3.

Nous ne sommes pas parvenu à prouver le sens de variation de la localisation résidentielle avec a . Toutefois, des simulations numériques ont confirmé le résultat du modèle monocentrique classique, à savoir la décroissance de la localisation résidentielle avec le coût marginal de transport dans la zone résidentielle.

La Figure 7.1 présente des courbes de localisation résidentielle, pour deux combinaisons des paramètres a et a' de coûts marginaux de transport. Les valeurs numériques retenues pour cette illustration sont détaillées dans l'Annexe 7.A.2. Elles ont été choisies de manière à faire ressembler la maquette que nous développons ici à l'Île-de-France. On constate que l'influence de a est largement prépondérante sur celle de a' . Cela est principalement dû au fait que a' ne joue un rôle que sur des courtes distances tandis que a joue sur de bien plus grandes distances.

L'influence des différents paramètres est relativement intuitive. Ainsi, on constate qu'augmenter le rayon d'origine de la zone résidentiel repousse l'ensemble des ménages plus loin du centre de la ville. Une hausse du revenu brut (ou une baisse du coût fixe de transport) permet aux ménages de supporter un coût de transport plus élevé pour bénéficier d'une surface de logement plus importante. Augmenter la capacité foncière diminue l'intensité de la concurrence pour le sol, et permet aux ménages de se rapprocher de leur emploi, sans trop y perdre en termes de surface

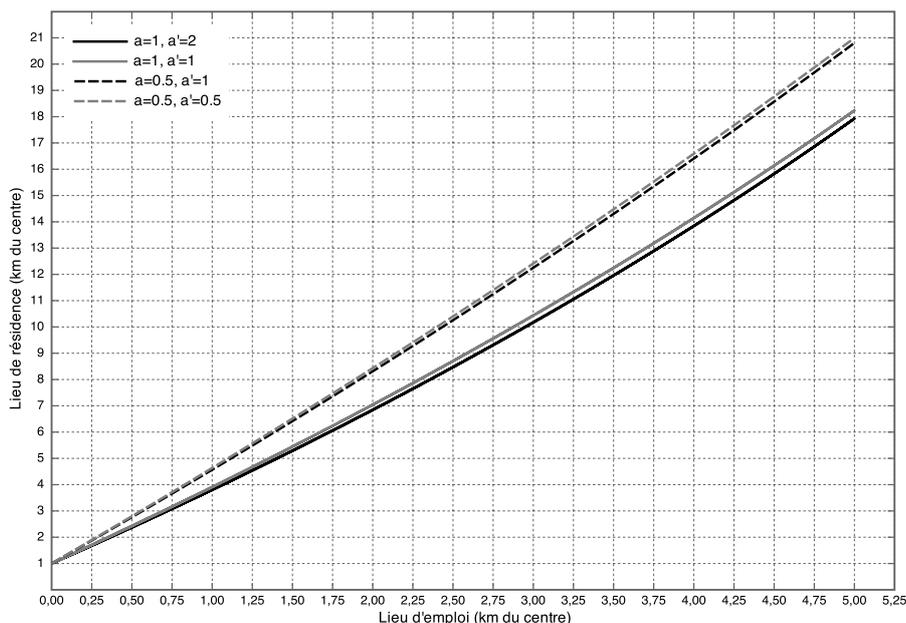


FIGURE 7.1.: Courbes de localisation résidentielle pour deux couples de valeurs de a et a' .

de logement. Le phénomène est inversé pour la rente alternative : lorsqu'elle augmente, le prix du sol augmente partout, si bien que les ménages sont contraints de réduire le coût de transport qu'ils subissent. Enfin, si la préférence des ménages pour la surface, mesurée par β , augmente, les ménages sont disposés à s'éloigner de leur emploi pour bénéficier d'une surface plus importante.

L'impact des paramètres ρ_f ou b doit être à travers une analyse à ménage fixé, et non à lieu d'emploi fixé. En effet, lorsque le paramètre étudié ne modifie pas la distribution des emplois, les variations observées sur la localisation des ménages correspondent bien à une trajectoire des ménages : c'est « le même » ménage qui travaille en ρ pour les différentes valeurs du paramètre étudié. En revanche, lorsque le paramètre étudié modifie la distribution des emplois, le ménage qui travaille en ρ change pour les différentes valeurs du paramètre. Autrement dit, ce qui s'applique à lieu d'emploi fixé ne s'applique pas nécessairement à *ménage* fixé. Par conséquent, la sensibilité du rayon limite d'urbanisation ou de la localisation moyenne des ménages aux paramètres qui modifient la distribution des emplois ne peut pas être directement déduite de la sensibilité de la localisation résidentielle à lieu d'emploi fixé. Il faut, pour traiter ces cas, étudier la sensibilité de la fonction de localisation résidentielle à ménage fixé, autrement dit $r(n)$, où n est une quantité cumulée de ménages. En supposant le nombre de ménages N dans la ville fixé, l'influence, pour un ménage donné, d'une variation de la distribution des emplois, c'est-à-dire de la

variation de b ou de ρ_f , est fournie par la proposition 7.2 suivante :

Proposition 7.2. *Si le nombre total de ménages est fixé à N , la localisation résidentielle $r(n)$ d'un ménage donné numéroté n diminue avec la densité des emplois b , et augmente avec le rayon limite d'emploi ρ_f .*

Nous n'avons pu validé cette proposition que par l'intermédiaire de simulations numériques.

Une conséquence directe de cette proposition est que le rayon limite d'urbanisation, $r_f = r(N)$ diminue avec b et augmente avec ρ_f , à N fixé. Son expression est :

$$r_f = \frac{Y_0}{a} + \frac{a'}{a} \left(\frac{\tilde{\rho}}{\alpha'} - \frac{\beta}{\alpha} \rho_f \right) - \left(\frac{Y_0}{a} + \frac{a'}{a} \frac{\tilde{\rho}}{\alpha'} - r_0 \right) \left(\frac{\lambda R_A}{\lambda R_A + aN} \right)^{\beta'}. \quad (7.20)$$

Par ailleurs, la localisation résidentielle moyenne des ménages est donnée par :

$$\begin{aligned} \bar{r} &= \frac{1}{\rho_f} \int_0^{\rho_f} r_\omega(\rho) d\rho \\ &= A + B \left(1 - \frac{\rho_f}{2\tilde{\rho}} \right) - \frac{C}{\rho_f} \left[- \left(1 - \frac{\rho}{\tilde{\rho}} \right)^{1+\beta'} \frac{\tilde{\rho}}{1+\beta'} \right]_0^{\rho_f} \\ &= \frac{\bar{Y}_0}{a} - \frac{a' - a}{a} \rho_f + \frac{a'}{a} \left(\frac{\tilde{\rho}}{\alpha'} - \frac{\beta}{\alpha} \frac{\rho_f}{2} \right) - \frac{\tilde{\rho}}{\rho_f} \left(\frac{\bar{Y}_0}{a} - \frac{a' - a}{a} \rho_f + \frac{a'}{a} \frac{\tilde{\rho}}{\alpha'} - r_0 \right) \frac{1 - \left(1 - \frac{\rho_f}{\tilde{\rho}} \right)^{1+\beta'}}{1 + \beta'}. \end{aligned} \quad (7.21)$$

Lorsque la densité des emplois diminue ou, ce qui est équivalent, le rayon limite des emplois s'étend, le rayon limite d'urbanisation s'étend également. Autrement dit, nous avons ici mis en évidence un rôle moteur potentiel de la dispersion des emplois sur l'étalement urbain, tel qu'il est évoqué par certains auteurs (Kahn, 2007, par exemple). Toutefois, nous verrons au chapitre suivant, que cet étalement urbain peut être vu comme un étalement « cohérent », qui se distingue, notamment, de celui provoqué par une baisse des coûts de déplacements.

Les Figures 7.2 et 7.3 permettent de constater l'influence de ces deux paramètres, coût unitaire de transport et rayon limite des emplois, sur le rayon limite d'urbanisation et la localisation résidentielle moyenne des ménages. On constate un effet de l'étalement des emplois beaucoup plus limité que celui d'une baisse des coûts de transport, sur l'étalement urbain.

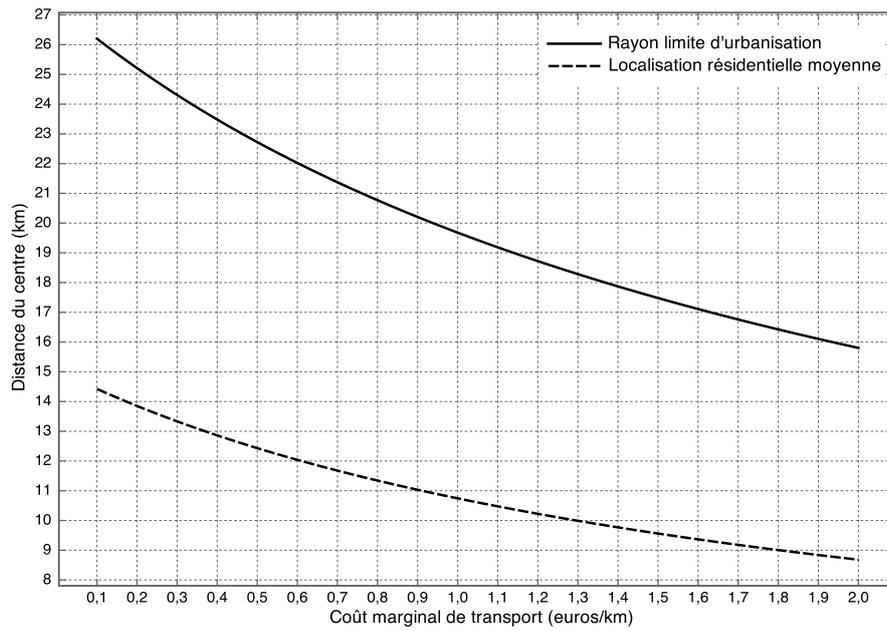


FIGURE 7.2.: Influence du coût marginal de transport sur le rayon d'urbanisation et la localisation résidentielle moyenne.

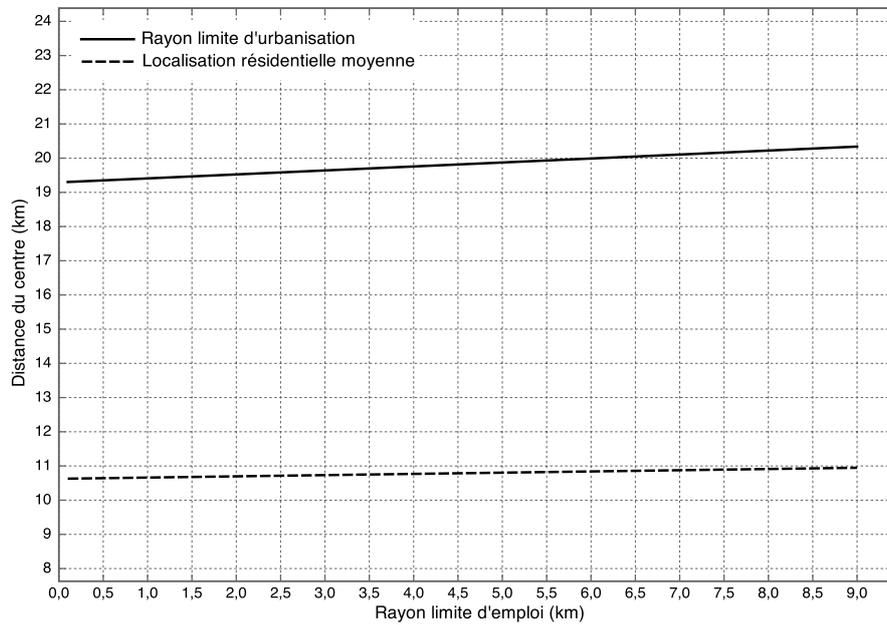


FIGURE 7.3.: Influence du rayon limite d'emploi sur le rayon d'urbanisation et la localisation résidentielle moyenne.

7.3.2. Densité résidentielle et taille des logements

La fonction de transition $r_\omega = H^{-1} \circ F$ donne accès à la densité résidentielle des ménages, en fonction du paramètre ρ également. Plus précisément, $h_\omega(\rho)$, densité résidentielle au point $r_\omega(\rho)$, s'écrit :

$$\begin{aligned} h_\omega(\rho) &= \frac{f(\rho)}{\dot{r}_\omega(\rho)} \\ &= \frac{b\tilde{\rho}}{\beta'C \left(1 - \frac{\rho}{\tilde{\rho}}\right)^{-\alpha'} - B} \end{aligned} \quad (7.22)$$

Compte tenu de son expression, h_ω est décroissant en ρ , donc également en r , puisque r_ω est croissant. On obtient donc que la densité résidentielle diminue avec la distance au centre³. On retrouve ainsi un résultat bien connu du modèle monocentrique classique.

La taille du logement se déduit simplement de la densité résidentielle, grâce à la relation $s(r) = L(r)/h(r)$. Par conséquent, en fonction du paramètre ρ :

$$\begin{aligned} s_\omega(\rho) &= \frac{\lambda}{b} \dot{r}_\omega(\rho) \\ &= \frac{\lambda}{b\tilde{\rho}} \left[\beta'C \left(1 - \frac{\rho}{\tilde{\rho}}\right)^{-\alpha'} - B \right] \end{aligned} \quad (7.23)$$

La surface des logements augmente donc avec la distance au centre.

La proposition suivante détaille l'influence des différents paramètres sur la densité résidentielle et la taille des logements :

Proposition 7.3 (Sensibilité de la densité résidentielle). *À lieu d'emploi ρ donné, la densité résidentielle h_ω par unité de distance radiale :*

- (i) *diminue avec le rayon résidentiel initial r_0 ;*
- (ii) *diminue avec le revenu brut Y et augmente avec le coût fixe de transport a_0 ;*
- (iii) *augmente avec a' ;*
- (iv) *augmente avec la capacité foncière λ ;*

3. Il convient de noter que h_ω désigne la densité *par unité de distance radiale*. Pour obtenir la densité par unité de surface, il suffit de rapporter h_ω à $L(r)$, la capacité foncière. Dans le cas qui nous occupe ici, $L(r)$ est supposée uniforme et égale à λ , si bien qu'à une constante multiplicative près, densité radiale et densité surfacique sont identiques. En revanche, si $L(r)$ est croissante avec r (rendant ainsi compte d'une augmentation de la surface au sol disponible avec l'éloignement au centre), le gradient de densité surfacique sera plus prononcé que le gradient de densité radiale ; à l'inverse, si $L(r)$ est décroissante avec r (rendant ainsi compte d'une prépondérance du rôle joué par l'aménagement vertical en zone centrale sur l'élargissement concentrique de la surface au sol), le gradient de densité surfacique sera moins prononcé que le gradient de densité radiale.

(v) augmente avec la rente alternative R_A ;

(vi) diminue avec β et croît avec α ;

La taille des logements réagit de manière opposée à la densité résidentielle aux variations des différents paramètres, à l'exception de l'influence de λ .

La preuve de cette proposition figure en Annexe 7.B.4.

Comme pour la localisation résidentielle, nous ne sommes pas parvenu à prouver le sens de variation de la densité résidentielle avec a . Toutefois, des simulations numériques ont confirmé le résultat du modèle monocentrique classique, à savoir la croissance de la densité résidentielle avec le coût marginal de transport dans la zone résidentielle.

La Figure 7.4 présente des courbes de surface des logements, pour deux combinaisons des paramètres a et a' .

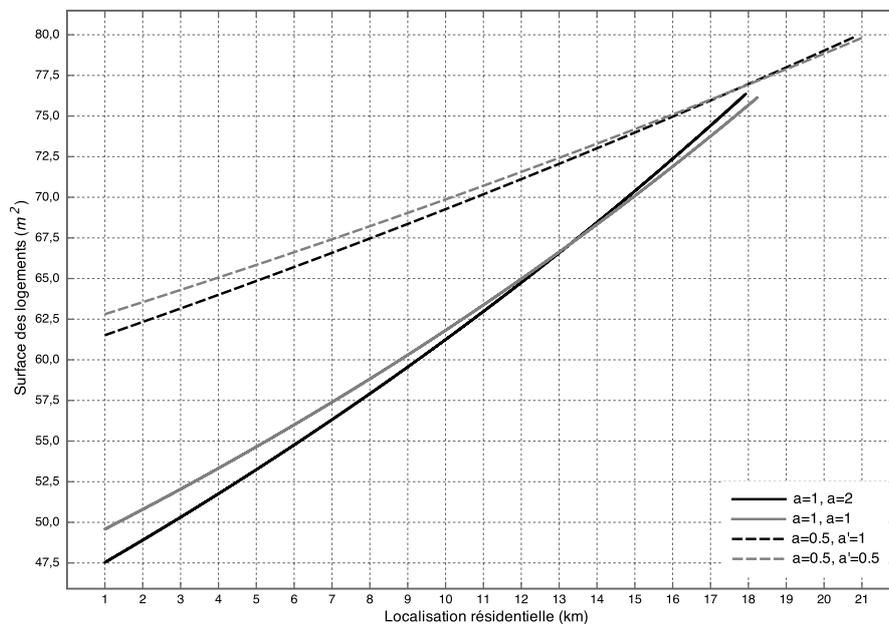


FIGURE 7.4.: Courbes de surface des logements pour deux couples de valeurs de a et a' .

Comme précédemment pour la localisation résidentielle, l'influence de a est prépondérante devant celle de a' .

L'influence de b ou ρ_f sur la densité, ainsi que l'influence de ces deux mêmes paramètres et de λ sur la taille des logements doit être déterminé à ménage fixé, comme pour la localisation résidentielle. En revanche, en supposant la population N de la ville fixée, on peut établir une proposition analogue à la Proposition 7.2 :

Proposition 7.4. *Si le nombre total de ménages est fixé à N , la surface du logement $s(n)$ d'un ménage donné numéroté n diminue avec la densité des emplois b , et augmente avec le rayon limite d'emploi ρ_f .*

Nous n'avons pu valider cette proposition que par l'intermédiaire de simulations numériques.

Par ailleurs, les simulations numériques que nous avons menées montrent que la taille des logements augmente avec la capacité foncière λ . Ce résultat, relativement intuitif, est illustré sur la Figure 7.5 qui présente la variation de la surface moyenne des logements lorsque λ , d'une part, et ρ_f , d'autre part, varient. On constate que l'influence de la capacité foncière est nettement plus marquée que celle du rayon limite des emplois, conséquence du fait que ce paramètre contraint fortement les ménages, les poussant à la fois à réduire leur consommation d'espace et à s'éloigner du centre. Autrement dit, à première vue, la distribution des emplois urbains semble nettement moins à même de constituer un outil efficace pour « juguler » un étalement urbain jugé néfaste. Toutefois, nous verrons dans le chapitre suivant que d'autres éléments, en lien avec les distances domicile-travail, peuvent remettre en cause cette conclusion.

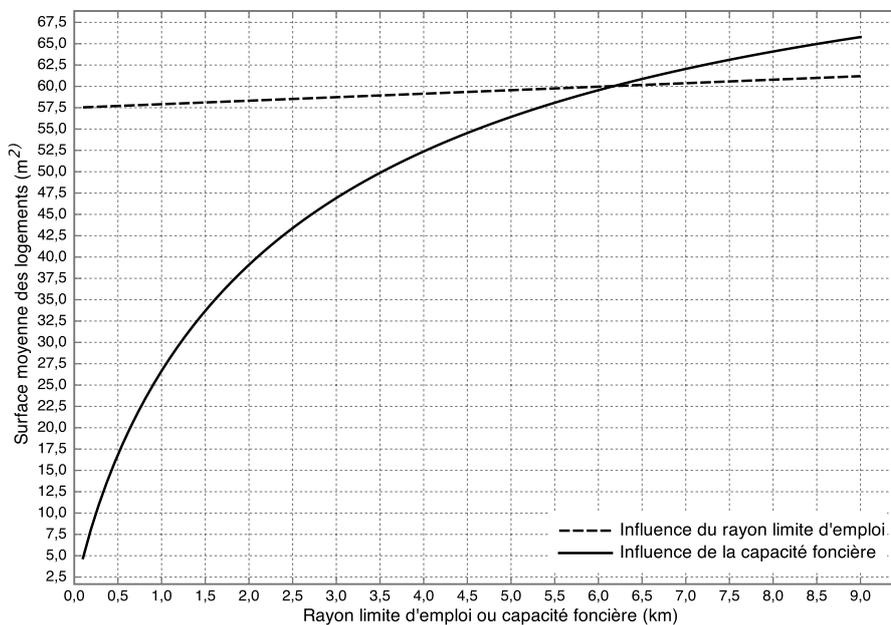


FIGURE 7.5.: Influences de λ et ρ_f sur la surface moyenne des logements.

7.3.3. Rente foncière

Les analyses menées sur la localisation résidentielle et la densité résidentielle peuvent être également menées sur la rente foncière. L'intérêt d'une telle analyse réside dans le fait que la rente foncière incarne les arbitrages des ménages (cf. chapitre 1) et qu'elle rend compte de la pression foncière s'exerçant aux différentes localisations urbaines. L'équation (7.18) relie de manière linéaire la rente foncière en $r_\omega(\rho)$ et ρ . En reformulant son expression de la manière suivante :

$$R_\omega(\rho) = R_A + \frac{a}{\lambda}(N - b\rho), \quad (7.24)$$

on peut étudier la sensibilité de la rente foncière aux paramètres qui agissent sur elle. La proposition suivante résume les résultats de sensibilité de la rente foncière :

Proposition 7.5 (Sensibilité de la rente foncière). *À lieu d'emploi ρ donné, la rente foncière R_ω :*

- (i) *est insensible au revenu brut Y , au coût fixe de transport a_0 , au coût marginal de transport dans le centre a' , et aux paramètres α et β ;*
- (ii) *augmente avec le coût marginal de transport dans la zone résidentielle a ;*
- (iii) *augmente avec R_A et diminue avec λ ;*
- (iv) *augmente avec b ainsi qu'avec ρ_f .*

La démonstration est directe, en tenant compte du fait que $N = b\rho_f$.

Toutefois, l'influence de b et de ρ_f telle qu'elle vient d'être déterminée n'est valable qu'à lieu d'emploi fixé d'une part, et avec un nombre de ménages variable, ce qui ne permet pas de « suivre » un ménage. La proposition suivante complète donc les résultats précédents, lorsque N est fixé :

Proposition 7.6. *Si le nombre total de ménages N est fixé :*

- (i) *la rente foncière $R_\omega(\rho)$ à lieu d'emploi fixé diminue avec b et augmente avec ρ_f ;*
- (ii) *la rente foncière $R(n)$ d'un ménage donné numéroté n est insensible à la densité des emplois b , ainsi qu'au rayon limite d'emploi ρ_f .*

La démonstration est également directe, compte tenu de l'expression de R_ω et que $R(n) = R_A + a(N - n)/\lambda$.

Le point (ii) est contre-intuitif, puisqu'il signifie que dans une ville fermée, un ménage donné fait toujours face au même niveau de rente foncière, quels que soient la densité des emplois ou le rayon limite des emplois. En réalité, il traduit le fait que le niveau d'enchère foncière d'un ménage ne dépend pas de la *concentration*

des emplois, mais uniquement de la *localisation* de son emploi (et des paramètres comme le coût de transport). Ce point est à souligner car il distingue la rente foncière des variables étudiées précédemment, comme la localisation et la densité résidentielles. Tout se passe comme si les variables d'ajustement des ménages étaient la localisation et la taille du logement, tandis que le prix auquel le ménage peut payer un mètre-carré resterait inchangé.

La Figure 7.6 présente les courbes de rente foncière correspondant à différentes valeurs du coût unitaire de transport dans la zone résidentielle a , tandis que la Figure 7.7 fait varier la valeur du rayon limite d'emploi ρ_f . Cette dernière figure met en évidence le fait que l'influence du rayon limite d'emploi sur le niveau de la rente foncière est très limitée, puisque lorsque ρ_f passe de 2 km à 8 km, l'augmentation de la rente atteint son maximum au niveau de la frontière urbaine, et représente une hausse de seulement 2,5 %. L'influence du coût marginal de transport est beaucoup plus significatif, et comparable aux résultats obtenus dans le cadre du modèle monocentrique standard.

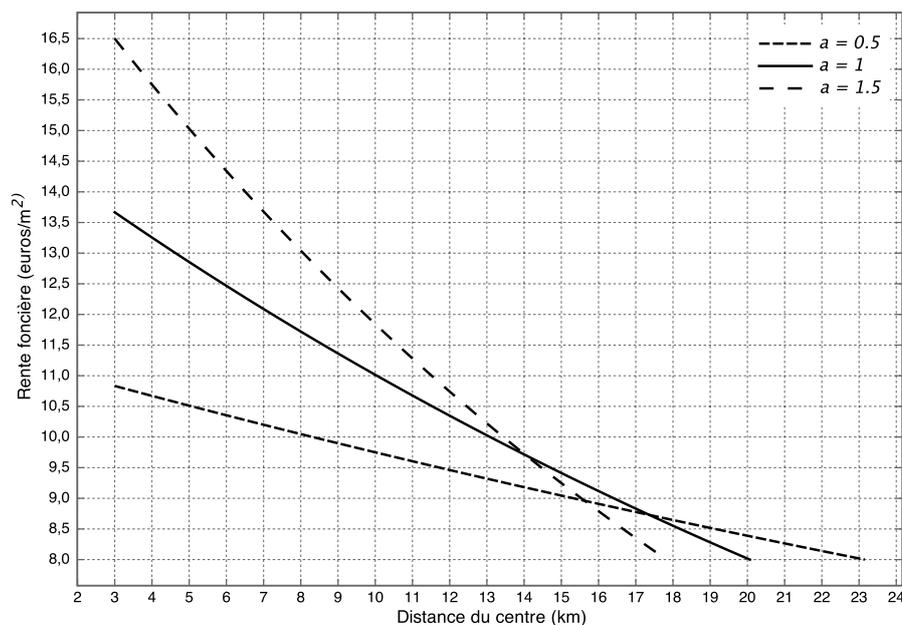


FIGURE 7.6.: Courbes de rente foncière pour différentes valeurs de a .

7.3.4. Revenu net des ménages

Par ailleurs, en notant que le revenu net des coûts de transport s'écrit $I_\omega(\rho) = Y_0 - ar_\omega(\rho) + a'\rho$, il apparaît que I_ω réagit en sens opposé à r_ω à des variations des

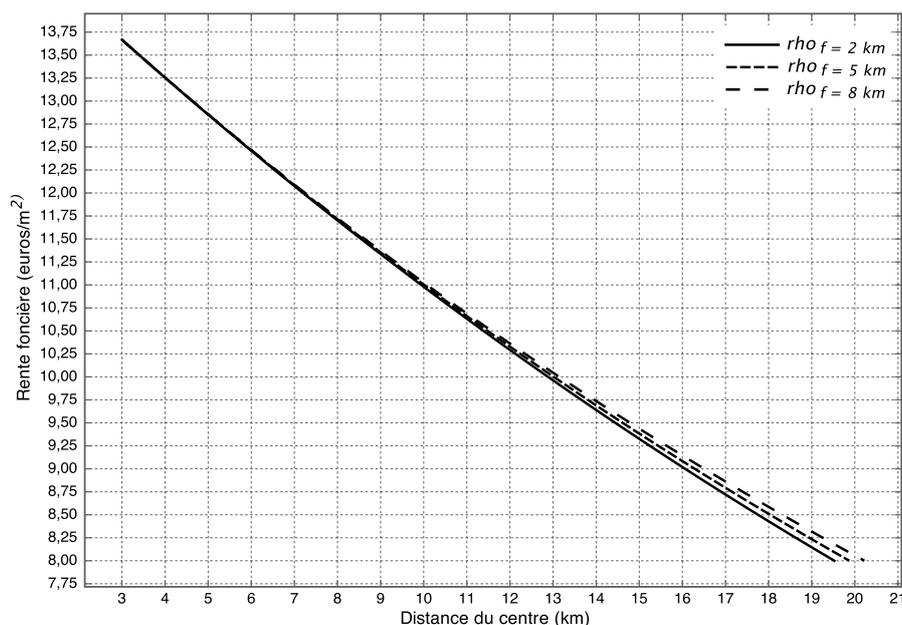


FIGURE 7.7.: Courbes de rente foncière pour différentes valeurs de ρ_f .

paramètres du modèle, autres que Y , a_0 , a , a' , b et ρ_f . Or, le revenu net se réécrit, compte tenu de l'expression de r_ω :

$$I_\omega(\rho) = \left(Y - a_0 + \frac{a'}{\alpha'} \tilde{\rho} - ar_0 \right) \left(1 - \frac{\rho}{\tilde{\rho}} \right)^{\beta'} - \frac{a'}{\alpha'} (\tilde{\rho} - \rho), \quad (7.25)$$

ce qui permet de vérifier que I_ω augmente avec Y et diminue avec a_0 . En revanche, nous n'avons pas été en mesure de déterminer, de manière analytique, le sens de variation du revenu net en réponse à des variations de a et a' . Seules des simulations numériques nous ont permis de constater que I_ω diminue avec a et a' , qu'il diminue avec b et augmente avec ρ_f . L'influence négative du coût unitaire de transport, dans la zone résidentielle comme dans la zone centrale, est relativement intuitive. Toutefois, il convient de souligner que ce résultat émerge de la nature mixte du revenu considéré ici : il s'agit en fait d'un revenu à la fois monétaire et temporel, dont la justification se trouve dans l'hypothèse implicite de substituabilité parfaite entre temps et revenu par l'intermédiaire d'un salaire horaire. Autrement dit, on suppose que le revenu Y est la somme d'un revenu monétaire et d'un revenu temporel éventuellement transformable en revenu monétaire. Cette hypothèse implicite est pratiquement toujours employée, souvent sans être même signalée, dans les modèles urbains théoriques. Fujita (1989) propose une extension au modèle monocentrique standard dans lequel les ménages disposent d'un budget monétaire et d'un budget temporel distincts.

7.3.5. Récapitulatif de l'analyse de sensibilité

La Table 7.1 offre un récapitulatif des différentes analyses de sensibilité que nous avons menées⁴. Plus précisément, le tableau présente les résultats pour un ménage donné et pour une population fixée. Nous avons vu en particulier que dans ces conditions, le rayon limite d'urbanisation réagit de la même manière que la localisation résidentielle, puisqu'il s'agit de la localisation résidentielle des ménages travaillant en ρ_f . Nous avons privilégié ce cas de figure car il facilite l'interprétation des résultats. En effet, il permet de rendre directement compte de l'influence des différents paramètres sur chacun des ménages, en observant la « trajectoire » d'un ménage donné. Les analyses ayant été menées en statique comparative, les résultats n'ont pas prétention à représenter une dynamique des ménages, mais renseignent sur le sens de variations des différentes variables attachées aux ménages avec les paramètres du modèle.

| | Loc. résidentielle | Surf. logement | Rente foncière | Revenu net |
|-----------|--------------------|----------------|----------------|------------|
| Y | + | + | 0 | + |
| β | + | + | 0 | - |
| α | - | - | 0 | + |
| a_0 | - | - | 0 | - |
| a | (-) | (-) | + | (-) |
| a' | - | - | 0 | (-) |
| r_0 | + | + | 0 | - |
| λ | - | (+) | - | + |
| R_A | - | - | + | + |
| b | (-) | (-) | 0 | (-) |
| ρ_f | (+) | (+) | 0 | (+) |

TABLE 7.1.: Tableau récapitulatif de l'analyse de sensibilité.

7.4. Conclusion

Nous avons, dans ce chapitre, mené les analyses de sensibilité du modèle développé au chapitre précédent. Pour ce faire, nous avons défini un cadre simplifié permettant une résolution analytique du modèle et, pour un certain nombre des paramètres et des variables, un traitement analytique de la sensibilité. Nous avons dû

4. Les parenthèses signalent un résultat obtenu uniquement de manière numérique.

nous contenter de simulations numériques pour d'autres paramètres. Les résultats sont semblables à ceux obtenus à partir du modèle monocentrique de base quant à l'influence des paramètres présents dans les deux modèles.

En revanche, nous avons obtenu des résultats originaux pour l'influence des paramètres liés à la distribution des emplois ou au coût de transport dans la zone centrale. En particulier, nous avons montré que toutes choses égales par ailleurs, un niveau élevé de congestion au centre de l'agglomération favorise la concentration des ménages. De même, nous avons confirmé le rôle moteur de la décentralisation des emplois sur l'étalement des ménages. Toutefois, nous avons également noté que le rôle joué par les paramètres liés à la distribution des emplois est nettement plus limité que celui joué par le coût de transport.

Le cadre, nécessairement limité, d'une analyse de sensibilité en statique comparative, ne permet cependant pas d'approfondir ces questions. Nous avons par conséquent, dans le chapitre suivant, cherché à compléter ces résultats et à les replacer dans une perspective plus large d'analyse du rôle de la décentralisation des emplois sur les trajets domicile-travail.

Annexe 7.A Hypothèses et valeurs numériques choisies

7.A.1 Interprétation complémentaire de la forme de la fonction de coût

Une interprétation complémentaire est possible pour la spécification particulière de la fonction de coût de transport que nous avons choisie⁵ : le modèle monocentrique classique suppose en effet que les travailleurs se rendent au CBD chaque jour pour travailler et pour effectuer leurs achats éventuellement. Si on suppose, dans le cadre de notre modèle qu'un travailleur se rend chaque jour ouvrable sur son lieu de travail en ρ , qu'il prolonge son trajet jusqu'au CBD s fois par semaine pour ses achats ($1 \leq s \leq 5$), et que le coût marginal de transport est uniforme sur l'ensemble de la ville, le coût journalier moyen de transport qu'il supporte s'écrit :

$$T(\rho, r) = \frac{1}{5}[5a_0 + sar + (5 - s)a(r - \rho)] = a_0 + ar - \frac{5 - s}{5}a\rho$$

On retrouve (7.6) avec $a' = \frac{5-s}{5}a$. Autrement dit, une augmentation de a' dans ce contexte correspond à une baisse de la fréquence à laquelle les ménages se rendent au CBD pour leurs achats. On remarquera néanmoins que cette interprétation n'est valable que pour des valeurs de a' telles que $a' \leq a$.

7.A.2 Valeurs numériques employées

Pour les simulations numériques, nous avons employé les valeurs suivantes des paramètres :

5. Nous sommes ici redevable envers Jacques-François Thisse qui, par une remarque lors d'une conférence à l'Université de Lille 3 en juin 2010, nous a suggéré cette seconde interprétation.

| Paramètre | Valeur | Unité |
|-----------|-----------|------------------------|
| N | 1 700 000 | ménages |
| Y | 2 500 | €/mois |
| α | 0,72 | — |
| β | 0,28 | — |
| a_0 | 5 | €/aller-retour |
| a | 1 | €/km aller-retour |
| a' | 1,5 | €/km aller-retour |
| r_0 | 3 | km |
| λ | 6 | km |
| R_A | 8 | €/m ² /mois |
| ρ_f | 5 | km |

Le nombre de ménages a été fixé à un peu moins de 2 millions. Nous avons en effet voulu principalement rendre compte de la localisation des ménages dont l'emploi est situé dans la zone d'emplois centrale de l'agglomération. En effet, nous avons considéré que les emplois constituant la « base résidentielle » (Davezies, 2008) du territoire ne répondait pas à la même logique de localisation que les emplois centraux. Nous avons ensuite ajusté la capacité foncière en conséquence (prise en compte d'environ un tiers des ménages seulement).

Annexe 7.B Annexes mathématiques

7.B.1 Résolution de l'équation différentielle (7.16)

Nous cherchons à résoudre l'équation différentielle suivante :

$$\dot{H}(r) = \frac{1}{\beta'} \frac{\lambda R_0 - aH(r)}{Y_0 - ar + a'H(r)/b}. \quad (7.16)$$

Changement de variable

On opère un changement de variable en définissant l'application suivante :

$$r : \begin{cases} [0, N] & \rightarrow [r_0, +\infty] \\ n & \mapsto H^{-1}(n) \end{cases}.$$

Par conséquent, $\dot{r}(n) = 1/\dot{H}(r(n))$. L'équation (7.16) devient :

$$\dot{r} = \beta' \frac{Y_0 - ar + \frac{a'}{b}n}{\lambda R_0 - an}.$$

Ce qui se réécrit encore :

$$\dot{r}(\lambda R_0 - an) + \beta' r = \beta' \left(Y_0 + \frac{a'}{b}n \right). \quad (7.26)$$

Solution de l'équation homogène associée

L'équation homogène associée à (7.26) est :

$$\dot{r}(\lambda R_0 - an) + \beta' r = 0$$

Elle s'intègre directement en :

$$\bar{r}(n) = \bar{r}_0 \left(\frac{\lambda R_0 - an}{\lambda R_0} \right)^{\beta'}$$

où \bar{r}_0 est une constante à déterminer.

Recherche d'une solution particulière

Nous employons la méthode de variation de la constante. Autrement, on cherche une solution à l'équation (7.26) sous la forme $\tilde{r}(n) = \nu(n)\bar{r}(n)$. Puisque \bar{r} est solution de l'équation homogène, on a :

$$\dot{\nu}\bar{r}(\lambda R_0 - an) = \beta' \left(Y_0 + \frac{a'}{b}n \right)$$

Ce qui donne, avec l'expression de \bar{r} :

$$\dot{\nu}(n) = \frac{\beta'}{\bar{r}_0} (\lambda R_0)^{\beta'} \left(Y_0 + \frac{a'}{b}n \right) (\lambda R_0 - an)^{-(\beta'+1)}$$

Quelques manipulations fournissent une expression directement intégrable, si bien que :

$$\tilde{r}(n) = \frac{\beta a'}{\alpha a} \tilde{\rho} \frac{\lambda R_0 - an}{\lambda R_0} + \left(\frac{Y_0}{a} + \frac{a'}{a} \tilde{\rho} \right)$$

où $\tilde{\rho} = \lambda R_0/ab$.

Solution générale de l'équation

En sommant cette solution particulière avec la solution de l'équation homogène, on obtient une expression de $r(n)$, solution de l'équation (7.26) :

$$r(n) = \bar{r}_0 \left(\frac{\lambda R_0 - an}{\lambda} \right)^{\beta'} + \frac{\alpha a'}{\beta a} \tilde{\rho} + \left(\frac{Y_0}{a} + \frac{a'}{a} \tilde{\rho} \right)$$

La valeur de \bar{r}_0 est obtenue en notant que $r(0) = r_0 = \bar{r}_0 + \frac{\alpha + \beta}{\alpha} \frac{a'}{a} \tilde{\rho} + \frac{Y_0}{a}$. Par ailleurs, étant donné que $\lambda R_0 - an = \lambda R$, on obtient une expression reliant la localisation résidentielle des ménages à la rente foncière :

$$r = A + B \frac{R}{R_0} - C \left(\frac{R}{R_0} \right)^{\beta'} \quad (7.17)$$

avec $A = \frac{Y_0}{a} + \frac{a'}{a} \tilde{\rho}$, $B = \frac{\beta a'}{\alpha a} \tilde{\rho}$, et $C = A + B - r_0$.

7.B.2 Relation analytique entre R et r

Dans le cas où $\alpha = \beta$, et donc $\alpha' = \beta' = 1/2$, la relation (7.17) se réduit à :

$$r = A + B \frac{R}{R_0} - C \sqrt{\frac{R}{R_0}}$$

$\sqrt{R/R_0}$ est donc solution de l'équation du second degré suivante en x :

$$x^2 - \frac{C}{B}x + \frac{A - r}{B} = 0$$

En retenant la solution décroissante en r , $x^* = \frac{C}{2B} \left(1 - \sqrt{1 - \frac{4B}{C^2}(A - r)} \right)$, et on obtient finalement :

$$R(r) = R_0 \left(\frac{C}{2B} \right)^2 \left(1 - \sqrt{1 - \frac{4B}{C^2}(A - r)} \right)^2$$

Les autres variables d'intérêt, comme la densité résidentielle, s'en déduisent ensuite.

7.B.3 Preuve de la Proposition 7.1

Démonstration. Avant tout, on peut traiter les points (iv) et (v). En effet, l'influence de R_A a été traitée dans le cas général, et le résultat reste donc valable avec les spécifications particulières. De même λ n'intervient, dans (7.17), que groupé avec

R_A sous la forme du produit λR_A . Par conséquent, son influence sur r_ω est identique à celle de R_A .

Pour révéler le sens de variation de r_ω en fonction des différents paramètres, on commence par adapter l'expression (7.17). Tout d'abord, on note $x = 1 - \rho/\tilde{\rho} \geq 0$ car $\rho \leq \rho_f \leq \tilde{\rho}$. Plus précisément, en notant $t = R_A/R_0 < 1$, on a : $t \leq x \leq 1$.

Pour étudier l'influence de r_0 , on remarque qu'on peut écrire $r_\omega(\rho) = K_0 + r_0 x^{\beta'}$, avec K_0 indépendant de r_0 . Donc la position résidentielle augmente avec r_0 : cela prouve le point (i) de la proposition. Ensuite, on peut écrire $r_\omega(\rho) = \frac{1}{a} Y_0 (1 - x^{\beta'}) + K_1$, avec K_1 indépendant de Y_0 , d'où le point (ii). Puis, $r_\omega(\rho) = a' K_2 + K_3$, avec $K_2 \leq 0$ (grâce à un développement limité en $\rho/\tilde{\rho}$) et K_3 indépendants de a' , ce qui prouve une partie du point (iii). Enfin, pour le point (vi), nous reformulons (7.17) de la manière suivante :

$$r_\omega - r_0 = (A - r_0)(1 - x^{\beta'}) + \frac{\beta}{\alpha} \frac{a'}{a^2 b} (x - x^{\beta'})$$

Comme la fonction $x^{\beta'}$ décroît avec β et croît avec α car $x \leq 1$, les fonctions $1 - x^{\beta'}$ et $x - x^{\beta'}$ croissent avec β et décroissent avec α , tout comme la fonction β/α . Cette influence se conserve par le produit et la somme de fonctions positives, donc r_ω est une fonction croissante de β et décroissante de α .

□

7.B.4 Preuve de la Proposition 7.3

Démonstration. Puisque $h_\omega = b/\dot{r}_\omega$ et $s_\omega = \lambda/b\dot{r}_\omega$, nous analysons la sensibilité de \dot{r}_ω à un paramètre θ en considérant la fonction dérivée $\theta \mapsto \frac{\partial}{\partial \theta} \dot{r}_\omega$. Étant donné que $\dot{r}_\omega(\rho) = \frac{\partial}{\partial \rho} r_\omega(\rho)$, on a, de manière formelle :

$$\frac{\partial \dot{r}_\omega}{\partial \theta} = \frac{\partial^2 r_\omega}{\partial \theta \partial \rho} = \frac{\partial^2 r_\omega}{\partial \rho \partial \theta}$$

par le théorème de Schwartz car r_ω est suffisamment régulière. Or nous avons déjà établi l'expression de $\partial^2 r_\omega / (\partial x \partial \theta)$ pour des variations de x correspondant uniquement à des variations de ρ , selon la relation de changement de variable $dx = -d\rho/\tilde{\rho}$. Donc :

$$\frac{\partial^2 r_\omega}{\partial \rho \partial \theta} = \frac{\partial}{\partial \rho} \frac{\partial r_\omega}{\partial \theta} = -\frac{1}{\tilde{\rho}} \frac{\partial}{\partial x} \frac{\partial r_\omega}{\partial \theta} = -\frac{1}{\tilde{\rho}} \frac{\partial^2 r_\omega}{\partial x \partial \theta}$$

Le signe de $\partial^2 r_\omega / (\partial \rho \partial \theta)$, donc de $\frac{\partial}{\partial \theta} \dot{r}_\omega$, est opposé à celui de $\partial^2 r_\omega / (\partial x \partial \theta)$, donc identique à celui de $\partial r_\omega / \partial \theta$. Ainsi la fonction $\theta \mapsto \dot{r}_\omega$ varie selon θ comme $\theta \mapsto r_\omega$, et les résultats de la Proposition 7.1 sont vrais aussi pour \dot{r}_ω .

Par conséquent, compte tenu des expressions respectives de h_ω et s_ω , l'analyse de

sensibilité est directement transposable à partir de celle de r_ω , sauf pour l'influence de λ pour s_ω .

□

8

Décentralisation des emplois et usage des transports

8.1. Introduction

Ce chapitre a pour objectif d'étudier, à partir du modèle d'équilibre urbain développé et analysé dans les deux chapitres précédents, la relation entre l'étalement des emplois et les trajets domicile-travail des ménages, en termes de distance, de coût, d'usage des transports et d'équité spatiale. Nous comparons pour ce faire différentes configurations urbaines et établissons un diagnostic.

La compréhension du phénomène d'étalement urbain constitue un des enjeux majeurs de l'économie urbaine. Mieux comprendre ses causes et ses conséquences est la clé pour lutter contre ses effets les plus néfastes sans tomber dans un pur débat idéologique. En particulier, l'évaluation des conséquences de l'étalement urbain interroge le lien entre ce phénomène et les distances parcourues par les ménages, parmi lesquelles les distances domicile-travail représentent une part toujours importante (Mills et Hamilton, 1994 ; Meyere *et al.*, 2005 ; De Solère *et al.*, 2010). En effet, si l'étalement urbain a de nombreuses autres conséquences (sur la consommation d'espace agricole, sur les performances thermiques des logements, voire sur la biodiversité), ce sont très souvent ses effets en termes d'usage des transport, et de la voiture en particulier, qui focalisent l'essentiel du débat.

Toutefois, dans ce contexte, le rôle de la décentralisation des emplois est souvent oublié ou négligé. Or, outre que ce mouvement d'étalement des activités est effectivement observé aux États-Unis (Glaeser et Kahn, 2001) ainsi qu'en France, et en particulier en Île-de-France (Hurriot et Bourdeau-Lepage, 2009), son rôle reste relativement indéterminé : ce phénomène constitue-t-il un moteur à l'étalement résidentiel ? Aggrave-t-il ou, au contraire, atténue-t-il certains effets de l'étalement urbain ? C'est au cœur de ce débat que prend place ce chapitre. Un de nos objectifs est donc, à partir du modèle d'équilibre urbain développé dans les deux chapitres précédents, de mettre en évidence un mécanisme par lequel l'étalement des emplois peut apparaître comme partiellement responsable de l'étalement urbain. Un second objectif est d'apporter des éléments de réflexions sur le fait que l'étalement urbain lié à l'étalement des emplois est potentiellement économe en termes de distances parcourues par les ménages pour se rendre à leur emploi.

La deuxième section de ce chapitre propose un bref rappel des principaux éléments du débat sur la question de l'étalement urbain et de ses liens avec les transports. Nous y insistons sur le rôle potentiel de la localisation des emplois et sur les apports de la modélisation urbaine sur cette question. La troisième section précise l'approche retenue, et rappelle les principales hypothèses et notations du modèle. La quatrième section présente une première analyse, visant à étudier le rôle de l'étalement des emplois et de la congestion dans l'étalement urbain. La cinquième section débute par une analyse, dans le même cadre de modélisation et à partir de plusieurs configurations types, de l'usage des transports et de ses variations dans l'espace, puis se focalise sur l'étude approfondie de deux configurations urbaines contrastées.

8.2. Étalement urbain et transport : un sujet à débat

L'objectif de cette section n'est pas d'analyser de manière exhaustive la littérature sur cette question de l'étalement urbain et de ses conséquences en matière de déplacements. Pouyanne (2004) en propose en effet une revue très complète. En revanche, nous chercherons à rappeler quelques faits saillants et les principaux termes du débat.

8.2.1. Qu'est-ce que l'étalement urbain et d'où vient-il ?

Il convient tout d'abord de faire un bref rappel sur cette notion complexe qu'est l'étalement urbain. Elle regroupe en effet deux phénomènes liés, mais bien distincts : l'étalement urbain désigne à la fois le fait que les villes s'étendent, c'est-à-dire qu'elles occupent, dans leur ensemble, une surface croissante, et le fait que

cette extension se fasse de manière plus que proportionnelle à la croissance de la population, *via* une diminution de la densité et/ou du gradient de densité. Ainsi, la croissance de la population d'une ville entraîne généralement une augmentation de la surface occupée par la ville, mais cette augmentation ne s'accompagne pas nécessairement d'une diminution du gradient de densité, qui peut fort bien augmenter partout. Au contraire, ce que l'on désigne par étalement urbain implique généralement un aplatissement de la courbe de densité, ou du moins un accroissement de la surface moyenne des logements. La Figure 8.1 donne une illustration des deux phénomènes dans un cadre monocentrique.

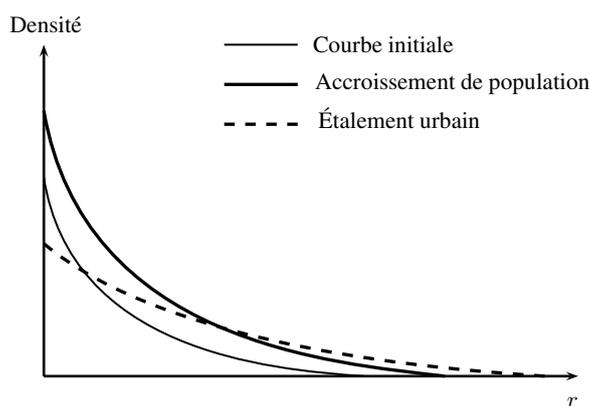


FIGURE 8.1.: Évolution de la courbe de densité en cas d'accroissement de la population et d'étalement urbain. *Source* : adapté de Huriot et Bourdeau-Lepage (2009)

Les causes de l'étalement sont nombreuses, et de différents ordres. Il ne s'agit pas ici d'en détailler les mécanismes, mais plutôt d'évoquer trois séries de facteurs jouant un rôle dans ce processus (Mieszkowski et Mills, 1993 ; Boiteux et Huriot, 2000) :

- L'évolution de la structure productive : le passage d'une économie industrielle vers une économie de services, alliée à la différenciation des produits et à la personnalisation des relations de service, limite l'importance des forces d'agglomération liées aux économies sur les coûts de transport, et met plutôt l'accent sur les conditions de circulation de l'information.
- La croissance de la population et la modification des préférences des ménages : l'augmentation des revenus¹ et les préoccupations des ménages pour la santé

1. On notera que l'effet de la hausse des revenus réels est difficile à évaluer, car si toutes choses égales par ailleurs un accroissement des revenus augmente la disposition à payer des ménages pour l'espace, il est probable que cette hausse s'accompagne d'un accroissement du coût d'opportunité du temps de transport, compensant la baisse du coût relatif de transport et poussant les ménages à chercher une localisation centrale.

et le bien-être ont tendance à accroître leur demande d'espace, tant privé que naturel. La croissance démographique pousse également à la hausse la consommation d'espace à l'échelle globale.

- L'amélioration des technologies de transport : la baisse des coûts de transport, en termes monétaires, mais surtout en termes temporels², ainsi que la diffusion de l'automobile qui homogénéise l'espace en terme d'accessibilité, rendent la périphérie des villes plus attractives (pour la France, voir Raux, 1993 ; Dupuy, 1995 ; Wiel, 1999).

8.2.2. Des conséquences de l'étalement urbain au débat

En ce qui concerne les conséquences de l'étalement urbain, comme le rappellent Huriot et Bourdeau-Lepage (2009, chap. 11),

« Les conséquences économiques de l'extension spatiale des villes sont nombreuses et importantes. Elles se traduisent en avantages et en coûts, en termes d'usage du sol, de transport, d'aménités et d'environnement, et elles sont à la source des débats récurrents sur le caractère souhaitable ou indésirable de l'étalement et sur la nécessité ou pas d'endiguer ce mouvement pour promouvoir une ville durable, assimilée d'emblée à la ville compacte. »

Les auteurs soulignent ainsi non seulement la multiplicité des effets de l'étalement urbain, mais également le fait que ces effets peuvent être positifs comme négatifs.

8.2.2.1. Un étalement nuisible

Aujourd'hui, l'étalement urbain est, en Europe comme aux États-Unis, généralement très négativement connoté, car synonyme de gaspillages dans toutes sortes de domaines — de sol, de ressources naturelles, d'énergie — et tenu pour responsable d'une baisse de la productivité et de la détérioration de la qualité de vie des urbains (voir par exemple Deakin, 2007 ; Glaeser et Kahn, 2008). L'argument principal contre l'étalement repose sur des comparaisons de consommation énergétique : l'étalement urbain entraînerait un accroissement de celle-ci. La courbe de Newman et Kenworthy (1989), malgré les nombreuses critiques qu'elle a pu recevoir, reste bien souvent le symbole de cette idée. Les faibles densités favorisent un habitat individuel, généralement moins bien isolé et plus spacieux, donc plus grand consommateur d'énergie, pour le chauffage en particulier. De plus, des travaux d'estimation des coûts d'urbanisation, et leurs répercussions sur les finances publiques, tendent à montrer qu'un mode de développement étalé est plus coûteux

2. La remarque précédente s'applique en partie, car les gains monétaires peuvent s'interpréter comme une hausse du revenu réel.

qu'un développement plus compact (voir par exemple Nouaille, 2009, ou Ecoffey et Pflieger, 2010, pour les services de provision d'eau). Mais, comme nous le verrons au 8.2.3, c'est dans le domaine du transport que les principaux arguments contre l'étalement urbain trouvent leur origine, puisque l'étalement urbain a généralement pour conséquence d'augmenter les consommations énergétiques liées au transport.

8.2.2.2. Un étalement partiellement revalorisé

Pourtant, le débat n'est pas tranché, et les arguments en faveur de la ville compacte, opposée à la ville étalée, apparaissent dès lors parfois d'ordre idéologique. Notons à ce propos que l'étalement urbain n'a pas toujours été vu comme une nuisance contre laquelle il faut lutter. Il a toujours accompagné la croissance des villes, tant démographique qu'économique, et a par conséquent souvent été vu comme un indicateur de la réussite économique d'une ville. Le fait que Paris se soit très fortement étendu depuis l'époque où la ville était cantonnée à l'Île de la Cité a longtemps été considéré comme un signe de sa puissance. De même, constatant la croissance démographique galopante de New York au début du XIX^e siècle, les autorités adoptent le *Commissioners' Plan*, qui prévoit la construction de rues sur la totalité de l'île de Manhattan, encourageant par là même l'occupation de cet espace disponible par la population entassée jusque-là dans la pointe sud de l'île.

Plus récemment, des éléments et des travaux ont partiellement réhabilité un étalement urbain pratiquement accusé par ailleurs de tous les maux urbains actuels. Remarquons d'abord que l'étalement urbain répond aux aspirations des populations à jouir de plus d'espace, et au refus souvent évoqué des fortes densités vécues comme une forme d'entassement. Bien entendu, les habitants de villes comme Paris et San Francisco sont souvent très attachés à la densité, comme le souligne Brueckner (2000) en indiquant que la densification peut améliorer la qualité de vie urbaine en favorisant les interactions sociales. Pourtant, le même auteur rappelle qu'entasser les populations dans des villes denses risque de détériorer la qualité de vie des habitants, et en particulier des Américains, plus habitués que les Européens au style de vie périurbain. De plus, il montre que la croissance urbaine, dont l'étalement est une des manifestations, n'est pas nécessairement indésirable, bien au contraire. Ce sont certaines de ses caractéristiques, liées à des défaillances de marché, qui le sont. Plus précisément, il cite trois externalités qui concernent respectivement l'appréciation de la valeur des espaces libres, l'estimation du coût de la congestion, et la prise en compte des coûts d'infrastructure des nouveaux développements.

Si l'aspiration croissante des ménages pour l'espace ne justifie pas à elle seule l'étalement urbain, ni ne supprime les nuisances dont il peut être à l'origine, il n'en demeure pas moins que ce constat nuance les jugements sévères dont il fait l'objet :

n'est-ce pas aussi le rôle d'une organisation sociale comme une ville de faire en sorte que ses habitants puissent vivre comme ils l'entendent ?

Par ailleurs, l'étalement est souvent accusé d'entraîner des coûts de viabilisation plus élevés que la compacité. Pourtant, en ce qui concerne les coûts de construction des logements, les travaux de Castel (2005, 2006) ont montré que ceux-ci augmentent avec la compacité de la forme urbaine.

Enfin, Kahn (2007) indique que les mesures visant à augmenter la compacité des villes favorisent les propriétaires de logements existants et défavorisent les locataires, en poussant la rente foncière à la hausse. Or, les propriétaires appartiennent à des catégories sociales souvent plus favorisées que les locataires si bien que les villes compactes ne facilitent pas la mixité sociale. Au contraire, en abaissant les prix fonciers, on peut penser que l'étalement urbain favorise l'accès à la propriété de populations qui ne pourraient l'envisager dans un autre contexte³.

8.2.3. Transport et étalement urbain : un lien complexe

Sur la question du transport plus spécifiquement, il semble que le lien entre déplacements quotidiens et étalement soit complexe et que le débat ne soit pas tranché non plus. Tout d'abord, selon nous, le débat est en partie faussé par une certaine confusion entre l'influence de la densité *au sein* d'une agglomération, et son influence *dans l'absolu*. Plus précisément, de nombreux travaux (par exemple, sur données françaises, Fouchier, 1997, 1998 ; Pouyanne, 2004 ; Blaudin de Thé, 2010) s'appuient sur des analyses intra-urbaines des distances parcourues (ou de la consommation de carburant) pour en déduire le rôle moteur des basses densités dans l'augmentation de ces distances. Autrement dit, ces auteurs mettent en avant l'existence d'un gradient des distances parcourues, en sens inverse du gradient de densité pour incriminer les basses densités. Mais il n'est pas acquis que le rôle absolu de la densité aille dans le même sens. Plus précisément, une densité moyenne faible est-elle nécessairement associée à des distances parcourues élevées ? Bruegman (2005) envisage le contraire, en s'appuyant sur le fait que l'étalement des emplois accompagne l'étalement des populations. Il constate que les temps de déplacement domicile-travail sont plus élevés dans les villes très denses.

Les travaux de Kahn (2007) et Glaeser et Kahn (2008) s'affranchissent de cette limite en comparant les agglomérations entre elles. Ils montrent ainsi que les villes américaines les moins denses sont également celles qui engendrent les plus fortes émissions de gaz à effet de serre dans les transports, ce qui semble accréditer la

3. Toutefois, comme l'ont montré Coulombel *et al.* (2007) pour l'Île-de-France, la mauvaise prise en compte, par ces ménages, des coûts liés à l'usage de la voiture particulière, les place parfois dans une situation financière délicate.

thèse d'un lien net entre densité et distances de transport. Melia *et al.* (2011) soulignent toutefois que cette conclusion, issue d'études transversales, ne permet pas de déterminer si les mêmes résultats pourraient être obtenus, dans une ville donnée, dans le cas d'une augmentation de la densité, qu'ils appellent *intensification*. L'exemple de Portland montre en effet que si une intensification urbaine réduit les distances parcourues, c'est dans des proportions relativement faibles, sans report modal significatif et au prix d'une congestion très forte, source de pertes de temps et d'externalités. À ce titre, comme le notent les auteurs, « À la lumière de ces résultats, il peut apparaître étrange que certains auteurs aient défendu l'intensification urbaine comme un moyen de réduire les externalités liées à la voiture⁴. »

De plus, la densité n'est pas, loin de là, la seule mesure de la forme urbaine. Les villes américaines les moins denses sont également souvent les plus jeunes et leur structure interne, leur forme urbaine, est très différente de celle des villes plus anciennes. Autrement dit, ce n'est pas uniquement leur densité qui différencie ces villes, mais également leur structure interne. Et la forme urbaine générique — monocentrique, duocentrique, polycentrique, etc. — elle-même ne suffit pas à caractériser correctement une agglomération et les conséquences, pour celle-ci, de l'étalement. Charron (2007) a ainsi montré, d'un point de vue théorique, qu'une même forme urbaine peut donner lieu à des résultats très différents en ce qui concerne les distances domicile-travail, en particulier du fait des différences dans les réseaux de transport et dans l'équilibre entre emplois et résidences. De même et d'un point de vue empirique cette fois, Aguiléra et Mignot (2007), citant Schwanen *et al.* (2004), rappellent que les résultats obtenus sont souvent contradictoires, du fait notamment de la multiplicité des formes urbaines, en particulier polycentriques. Leurs propres résultats, portant sur les agglomérations de Lille, Lyon et Marseille, indiquent que la forme polycentrique, partagée par Lille et Marseille, entraîne dans les deux cas des longueurs de déplacements domicile-travail très différentes, Lille étant relativement « économe » en distance, tandis que Marseille est caractérisée au contraire par de très longs déplacements. Lyon, pour sa part, de forme plutôt monocentrique, produit des longueurs de déplacement intermédiaires entre les deux précédentes.

Enfin, si la question des distances parcourues est délicate, celle des temps de parcours, et donc du coût économique de la compacité ou de l'étalement en termes de transport l'est également. Plusieurs auteurs, que nous avons déjà cités (Fouchier, 1997 ; Bruegman, 2005 ; Kahn, 2007) signalent que les temps de transport subis par les ménages sont plus faibles dans les zones moins denses, et cela, en comparaison intra-urbaine comme en comparaison inter-urbaine. Autrement dit, ce constat rejoint celui de Brueckner (2000) : le coût économique de la compacité n'est pas négligeable. Si certains auteurs (Kahn, 2007, par exemple) soulignent que le fait que

4. Traduction de l'auteur.

les villes compactes connaissent de hauts niveaux de congestion encourage l'usage des transports publics, moins consommateurs d'énergie, d'autres (Melia *et al.*, 2011, par exemple) sont, comme nous l'avons vu, d'un autre avis.

8.2.4. Notre positionnement au sein du débat

Nous venons de le rappeler, le débat n'est pas clos, et ce chapitre ne vise pas cet objectif, puisque, par la nature multiforme et en perpétuelle évolution des objets observés — les villes — il ne peut l'être. Il s'agit plutôt pour nous d'apporter une contribution d'économie théorique à cette question, à l'aide du modèle développé dans les chapitres précédents. Nous souhaitons montrer, par cette contribution, qu'en matière de forme de développement urbain il est sage de rester prudent face aux généralisations hâtives sur des bases partielles. Plus précisément, au sein du débat sur les liens entre étalement urbain et transport, deux questions nous apparaissent comme saillantes et nous sommes en mesure d'y apporter une contribution :

- Les distances parcourues, les coûts privés et les externalités du transport évoluent-ils dans le même sens en fonction de la densité ?
- L'étalement urbain est-il synonyme d'un accroissement des trafics ?

De la réponse à la première question découlent les liens entre les différentes mesures retenues pour évaluer l'impact de l'étalement urbain et déterminer son caractère néfaste ou non. De plus, la réponse à cette question peut mettre en évidence un découplage entre coût privé et coût externe, fortement préjudiciable. La seconde question interroge quant à elle plus directement le caractère néfaste de l'étalement urbain.

8.3. L'approche retenue et le modèle

8.3.1. Pourquoi étudier les déplacements domicile-travail ?

Notre contribution porte sur les déplacements domicile-travail. Ce choix, qui peut sembler fortement réducteur, se justifie : les trajets domicile-travail constituent aujourd'hui encore une partie considérable et relativement régulière, tant en structure origine-destination qu'en termes d'horaires, des trafics en milieu urbain. Mills et Hamilton (1994, chap. 13) le rappellent, si les trajets domicile-travail ne représentent, en distance parcourue, qu'environ 25 % des déplacements personnels des ménages urbains américains, ils constituent en revanche la très grande majorité des déplacements aux heures de pointe. En Île-de-France, un calcul à partir de l'Enquête Globale Transport de 2001 fournit le chiffre, plus élevé, de 37,9 % des distances parcourues. De plus, les déplacements pour motif travail (ou scolaire) représentent près de 80 % des déplacements à l'heure de pointe du matin (Meyere *et al.*, 2005). À

partir de l'analyse d'enquêtes Ménages Déplacements sur l'ensemble de la France, De Solère *et al.* (2010) montrent que si l'on ne considère que les déplacements en voitures réalisés à l'heure de pointe du matin, le motif travail représente 45 % des déplacements. Ils sont souvent le support d'autres types de déplacements, comme les déplacements pour motifs achats sur le trajet de retour du travail, ou la dépose des enfants à l'école le matin (Kwan, 1999). Plus généralement d'ailleurs, le lieu de travail polarise l'espace de vie quotidienne des actifs (Boulaïbal, 2001) si bien que la mobilité vers le travail reste structurante de la mobilité globale (Orfeuil, 1995). Ainsi, Meyere *et al.* (2005) montrent que plus de 80 % des déplacements pour motifs privés à l'heure de pointe du matin sont des déplacements d'accompagnement. Autrement dit, seuls 4 % des déplacements à l'heure de pointe du matin seraient liés à des motifs autres qu'assimilés au travail ou aux études.

Par ailleurs, du fait des volumes de trafic très élevés présents sur les réseaux aux heures de pointe (cf. chapitre 4), ce sont finalement les trajets domicile-travail qui sculptent la capacité des réseaux de transport, qui la consomment majoritairement et qui contribuent donc au dimensionnement des besoins d'infrastructures et de services de transport public (Orfeuil, 1995 ; De Solère *et al.*, 2010). Ils forment donc une base structurante pour la planification des transports. Ils sont également plus faciles à étudier que les autres types de déplacements, du fait de leur stabilité spatiale et temporelle, du moins à court terme. L'enjeu de ces déplacements en termes de consommation énergétique et d'émission de gaz à effet de serre est également à signaler. Étudier cette catégorie particulière⁵ de déplacements dans leur inscription urbaine constitue un enjeu de recherche important.

C'est pourquoi ces trajets, et même plus précisément la distance séparant le lieu de résidence et le lieu de travail, et ses variations dans l'espace, ont fait l'objet de nombreuses recherches, tant en géographie et science régionale (par exemple, Korsu et Massot, 2006 ; Shearmur, 2006 ; Boussauw *et al.*, 2010) qu'en économie urbaine. De plus, comme le soulignent Aguiléra et Mignot (2002), il semble important de s'intéresser à l'influence de la configuration urbaine sur chaque type de déplacements, et en particulier sur les déplacements domicile-travail, avant de pouvoir apporter une réflexion globale.

8.3.2. Une analyse à partir d'un modèle d'équilibre urbain

Le modèle monocentrique, dont nous avons déjà indiqué qu'il avait été intensivement employé, a également servi de bases aux discussions sur le sujet des trajets domicile-travail. La faiblesse de sa description des localisations de l'emploi, que nous avons déjà évoquée au chapitre 6, ainsi que son hypothèse d'optimisation par les ménages de la distance domicile-travail ont toutefois amené de nombreuses critiques,

5. Dans tous les sens du terme.

entre autres sur la base des distances parcourues par les ménages. En omettant de prendre en compte la dispersion des emplois, le modèle monocentrique standard limite fortement les possibilités d'étude des distances domicile-travail, et donc des coûts de transport, ainsi que des politiques urbaines visant à limiter l'étalement urbain, comme les taxes, péages, et les restrictions ou incitations à certains usages du sol. C'est pourquoi le modèle développé dans les deux chapitres précédents nous semble plus approprié pour explorer ces questions.

Nous reprenons ainsi le modèle développé au chapitre 6, dans sa version simplifiée du chapitre 7. Autrement dit, nous considérons une ville linéaire, dans laquelle la localisation des emplois est exogène. Par ailleurs, chaque ménage, qui ne comprend qu'un seul travailleur⁶, choisit sa localisation résidentielle en prenant son lieu d'emploi comme fixé. Comme nous l'avons déjà mentionné dans le chapitre 1, l'utilisation d'un modèle dans lequel les ménages se localisent à travers un critère de minimisation du coût de transport est un choix qui peut être critiqué (voir par exemple le corpus bibliographique sur la question du *wasteful commuting* : Hamilton, 1982 ; White, 1988b ; Hamilton, 1989 ; Cropper et Gordon, 1991 ; Small et Song, 1992 ; Ma et Banister, 2006). Même si nous ne chercherons pas, dans ce chapitre, à justifier de manière empirique la validité de ce type de modèle, il nous semble que si des éléments en sa défaveur ont été apportés par les travaux que nous venons de citer, d'autres travaux et éléments militent en sa faveur. En dehors de ceux que nous avons évoqués dans le chapitre 1, McDonald (2009), par exemple, obtient un calibrage satisfaisant d'un modèle monocentrique sur la ville de Chicago, tandis que De Palma *et al.* (2008) réalisent la même chose pour la région parisienne. Enfin, sa simplicité d'usage et sa capacité à faire émerger des phénomènes réellement observés sont pour nous des éléments décisifs. De même, Verhetsel *et al.* (2010) montrent, à partir du cas belge, que la distance et le temps de transport jouent, aujourd'hui encore, un rôle majeur dans la structure des déplacements domicile-travail. Enfin, Delons *et al.* (2008) développent un modèle d'interaction transport et usage du sol à l'échelle de l'Île-de-France — Pirandello — fondé sur les mêmes hypothèses que le modèle monocentrique.

Le modèle employé repose sur un certain nombre d'hypothèses et de notations qui doivent être rappelées.

6. Cette hypothèse est relativement forte, même si, en France en 2002, près de la moitié (42,3 % exactement) des couples ne comportaient qu'un actif (Stancanelli, 2006). Elle peut être relâchée sans que les résultats ne soient qualitativement modifiés, en considérant la localisation moyenne d'emploi des deux membres d'un couple biactif. Cela modifie la distribution des emplois, rendant la résolution analytique impossible.

Variables et paramètres exogènes

| | |
|----------------------------|-------------------------------------------------------------------------------------------------------|
| Rayon limite des emplois : | ρ_f |
| Distribution des emplois : | $f(\rho) = \begin{cases} b & \text{pour } \rho \in [0, \rho_f] \\ 0 & \text{ailleurs} \end{cases}$ |
| Capacité foncière : | $L(r) = \begin{cases} \lambda & \text{pour } r \in [r_0, +\infty[\\ 0 & \text{ailleurs} \end{cases}$ |
| Coût de transport : | $T(\rho, r) = a_0 + ar - a'\rho$ |
| Fonction d'utilité : | $U(z, s) = U_0 z^\alpha s^\beta$ |
| Population de la ville : | N |

Variables endogènes

| | |
|-------------------------------|------------------------------------------------------------------------------------------------------------------------|
| Rayon limite d'urbanisation : | r_A |
| Distribution des ménages : | $h(r) \begin{cases} = 0 & \text{pour } r < r_0 \text{ et } r > r_A \\ > 0 & \text{pour } r \in [r_0, r_A] \end{cases}$ |
| Rente foncière : | $R(r) \begin{cases} = R_A & \text{pour } r \geq r_A \\ > R_A & \text{pour } r \in [r_0, r_A[\end{cases}$ |

La modélisation que nous employons, ainsi que les analyses que nous proposons de réaliser comportent des limites, dont nous sommes conscients. Tout d'abord, la représentation du marché du travail est pratiquement absente de notre modèle. Intégrer une modélisation, même simple, de ce marché, permettrait d'analyser plus finement certaines des questions que nous abordons. Il s'agit là clairement d'une piste d'amélioration de notre modèle. Par ailleurs, si nous avons souhaité insister sur la questions des distances parcourues et des coûts privés de transport, d'autres facettes mériteraient également une analyse. En particulier, la question des externalités liées au transport (pollution, bruit, gaz à effet de serre) n'est pas traitée au delà du *proxy* que représente la distance parcourue. De plus, les questions d'intermodalité, ou même d'orientation modale, ne sont pas abordées dans notre analyse.

Notre analyse des liens entre étalement des emplois, étalement urbain et déplacements domicile-travail repose sur la détermination et l'étude de plusieurs configurations urbaines. Nous adoptons donc une démarche en deux temps : identification des configurations typiques dans un premier temps, analyse économique de ces configurations dans un second temps.

8.4. L'étalement des emplois, moteur de l'étalement urbain ?

L'étude des liens entre distance parcourue et coût du transport (en particulier temporel) permet de mettre en évidence un mécanisme à l'origine de l'étalement urbain. Cette idée s'appuie sur un certain nombre de travaux menés sur cette question. Ainsi, Fouchier (1997), dans une analyse sur l'Île-de-France, montre que si la distance parcourue chaque jour par les ménages franciliens diminue avec la densité de leur lieu d'habitation, le temps passé dans les transports est relativement stable, et même tend à augmenter avec la densité (Cf. Figure 8.2). De manière complémentaire, Genre-Grandpierre (2007) montre que la vitesse moyenne de déplacement augmente avec la portée des déplacements, et que cette dernière diminue avec la densité. Au total, le constat est le suivant : les ménages localisés dans des zones peu denses, donc loin du centre de la ville, parcourent des distances plus grandes, mais passent autant voire moins de temps en déplacement que les ménages localisés au centre de la ville. Ce constat est important, car il met en évidence un moteur possible à l'étalement urbain : lorsqu'en s'éloignant du centre les ménages bénéficient non seulement de prix immobiliers plus faibles mais encore d'un coût privé de transport plus faible. En Île-de-France les distances parcourues sont plus élevées en s'éloignant du centre, ainsi que les nuisances associées aux moyens de transport utilisés (pollution, gaz à effet de serre, plus intensément pour la voiture individuelle que les transports collectifs).

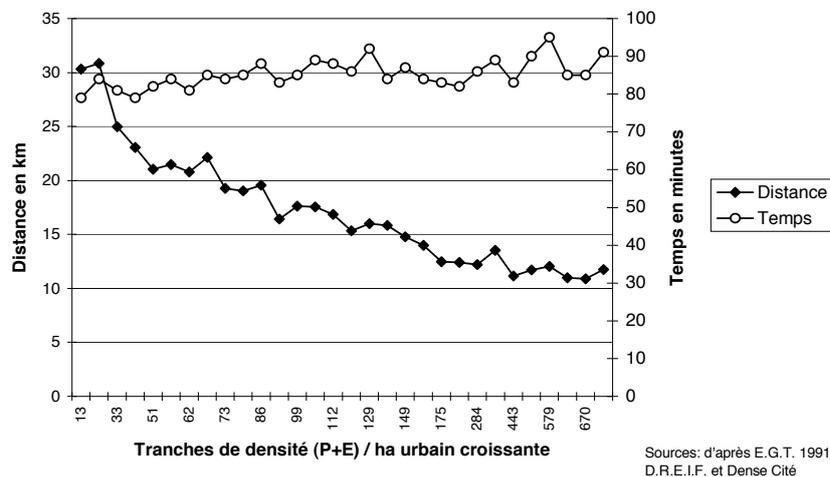


FIGURE 8.2.: Distance et temps de déplacement par individu et par jour en Île-de-France. *Source* : Fouchier (1997)

8.4.1. Identification de configurations urbaines typiques

8.4.1.1. Des critères aux configurations

Plaçons-nous dans un premier temps dans le cadre général du modèle, et intéressons-nous à la distance parcourue et au coût du transport subi par les ménages. La distance, fonction de r , s'écrit $\hat{D}(r) = r - \rho_\omega(r)$, tandis que le coût s'exprime par $\hat{T}(r) = T(\rho_\omega(r), r)$. Ce dernier varie avec la localisation résidentielle r d'une part de manière directe et croissante par l'intermédiaire de $T(\cdot, r)$, d'autre part de manière indirecte et décroissante par l'intermédiaire de $\rho_\omega(r)$. Les deux influences antagonistes se superposent avec une résultante variable, que nous allons illustrer dans des situations caractéristiques.

Ainsi, l'influence totale de r est déterminée par les dérivées droites de $\hat{D}(r)$ et $\hat{T}(r)$:

$$\frac{d\hat{D}(r)}{dr} = 1 - \dot{\rho}_\omega(r) \quad (8.1)$$

$$\begin{aligned} \frac{d\hat{T}(r)}{dr} &= \dot{\rho}_\omega(r) \frac{\partial}{\partial \rho} T(\rho_\omega(r), r) + \frac{\partial}{\partial r} T(\rho_\omega(r), r) \\ &= \frac{\partial T}{\partial \rho} \left[\dot{\rho}_\omega(r) + \frac{\partial T / \partial r}{\partial T / \partial \rho} \right] \end{aligned} \quad (8.2)$$

Par hypothèse, $\partial T / \partial \rho \leq 0$ et $\partial T / \partial r \geq 0$, si bien que $d\hat{T} / dr$ est négativement proportionnel à la somme d'un terme positif, $\dot{\rho}_\omega(r)$, et d'un terme négatif, $\frac{\partial T / \partial r}{\partial T / \partial \rho}$, tandis que le signe de $d\hat{D} / dr$ ne dépend que de $\dot{\rho}_\omega(r)$.

En ce qui concerne le terme $\dot{\rho}_\omega(r)$, on peut définir une situation de *resserrement* des logements relativement aux emplois lorsque la condition $\dot{r}_\omega \leq 1$ est vérifiée, ce qui équivaut à $\dot{\rho}_\omega \geq 1$. Cela signifie que deux travailleurs dont les emplois sont séparés de $d\rho$ résident en des lieux séparés d'une distance $dr < d\rho$. Symétriquement, une situation de *deserrement* des logements relativement aux emplois est traduite par la condition que $\dot{r}_\omega \geq 1$, équivalente à $\dot{\rho}_\omega \leq 1$.

Concernant le terme $\frac{\partial T / \partial r}{\partial T / \partial \rho}$, remarquons que pour de nombreux réseaux routiers urbains on observe une situation *congestionnée* dans laquelle les coûts (temporel et monétaire), par déplacement et par unité de distance, sont plus élevés dans le centre qu'en périphérie : alors $|\partial T / \partial \rho| \geq \partial T / \partial r$ donc $\frac{\partial T / \partial r}{\partial T / \partial \rho} \geq -1$. Inversement, il est théoriquement possible d'investir dans le système de transport en privilégiant des modes très massifiés et très performants près du centre de l'agglomération, là où les volumes sont plus importants, afin d'assurer $|\partial T / \partial \rho| \leq \partial T / \partial r$, donc $\frac{\partial T / \partial r}{\partial T / \partial \rho} \leq -1$. Nous parlerons de situation de *massification*.

À partir du croisement de ces quatre conditions, il est possible de construire quatre situations typiques mettant en évidence un comportement différent de la distance et du coût de transport (et donc du revenu net) avec la localisation résidentielle. La Table 8.1 résume ces croisements. Sont en *italique* les configurations que nous estimons peu probables, tandis que sont en **gras** les configurations plus probables.

| | Desserrement | Resserrement |
|---------------|----------------------------|--------------------|
| Massification | Quasi-monocentrique | <i>Recentrée</i> |
| Congestion | Déconcentrée | <i>Excentrique</i> |

TABLE 8.1.: Les quatre configurations, résultat du croisement des quatre caractéristiques.

Les données empiriques suggèrent que le desserrement des logements relativement aux emplois restent généralement la règle (pour une revue de ces travaux, voir Huriot et Bourdeau-Lepage, 2009). C'est pourquoi nous avons privilégié ce critère. En le croisant avec les deux critères relatifs au système de transport, nous obtenons deux configurations typiques.

Tout d'abord, une agglomération caractérisée par un *desserrement* des logements relativement aux emplois et une *massification* du transport donne lieu à une configuration que nous avons appelée « *quasi-monocentrique* » : dans ce cas, la distance comme le coût de transport augmentent quand la localisation résidentielle s'éloigne du centre. En effet, $d\hat{D}(r)/dr \geq 0$ et $d\hat{T}(r)/dr \geq 0$ puisque $\dot{\rho}_\omega \leq 1 \leq -\frac{\partial T/\partial r}{\partial T/\partial \rho}$. Autrement dit, lorsque ces conditions sont réunies, distance et coût de transport se comportent, en fonction de r , de la même manière que dans le modèle monocentrique standard⁷.

Ensuite, une agglomération caractérisée par un *desserrement* des logements et une *congestion* du système de transport au centre produit une configuration que nous avons appelée « *déconcentrée* ». Elle présente un intérêt tout particulier puisqu'elle rend compte du constat de Fouchier (1997). La distance y augmente avec la localisation résidentielle, tandis que dans le même temps, le coût y diminue. En effet, $\dot{\rho}_\omega \leq 1$ et $-\frac{\partial T/\partial r}{\partial T/\partial \rho} \leq 1$.

La Figure 8.3 illustre les deux configurations typiques quasi-monocentrique et déconcentrée, telles qu'elles peuvent être obtenues par notre modèle.

Il est toutefois possible d'envisager un resserrement des logements relativement aux emplois. Dans une telle agglomération, les emplois auraient une plus forte

7. Notons d'ailleurs que, dans le modèle monocentrique standard, tous les emplois sont concentrés en un point, si bien que les logements sont nécessairement desserrés par rapport aux emplois. De même, le coût de transport est nul dans le CBD, ce qui correspond à une massification infinie.

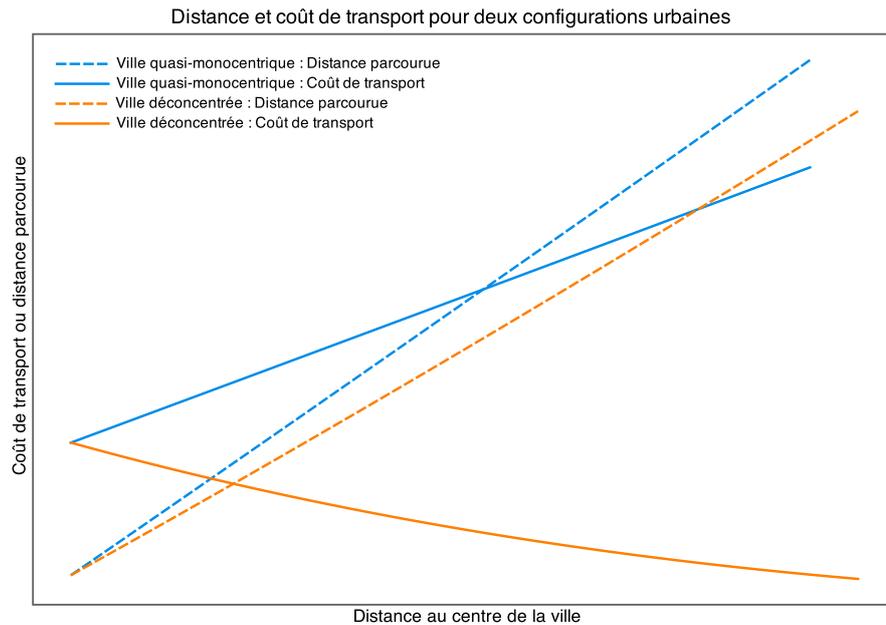


FIGURE 8.3.: Courbes de la distance parcourue et du coût de transport dans deux configurations urbaines typiques.

tendance à l'étalement que les logements. Comme précédemment, deux cas peuvent se présenter.

Tout d'abord, une agglomération présentant un *resserrement* des logements relativement aux emplois et une *congestion* du réseau de transport engendre une configuration que nous avons appelée « *excentrique* » : la distance et le coût de transport y diminuent avec l'éloignement du centre. Autrement dit, cette configuration représente une forte originalité par rapport au modèle standard. Ce type d'organisation est susceptible d'émerger dans une situation où le coût unitaire de transport est partout très élevé et encore plus dans le centre, les emplois étant pour leur part relativement dispersés.

Ensuite, une agglomération caractérisée par un *resserrement* logements et une massification des transports a été appelée « *recentrée* ». Dans ces conditions la distance parcourue par les ménages diminue avec la localisation résidentielle, tandis que le coût de transport augmente.

Des situations mixtes sont envisageables et même probables ; elles conduisent à un coût de transport qui évolue de manière non monotone avec la localisation résidentielle, commençant par décroître, puis augmentant plus loin. Par ailleurs, les conditions que nous avons définies sont suffisantes pour assurer la croissance ou la décroissance du coût de transport avec la localisation résidentielle, mais elles ne sont pas nécessaires, si bien qu'il est possible d'obtenir l'une comme l'autre sans

que l'ensemble des conditions soit réuni.

8.4.2. Villes quasi-monocentrique et déconcentrée dans le cadre simplifié

Avec les hypothèses spécifiques rappelées au 8.3.2, la distance parcourue par un ménage travaillant en ρ s'écrit, en fonction du paramètre ρ :

$$\begin{aligned} D_\omega(\rho) &= r_\omega(\rho) - \rho \\ &= \frac{Y_0}{a} + \frac{a'}{\alpha'a} \tilde{\rho} - \frac{\alpha'a + \beta'a'}{\alpha'a} \rho - \left(\frac{Y_0}{a} + \frac{a'}{\alpha'a} \tilde{\rho} - r_0 \right) \left(1 - \frac{\rho}{\tilde{\rho}} \right)^{\beta'} . \end{aligned} \quad (8.3)$$

De même, le coût de transport subi par un ménage travaillant en ρ s'écrit, en fonction du paramètre ρ :

$$\begin{aligned} T_\omega(\rho) &= a_0 + ar_\omega(\rho) - a'\rho \\ &= Y + \frac{a'}{\alpha'}(\tilde{\rho} - \rho) - \left(Y_0 + \frac{a'}{\alpha'}\tilde{\rho} - ar_0 \right) \left(1 - \frac{\rho}{\tilde{\rho}} \right)^{\beta'} . \end{aligned} \quad (8.4)$$

Les deux expressions (8.3) et (8.4) permettent d'étudier à quelles conditions sur les paramètres les deux situations caractéristiques mises en évidence précédemment dans le cas général peuvent être rencontrées dans ce cadre d'analyse simplifié. En effet, il suffit pour cela d'étudier le signe de $dD_\omega/d\rho$ et $dT_\omega/d\rho$, puisque $r_\omega(\rho)$ est croissante. Or :

$$\frac{dD_\omega}{d\rho} = -\frac{B}{\tilde{\rho}} + \beta' \frac{C}{\tilde{\rho}} \left(1 - \frac{\rho}{\tilde{\rho}} \right)^{-\alpha'} - 1, \quad (8.5)$$

$$\frac{dT_\omega}{d\rho} = \frac{\beta'}{\tilde{\rho}} aC \left(1 - \frac{\rho}{\tilde{\rho}} \right)^{-\alpha'} - \frac{a'}{\alpha'}. \quad (8.6)$$

À partir de ces expressions, on cherche des conditions suffisantes sur les coûts marginaux de transport a et a' et sur le rayon limite des emplois ρ_f pour qu'émergent les deux configurations typiques quasi-monocentrique et déconcentrée. En effet, ces trois paramètres sont ceux qui permettent le plus simplement de rendre compte des caractères portant sur le réseau (massification ou congestion) et sur l'étalement relatif des emplois et des logements (desserrement ou resserrement).

Les deux configurations retenues étant caractérisées par le fait que la distance parcourue augmente avec la localisation résidentielle, on cherche des conditions

suffisantes pour que $dD_\omega/d\rho \geq 0$, c'est-à-dire :

$$\frac{dD_\omega}{d\rho} = \dot{r}_\omega(\rho) - 1 \geq \dot{r}_\omega(0) - 1 = \frac{\beta'}{a\tilde{\rho}}(Y - a_0 - ar_0) - 1 \geq 0 \quad (8.7)$$

Autrement dit, une condition suffisante est :

$$\rho_f \left(\frac{\lambda R_A}{aN} + 1 \right) \leq \beta' \frac{Y - a_0 - ar_0}{a} \quad (8.8)$$

En choisissant une valeur faible de a , la condition peut être remplie. Plus précisément, il faut :

$$a \leq \frac{\beta' Y_0 - \frac{\lambda R_A}{b}}{\rho_f + \beta' r_0} \quad (8.9)$$

Ville quasi-monocentrique Cette configuration est caractérisée par le fait que le coût de transport augmente avec la localisation résidentielle. Or, pour que $dT_\omega/d\rho \geq 0$, il suffit que :

$$\frac{\beta'}{\tilde{\rho}} a C x^{-\alpha'} - \frac{a'}{\alpha'} \geq 0, \quad (8.10)$$

avec $x = 1 - \rho/\tilde{\rho}$. Cela revient à $a'\tilde{\rho}x^{\alpha'} \leq \alpha'\beta'aC$ en tout point x . Puisque $x \leq 1$, il suffit que $a'\tilde{\rho} \leq \alpha'\beta'aC$, soit encore :

$$\rho_f \left(\frac{\lambda R_A}{aN} + 1 \right) \leq \beta' \frac{Y - a_0 - ar_0}{a'} \quad (8.11)$$

En choisissant de n'agir que sur les coûts marginaux de transport et sur le rayon limite des emplois, il apparaît que la condition est remplie si a' est suffisamment faible par rapport à a , ce qui correspond bien à une forme de *massification* des transports dans le centre comme nous l'avons identifié dans le cas général, et ρ_f suffisamment faible, entraînant un *desserrement* des logements par rapport aux emplois.

Ville déconcentrée La ville déconcentrée étant caractérisée par le fait que le coût de transport est décroissant avec la localisation résidentielle, on cherche maintenant la condition pour laquelle $dT_\omega/d\rho \leq 0$, donc $a'\tilde{\rho}x^{\alpha'} \leq \alpha'\beta'aC$ en tout point x . Puisque $x \geq t = R_A/R_0 = R_A/(R_A + \frac{a}{\lambda}N)$, il suffit que $a'\tilde{\rho}t \leq \alpha'\beta'aC$, ce qui s'écrit encore :

$$\rho_f \left(\frac{\lambda R_A}{aN} + 1 \right) (t^{\alpha'} - \beta') \geq \alpha'\beta' \frac{Y - a_0 - ar_0}{a'} \quad (8.12)$$

En choisissant là aussi de n'agir que sur a et a' d'une part, et sur ρ_f d'autre part, il apparaît que la condition est remplie d'abord si a n'est pas trop élevé (pour assurer

$t^{\alpha'} - \beta' \geq 0$) et ensuite, si a' est suffisamment élevé par rapport à a ce qui traduit la présence d'un réseau de transport *congestionné* au centre de la ville, en tenant compte du fait que ρ_f est relativement faible, pour assurer un *desserrement* des logements par rapport aux emplois.

8.4.2.1. Distance et coût de transport dans une ville « déconcentrée » : un paradoxe expliqué ?

On vient de le voir, le coût de transport, dans une ville « déconcentrée », est plus élevé (ou sensiblement identique pour les cas intermédiaires) pour les ménages résidant au centre de l'agglomération que pour ceux résidant à distance du centre, autrement dit en zone périurbaine. Les conditions suffisantes (8.12) et (8.8) sont compatibles et il est donc tout à fait possible de trouver un triplet (a, a', ρ_f) tel que les conditions soient remplies simultanément. C'est ce qu'illustre la Figure 8.4.

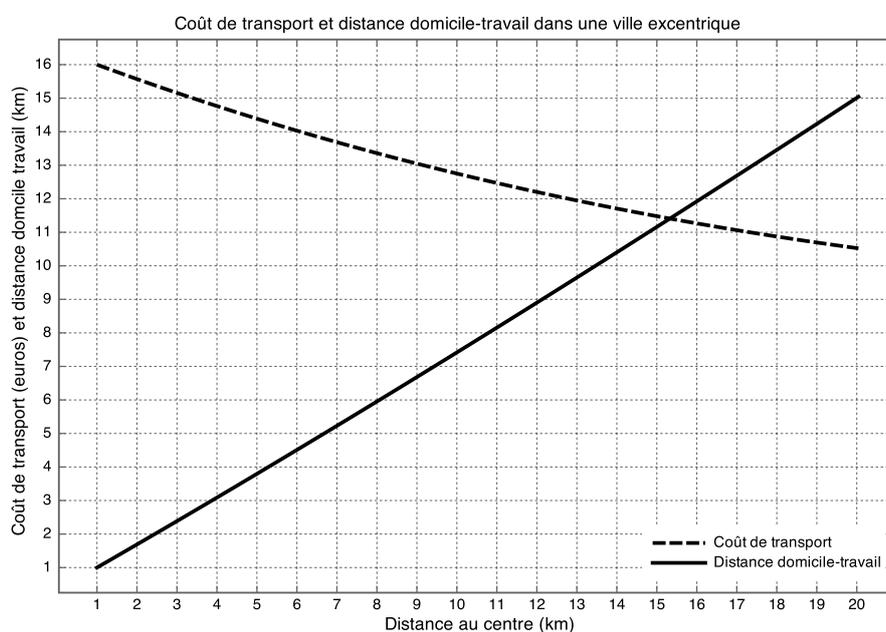


FIGURE 8.4.: Courbe de coût de transport et de distance domicile-travail avec les paramètres $(a, a', \rho_f) = (0,5 \text{ €/km} ; 3 \text{ €/km} ; 5 \text{ km})$.

Le coût de transport tel qu'il est modélisé ici mesure surtout les coûts temporels liés au transport, ainsi que les coûts monétaires variables liés à la distance parcourue et à la congestion, tels que la surconsommation de carburant dans les embouteillages. Le résultat que nous venons de présenter montre donc qu'avec une prise en compte explicite de la distribution des emplois dans l'espace, dans un cadre monocentrique (opposé à polycentrique), et une modélisation, même extrêmement

simple comme ici, d'un coût unitaire de transport variable, il est possible de fournir une base théorique aux résultats empiriques de Fouchier (1998) pour l'Île-de-France, ou de Kahn (2007) pour les villes américaines. Plus les ménages vivent loin du centre de l'agglomération, c'est-à-dire dans un environnement à faible densité, et plus les distances parcourues pour se rendre sur le lieu d'emploi sont élevées, mais dans le même temps, les temps de parcours diminuent ou du moins restent stables. Ces auteurs justifient ce résultat empirique en notant, comme Genre-Grandpierre (2007), que les vitesses de déplacement en périphérie sont bien plus élevées que celles du centre, du fait d'une part des infrastructures de transport mises en œuvre au centre (rues et boulevards urbains) et en périphérie (voies rapides), et d'autre part de la présence de hauts niveaux de congestion dans le centre des agglomérations. Pourtant, le modèle monocentrique standard, pierre angulaire de l'économie urbaine théorique, est incapable de rendre compte de cette dissociation entre distance et temps de transport.

Notre modèle permet, comme nous venons de le voir, de mettre en évidence l'apparition de cette dissociation qui peut constituer un moteur à l'étalement urbain. En effet, en s'éloignant du centre, les ménages bénéficient d'un temps de transport plus faible⁸ et sont donc incités à s'éloigner du centre, afin de bénéficier de plus grandes surfaces de logements, ainsi que de temps de transport réduits. En fait, notre modèle le confirme, c'est bien la présence de hauts niveaux de congestion dans le centre ou d'un réseau de transport public lent et à faible capacité d'une part, et d'autre part l'existence d'un coût marginal faible en périphérie, du fait de la présence d'infrastructures de grande capacité et à haute vitesse de circulation qui engendrent ce résultat.

En termes de politiques de transport, ce résultat⁹ suggère donc que l'amélioration des conditions de circulation dans le centre de l'agglomération, en particulier par la mise en place de réseaux de transport collectif performant peut être de nature à atténuer voire à supprimer cette décroissance des coûts de transport avec la distance. Dit autrement, en rendant au centre de l'agglomération un avantage qu'elle a en partie perdu, il est possible d'inciter les ménages à s'en rapprocher sans pour autant les contraindre par des mesures, généralement mal vécues, de hausse des coûts de transport. Au contraire, il nous semble que des mesures aboutissant à une diminution globale de la vitesse de déplacements au centre des villes (réduction de la capacité routière se traduisant exclusivement par une augmentation de l'es-

8. Il peut être éventuellement compensé par des coûts monétaires croissants, en particulier en termes de coûts fixes liés à l'achat d'un véhicule, à son entretien, etc. Mais Coulombel *et al.* (2007) ont justement montré que les ménages périurbains sous-estiment le coût de transport complet et sont donc incités à s'éloigner du centre plus qu'ils ne le devraient.

9. Toutefois, compte tenu du caractère théorique du modèle, la prudence est de mise vis-à-vis de ces recommandations opérationnelles.

pace réservé aux piétons, sans amélioration de la qualité de service en transports collectifs) est de nature à accentuer l'étalement urbain.

8.5. Un étalement des emplois favorisant un usage raisonné des transports ?

Nous avons, dans la section précédente, montré que la prise en compte de la dispersion des emplois permettait de mettre en évidence un mécanisme favorisant potentiellement l'étalement urbain. Nous allons maintenant nous intéresser aux conséquences de l'étalement des emplois, notamment en termes d'usage des transports. Notre analyse se focalise notamment sur la question des distances parcourues, puisqu'il s'agit, comme nous l'avons noté au 8.2.2 certains auteurs suggèrent que la dispersion des emplois, souvent mal pris en compte par des analyses focalisées sur la question des ménages, peut avoir des effets importants sur le phénomène d'étalement urbain.

8.5.1. L'étalement des emplois modifie l'intensité d'usage des transports

Nous venons de le voir, dans un contexte où la distribution des emplois urbains est explicitement décrite, distance parcourue et temps de transport n'évoluent pas nécessairement de la même manière. De même, l'intensité d'usage des transports, c'est-à-dire le nombre d'usagers transitant par le réseau de transport à une certaine distance du centre, peut présenter un profil particulier. Dans le cadre du modèle monocentrique standard, le profil est simple : l'intensité augmente avec la proximité du centre, de plus en plus de ménages se trouvant sur le réseau à mesure que l'on se rapproche du centre.

Au contraire, comme l'indique Wheaton (2004), la distribution des emplois modifie ce schéma simple : en reprenant les notations précédentes, le nombre de ménages utilisant le réseau à une distance r du centre est :

$$Q(r) = [N - \tilde{H}(r)] - [N - \tilde{F}(r)] = \tilde{F}(r) - \tilde{H}(r) \quad (8.13)$$

Où les fonctions \tilde{F} et \tilde{H} désignent les fonctions cumulées des emplois et des rési-

dences, respectivement, prolongées sur $[0, +\infty[$, autrement dit : $Q(r)$ s'écrit :

$$Q(r) = \begin{cases} F(r) & \text{sur } [0, r_0[\\ F(r) - H(r) & \text{sur } [r_0, \rho_f] \\ N - H(r) & \text{sur }]\rho_f, r_A] \\ 0 & \text{sur }]r_A, +\infty[\end{cases} \quad (8.14)$$

Q est continue sur $[0, +\infty[$ et dérivable par morceaux sur chacun des intervalles ci-dessus, et la dérivée prend les valeurs suivantes :

$$Q'(r) = \begin{cases} f(r) > 0 & \text{sur } [0, r_0[\\ f(r) - h(r) & \text{sur } [r_0, \rho_f] \\ -h(r) < 0 & \text{sur }]\rho_f, r_A] \\ 0 & \text{sur }]r_A, +\infty[\end{cases} \quad (8.15)$$

Comme le montrent les relations précédentes, le profil de l'intensité d'usage des transports dépend des valeurs relatives de f et de h , pour les différentes localisations r . Nous avons choisi de nous focaliser sur deux hypothèses particulières concernant la distribution des emplois : f constante sur l'ensemble de la zone d'emploi d'une part, et f décroissante jusqu'à 0 en ρ_f d'autre part.

8.5.1.1. Différentes configurations urbaines

Les deux hypothèses donnent lieu à trois modalités. La Figure 8.5 illustre, de manière qualitative, les trois situations possibles dans le cas de l'hypothèse de distribution homogène, tandis que la Figure 8.6 fait de même pour les trois situations possibles dans le cas de l'hypothèse de distribution décroissante. Dans les deux cas, nous avons représenté, sous les figures, une vue en plan de l'organisation spatiale de la ville, qui permet de mieux caractériser les différents types de structures. Dans les figures, la légende E désigne une zone exclusive d'emploi, ME, une zone mixte à dominante emploi, MR une zone mixte à dominante résidentielle, et R une zone résidentielle exclusive.

Les Figures 8.5(b) et 8.6(b) mettent en évidence des structures urbaines semblables. En revanche, les autres configurations sont toutes différentes, et correspondent à des fonctionnements urbains distincts, amenant un usage des transports également différent. On notera en particulier que pour des valeurs moyennes de b , la configuration urbaine comprend un CBD d'une part, et une couronne où l'emploi domine ; entre ces deux zones d'emplois se situe une zone à dominante résidentielle (Figure 8.5(c)). Cette dernière configuration n'est pas sans rappeler certaines configurations obtenues par Ota et Fujita (1993). La Figure 8.6(c) met en évidence deux

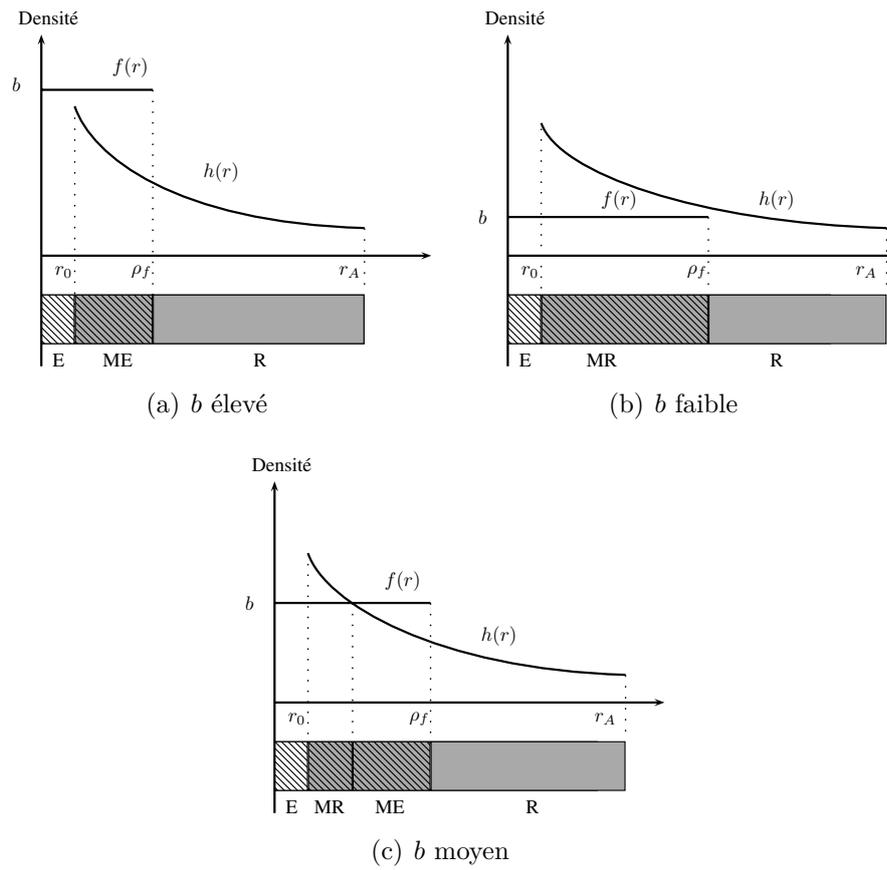


FIGURE 8.5.: Trois situations typiques lorsque f est constante.

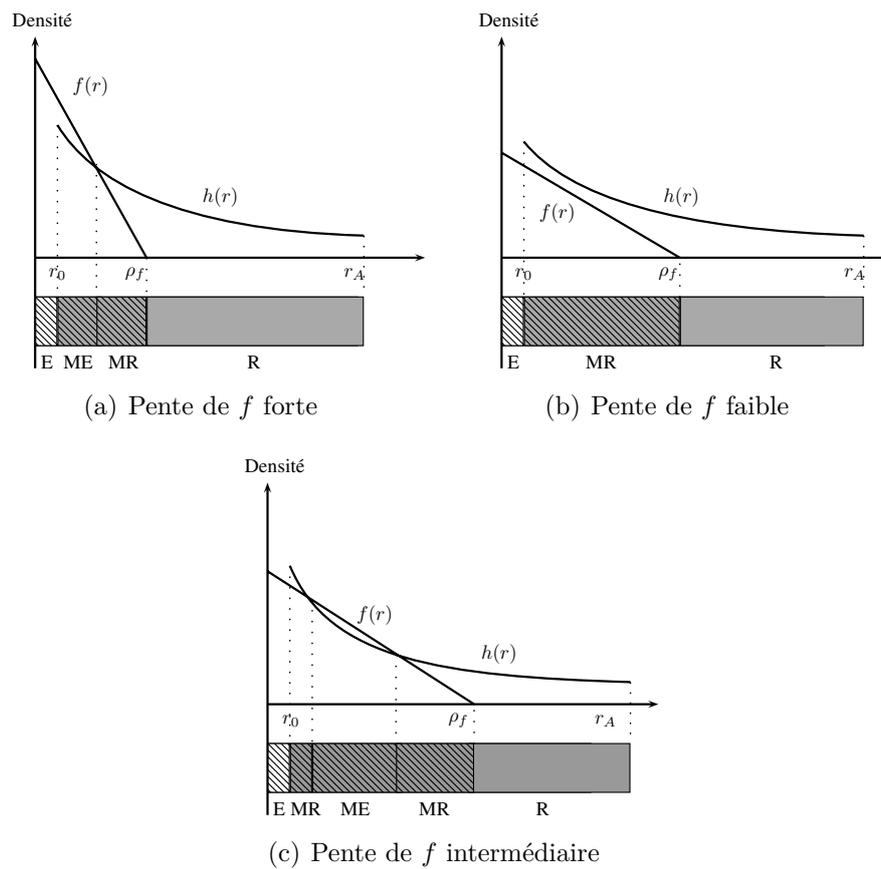


FIGURE 8.6.: Trois situations typiques lorsque f est décroissante.

points de changement de dominante :

8.5.1.2. Profils d'intensité d'usage des transports

On peut, à partir de la relation (8.15), obtenir les courbes de profils d'intensité d'usage du réseau de transport au sein de l'agglomération. Ainsi, les Figures 8.7 et 8.8 présentent les profils d'intensité d'usage pour les hypothèses et modalités précédentes¹⁰.

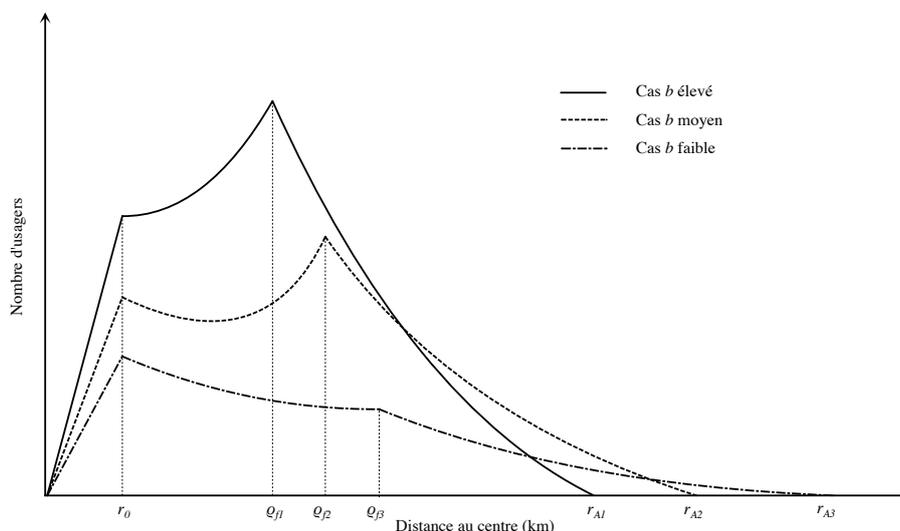


FIGURE 8.7.: Profils d'intensité d'usage pour f constante et trois valeurs de b .

On constate dans un premier temps que l'existence d'une distribution des emplois modifie en profondeur le profil de l'intensité d'usage, puisque le point d'intensité maximale ne se situe plus au centre de la ville, comme dans le modèle monocentrique, mais à une distance variable, dépendant des paramètres affectant le niveau de décentralisation des emplois. Ainsi, dans le cas d'une distribution homogène des emplois, le point d'intensité maximale d'une part diminue en niveau lorsque le rayon de la zone d'emploi augmente, et d'autre part se déplace du point marquant la fin de la zone d'emploi au point marquant le début de la zone résidentielle.

Autrement dit, l'étalement des emplois entraîne deux effets conjoints : tout d'abord, en dispersant les lieux de destination des travailleurs, l'étalement diminue globalement la concentration des usagers sur le réseau de transport. Celui-ci sera vraisemblablement moins congestionné que lorsque l'emploi est concentré. Ensuite, la zone la plus congestionnée, située au niveau de l'accès à la zone d'emploi

10. Il s'agit de schémas théoriques et non de profils réels : l'objectif est de présenter les différents profils possibles de la manière la plus claire afin d'insister sur les caractéristiques qui les distinguent. Par conséquent, les échelles ne sont pas respectées, seul l'ordre des variables et des fonctions l'est, ainsi que les sens de variation.

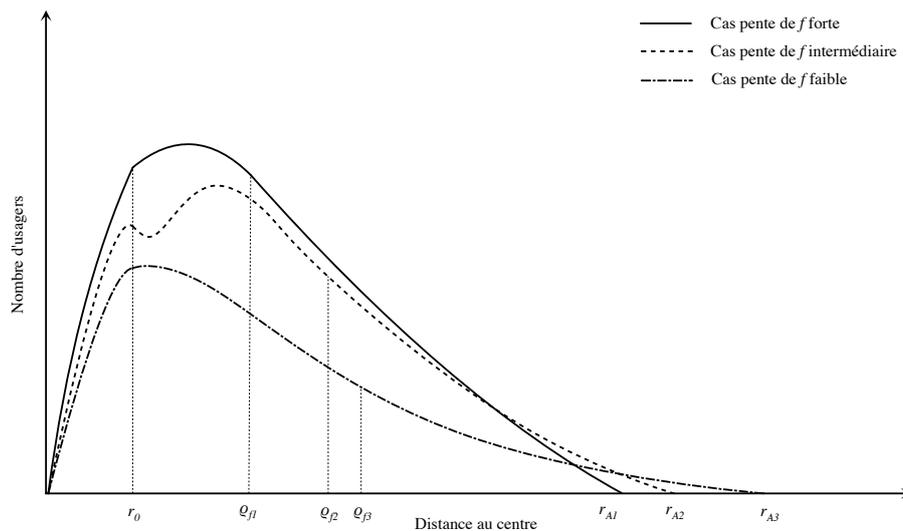


FIGURE 8.8.: Profils d'intensité d'usage pour f linéairement décroissante et deux valeurs de la pente.

lorsque celui-ci est concentré, se déplace au niveau de l'accès à la zone exclusive des emplois, le CBD.

C'est finalement toujours l'accès au lieu rassemblant le plus d'emplois qui constitue un goulot d'étranglement. Toutefois, ce déplacement a son importance, car ce ne sont pas les mêmes travailleurs qui subissent les plus hauts niveaux de congestion dans les deux cas. Lorsque l'emploi est concentré, ce sont majoritairement les travailleurs résidant en périphérie, et travaillant en bordure de la zone d'emplois qui subiront la plus forte congestion, tandis que les usagers travaillant et résidant plus au centre, rencontreront des niveaux moins élevés. Au contraire, en présence d'un étalement des emplois, les travailleurs du centre sont moins bien lotis que ceux de la périphérie. Dit autrement, si l'étalement semble, dans ce cadre, bénéfique pour tous (en termes de niveaux de congestion, et donc de perte de temps), les gains ne sont cependant pas répartis de manière égale pour tous les ménages. Nous avons vu au chapitre 6 que le niveau d'utilité des ménages était croissant avec la localisation résidentielle, et ce d'autant plus que les emplois étaient étalés. Or ce résultat était obtenu avec une technologie de transport sans congestion. Nous voyons ici que la congestion, considérée de manière exogène, accentue encore ce phénomène.

8.5.2. Étalement des emplois et déplacements domicile-travail

Les différences constatées en ce qui concerne l'intensité d'usage des transports mettent en évidence un avantage potentiel d'une ville où l'emploi serait étalé : la congestion y apparaît plus faible. Toutefois, nous n'avons pas encore abordé

la question des distances domicile-travail, des coûts de transport associés, ainsi que de l'évolution du surplus des ménages dans le cadre d'une comparaison des développements urbains étalé et compact. C'est ce que nous nous proposons de faire ici. Notre analyse porte essentiellement sur la question des distances parcourues et des coûts de transport associés. Elle est complétée par une analyse des questions d'équité spatiale qui y sont liées.

À partir d'une situation d'équilibre de notre modèle pour des valeurs données de la population et des caractéristiques de la distribution des emplois (densité et étendue de la zone d'emplois), nous nous proposons de comparer deux nouvelles situations d'équilibre correspondant à des configurations urbaines différentes.

La première correspond à un développement compact : le rayon de la zone d'emplois est réduit, tandis que sa densité augmente. La seconde correspond à un développement étalé, où la zone d'emplois s'étend, avec une baisse de la densité des emplois. La Figure 8.9 illustre les deux types d'évolution envisagés.

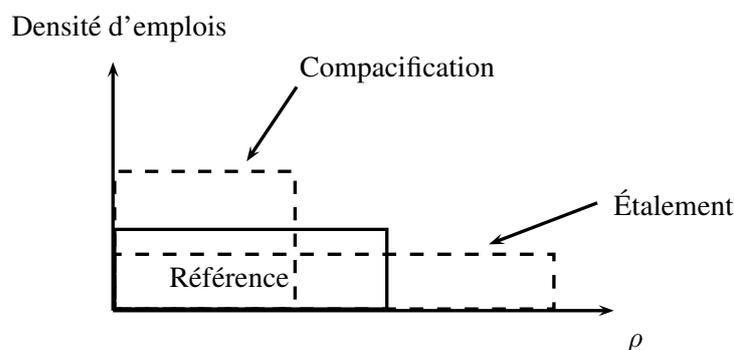


FIGURE 8.9.: Schéma présentant les deux évolutions envisagées de la zone d'emploi.

En termes de notations, nous comparons d'une part un équilibre initial, défini par une population N , et une zone d'emplois caractérisée par ρ_{f0} et b_0 , et d'autre part deux équilibres : le premier pour lequel la zone d'emplois est caractérisée par $\rho_{f1} < \rho_{f0}$ et $b_1 > b_0$ (compacification des emplois) et le second pour lequel elle est caractérisée par $\rho_{f2} > \rho_{f0}$ et $b_2 < b_0$ (étalement des emplois).

8.5.2.1. Le cadre d'analyse : du surplus individuel à la variation compensatrice globale

Nous souhaitons disposer d'un cadre d'analyse permettant de comparer les configurations urbaines définies plus haut. En effet, la structure spatiale des emplois a potentiellement une influence forte sur le revenu des ménages, par l'intermédiaire

des économies d'agglomération d'une part ¹¹, et par l'intermédiaire des coûts de transport subis par les ménages d'autre part. De manière à mesurer, en termes monétaires, l'incidence, pour un ménage ou pour l'ensemble des ménages, d'une modification de la structure urbaine, nous mobilisons la notion de surplus. Celui-ci s'exprime comme la différence entre la variation de revenu d'un ménage (que l'on suppose liée au passage d'une structure urbaine à l'autre, et qui est éventuellement nulle) et la variation compensatrice de revenu (*VCR*), c'est-à-dire la variation de revenu qui laisserait l'utilité du ménage inchangée.

Surplus individuel Autrement dit, le surplus σ qu'un ménage indexé n dégage du passage d'une situation 1 à une situation 2 s'écrit ¹² :

$$\sigma(n) = Y_2 - (Y_1 + VCR(n)) \tag{8.16}$$

Or, la variation compensatrice de revenu s'exprime grâce à la fonction d'utilité indirecte. En effet, avec une fonction d'utilité Cobb-Douglas, on peut écrire, pour un ménage indexé n :

$$\begin{aligned} (Y_1 + VCR(n)) - T(\rho_2(n), r_2(n)) &= R_2(n)^{\beta'} \left(\frac{V_1}{v_0} \right)^{1/(\alpha+\beta)} \\ Y_2 - T(\rho_2(n), r_2(n)) &= R_2(n)^{\beta'} \left(\frac{V_2}{v_0} \right)^{1/(\alpha+\beta)} \end{aligned}$$

On en tire la valeur de la variation compensatrice de revenu :

$$VCR(n) = T(\rho_2(n), r_2(n)) - Y_1 + (Y_2 - T(\rho_2(n), r_2(n))) \left(\frac{V_1}{V_2} \right)^{1/(\alpha+\beta)} \tag{8.17}$$

En introduisant l'équation (8.17) dans (8.16), on obtient la valeur du surplus dégagé par le ménage n , lors du passage d'une situation 1 à une situation 2 :

$$\sigma(n) = \left(Y_2 - T(\rho_2(n), r_2(n)) \right) \frac{V_2^{1/(\alpha+\beta)} - V_1^{1/(\alpha+\beta)}}{V_2^{1/(\alpha+\beta)}} \tag{8.18}$$

11. Nous présenterons au 8.5.2.4 une modélisation du rôle joué par les économies d'agglomération sur le revenu des ménages.

12. La raison du choix de l'index des ménages plutôt que la localisation résidentielle ou de l'emploi est double. Tout d'abord, elle tient au fait que dans le cas d'une modification de la distribution des emplois, le ménage qui occupe l'emploi situé en ρ ou qui réside en r n'est pas le même, dans l'une ou l'autre des situations. Par ailleurs, contrairement au modèle monocentrique standard dans lequel tous les ménages atteignent le même niveau à l'équilibre, et sont donc indifférents entre toutes les localisations résidentielles (ce qui autorise à supposer qu'ils restent à la même place, cf. Bernard, 1998), ce n'est pas le cas dans notre modèle.

De plus, comme nous l'avons vu au chapitre précédent, avec les spécifications rappelées au 8.3.2, la rente foncière unitaire payée par un ménage donné ne dépend ni du salaire ni de la distribution des emplois. Autrement dit, ici : $R_2(n) = R_1(n)$. Ceci permet de simplifier singulièrement l'expression du surplus d'un ménage :

$$\sigma(n) = [Y_2 - Y_1] - [T(\rho_2(n), r_2(n)) - T(\rho_1(n), r_1(n))] \quad (8.19)$$

Pour étudier plus en détail l'expression du surplus, il est utile de remplacer le coût de transport par son expression dans le cadre des hypothèses du 8.3.2, à savoir $T(\rho, r) = a_0 + ar - a'\rho$. On obtient donc, tous calculs faits :

$$\sigma(n) = (Y_2 - Y_1) + (\rho_{f2} - \rho_{f1}) \frac{a' \lambda R_0 - an}{a' a N} \left[\left(\frac{\lambda R_0}{\lambda R_0 - an} \right)^{\alpha'} - 1 \right] \quad (8.20)$$

On peut, à partir de cette expression, tirer une première information : en supposant que le salaire des ménages est inchangé entre les situations 1 et 2, alors $\sigma(n) \geq 0$ si et seulement si $\rho_{f2} \geq \rho_{f1}$, car le terme en $\rho_{f2} - \rho_{f1}$ est positif. Autrement dit, toutes choses égales par ailleurs, le surplus individuel augmente avec l'étalement des emplois.

En ce qui concerne l'évolution du surplus avec n , c'est-à-dire la manière dont le surplus évolue avec l'éloignement au centre des ménages, on montre aisément (les calculs figurent en Annexe 8.A.1) que le surplus croît jusqu'à une valeur n^* puis décroît, avec :

$$n^* = \frac{\lambda R_0}{a} (1 - \beta^{1/\alpha'}) \quad (8.21)$$

qui est positif, car $\beta' < 1$. De plus, l'Annexe 8.A.1 montre également que $n^* < N$ à la condition que :

$$N > \frac{\lambda R_A}{a} \left(\frac{1}{\beta^{1/\alpha'}} - 1 \right) \quad (8.22)$$

Cette condition est donc remplie pour une ville très peuplée, et dont le coût unitaire de transport dans la zone résidentielle est élevé, correspondant à un niveau élevé de congestion, même en dehors du centre de la ville. Dans une telle agglomération, donc, l'avantage que retireraient les ménages d'un étalement des emplois, toutes choses égales par ailleurs, augmente d'abord avec la distance au centre, pour diminuer ensuite. Il existe donc une distance pour laquelle les avantages attendus sont maximaux.

Surplus collectif On calcule ensuite le surplus agrégé des ménages, en intégrant sur l'ensemble des ménages¹³ :

$$\begin{aligned}
 \mathbf{S} &= \int_0^N \sigma(n) dn \\
 &= N(Y_2 - Y_1) + \dots \\
 &\quad \dots (\rho_{f2} - \rho_{f1}) \frac{a'}{\alpha'} \int_0^N \frac{\lambda R_0 - an}{aN} \left[\left(\frac{\lambda R_0}{\lambda R_0 - an} \right)^{\alpha'} - 1 \right] dn \quad (8.23)
 \end{aligned}$$

Le calcul explicite de l'intégrale donne le résultat suivant :

$$\int_0^N \frac{\lambda R_0 - an}{aN} \left[\left(\frac{\lambda R_0}{\lambda R_0 - an} \right)^{\alpha'} - 1 \right] dn = \frac{\eta(\alpha') - \eta(0)}{aN} \geq 0$$

où la fonction η est définie par :

$$\epsilon \mapsto \eta(\epsilon) = -\frac{(\lambda R_0)^\epsilon}{a(2-\epsilon)} \left[(\lambda R_0)^{2-\epsilon} - (\lambda R_A)^{2-\epsilon} \right]$$

η est négative et décroissante en valeur absolue pour $\epsilon \leq 1$.

La remarque formulée pour le surplus individuel se maintient pour le surplus collectif : si le salaire des ménages est inchangé, le surplus augmente si $\rho_{f2} \geq \rho_{f1}$.

Variation compensatrice de revenu globale De manière à établir une référence pour la comparaison des deux scénarios d'évolution urbaine que sont la ville compacte et la ville étalée, nous définissons la variation compensatrice de revenu globale (\overline{VCR}) comme la somme (identique pour tous les ménages) qu'il faut ajouter au salaire des ménages en situation 1 pour annuler le surplus agrégé du passage de 1 à 2. Or, $\mathbf{S} = 0$ se traduit par :

$$Y_1 = Y_2 + \frac{\rho_{f2} - \rho_{f1}}{N} \frac{a'}{\alpha'} \frac{\eta(\alpha') - \eta(0)}{aN}$$

13. En toute rigueur, le surplus agrégé des ménages ne peut être retenu comme mesure du bien-être collectif que si l'on dispose d'une fonction d'utilité collective de type Rawlsien, d'une part, et d'autre part s'il est possible d'effectuer une redistribution entre les ménages par des transferts forfaitaires de revenu, sans distorsion du système de prix. C'est le cas dans le cadre du modèle monocentrique standard, puisque tous les ménages y atteignent le même niveau d'utilité et touchent le même revenu (aucune redistribution n'est donc nécessaire). Dans notre modèle, un transfert monétaire entre les ménages modifierait nécessairement l'équilibre des localisations. Toutefois, nous conservons le surplus comme mesure du bien-être, tout en lui adjoignant une analyse des inégalités à l'équilibre.

Par conséquent, pour le passage d'une situation 1 caractérisée par ρ_{f1} à une situation 2 caractérisée par ρ_{f2} , la variation compensatrice de revenu globale est :

$$\overline{VCR} = \frac{\rho_{f2} - \rho_{f1}}{N} \frac{a' \eta(\alpha') - \eta(0)}{\alpha' a N} \quad (8.24)$$

Cette valeur va nous permettre de comparer les deux situations envisagées au début de cette section, à savoir pour une population N donnée, une configuration des emplois plus compacte et une configuration plus étalée que la situation de référence. En effet, elle permet d'analyser les différences entre plusieurs configurations entre lesquelles les ménages, pris dans leur ensemble, sont indifférents.

La Table 8.2 présente, par rapport à la situation de référence, les valeurs de la variation compensatrice de revenu et du revenu brut correspondant dans les deux situations contrastées décrites précédemment. Nous avons fait figurer les résultats pour deux valeurs du coût marginal de transport dans le centre, correspondant, pour $a' = a = 1$ €/km, à une absence de congestion dans le centre, ce qui mène à une ville quasi-monocentrique et, pour $a' = 2$ €/km, à un niveau important de congestion centrale, qui mène potentiellement à une ville déconcentrée, suivant les valeurs de ρ_f . Le coût fixe est, conformément aux éléments présentés au 7.2.1.2 du chapitre 7, supposé dépendre de la différence entre le coût marginal de transport dans le centre et dans la zone résidentielle et de la valeur de ρ_f : $a_0 = \bar{a}_0 + \rho_f(a' - a)$, avec $\bar{a}_0 = 3$ €. On prend $\rho_f = 5$ km pour la situation de référence, et on suppose que la taille de la zone d'emplois joue sur le coût fixe de transport \bar{a}_0 par l'intermédiaire d'un coût de stationnement (temporel autant que monétaire) de telle sorte que a_0 reste constant lorsque ρ_f varie. La configuration compacte est caractérisée par $\rho_f = 2$ km et la configuration étalée par $\rho_f = 8$ km. Enfin, nous avons fixé $r_0 = 1$ km ; les autres valeurs numériques sont identiques à celles employées dans le chapitre précédent.

| Réseau | Situation | \overline{VCR} (€/mois) | Y (€/mois) |
|-----------------|-----------|---------------------------|--------------|
| Sans congestion | Référence | 0 | 2 500 |
| | Compacte | 26,5 | 2 526,5 |
| | Étalée | -26,5 | 2 473,5 |
| Congestion | Référence | 0 | 2 500 |
| | Compacte | 53,0 | 2 553 |
| | Étalée | -53,0 | 2 447 |

TABLE 8.2.: Variation compensatrice de revenu globale et revenu brut en situation de référence et pour les deux situations équivalentes.

Cette table se lit de la manière suivante : en présence de congestion dans le centre de la ville, pour que, dans leur ensemble, les ménages soient indifférents entre la

configuration étalée et la configuration compacte, il faut que le revenu mensuel qu'ils recevraient dans la ville compacte soit supérieur d'environ 100 € à celui qu'ils toucheraient dans la ville étalée. Rappelons ici que l'existence d'économies d'agglomération liées à la proximité géographique peut justifier une telle différence de revenu entre une ville à zone d'emplois compacte et une ville à zone d'emplois étalée.

8.5.2.2. Comparaison des trois configurations

Il est possible, à l'aide de la variation compensatrice de revenu globale, de comparer l'extension urbaine et les caractéristiques de l'usage des transports — distance domicile-travail, coût de transport, intensité d'usage — pour des configurations compacte ou étalée.

La Table 8.3 fournit les valeurs du rayon d'urbanisation, de la distance domicile-travail moyenne et du coût de transport moyen des ménages, en absolu ainsi que relativement au revenu brut. On constate que pour des transformations à surplus

| Réseau | Situation | r_A (km) | \bar{D} (km) | \bar{T} (€) | Part revenu |
|-----------------|-----------|------------|----------------|---------------|-------------|
| Sans congestion | Référence | 18,24 | 6,51 | 9,51 | 7,6 % |
| | Compacte | 18,19 | 8,02 | 11,02 | 8,7 % |
| | Étalée | 18,28 | 4,99 | 7,99 | 6,4 % |
| Congestion | Référence | 17,92 | 6,30 | 11,80 | 9,4 % |
| | Compacte | 17,83 | 7,83 | 14,83 | 11,6 % |
| | Étalée | 18,02 | 4,77 | 8,77 | 7,2 % |

TABLE 8.3.: Rayon d'urbanisation et caractéristiques de l'usage des transports en situation de référence et pour les deux situations équivalentes.

nul, le rayon d'urbanisation varie peu d'une situation à l'autre, même si la configuration à emploi compact tend à donner lieu à une ville moins étendue que la configuration à emploi étalé. En revanche, la distance domicile-travail moyenne tend clairement à diminuer avec l'étalement des emplois, de même que le coût de transport, et ce, en valeur absolue ainsi qu'en valeur relative par rapport au revenu.

Ces résultats se limitent aux déplacements domicile-travail, et ne présentent pas de ce que pourrait donner une évaluation des distances quotidiennes parcourues par les ménages, tenant compte des trajets liés aux achats et aux loisirs. Toutefois, nos résultats suggèrent que pour des situations *indifférentes pour les ménages considérés dans leur ensemble*, la configuration compacte engendre des déplacements domicile-travail plus longs et plus coûteux (en particulier en présence de congestion dans le centre). Par ailleurs, le gain, sur l'extension urbaine, procuré par une configuration compacte est relativement minime.

Les Figures 8.10 et 8.11¹⁴ présentent, pour le cas avec congestion centrale, les distances domicile-travail et le coût de transport des ménages, en fonction de leur localisation résidentielle. Ces résultats peuvent sembler en apparence contradiction avec ceux de Talbot (2001), qui montrent un accroissement des distances domicile-travail constatées. Toutefois, d'une part ces résultats portent sur l'ensemble du pays, et non seulement sur la région Île-de-France qui présente, plus que d'autres agglomérations, des caractéristiques monocentriques. D'autre part, notre étude porte sur le rôle de l'étalement des emplois dans l'évolution des distances domicile-travail, et nous comparons des situations équivalentes, toutes choses égales par ailleurs. Or, les choses sont rarement égales par ailleurs. D'autres évolutions conjointes sont possibles, en particulier des coûts de transport et des goûts des ménages : préférence pour la nature ou au contraire pour la centralité, qui engendrent de nombreux trajets pendulaires inverses (Aguiléra *et al.*, 2009).

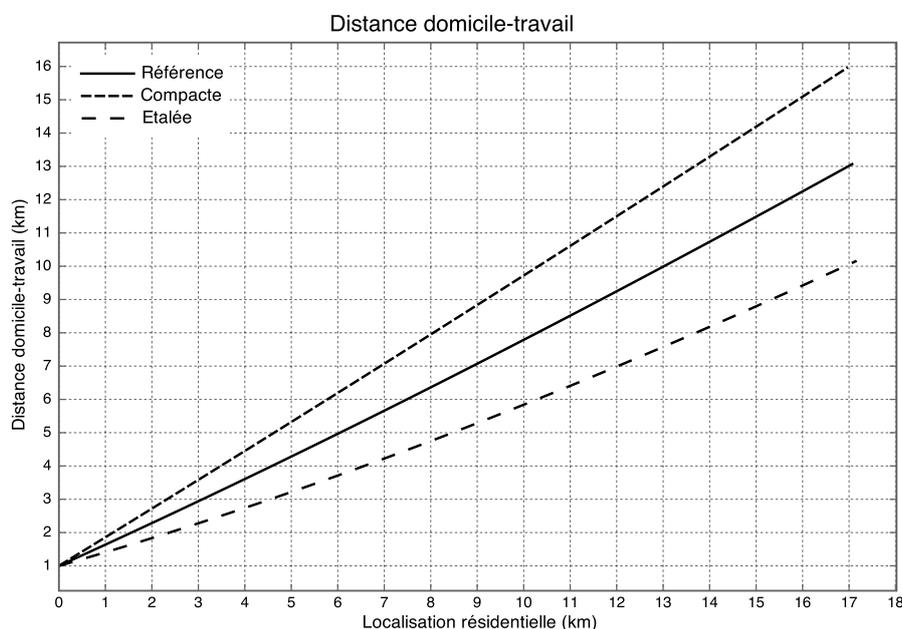


FIGURE 8.10.: Distance domicile-travail en situation de référence et pour les deux situations équivalentes.

La Figure 8.12 présente le coût unitaire moyen de transport, c'est-à-dire le rapport entre le coût de transport (plus précisément sa partie variable) et la distance parcourue par le ménage. On constate que ce coût unitaire moyen est élevé pour les ménages centraux, puis il diminue pour augmenter légèrement ensuite. Nous avons

14. Dans ces figures et les suivantes, nous avons fixé l'origine des distances en $r_0 = 1$ km. Autrement dit, les localisations résidentielles sont à comprendre comme des distances au centre de la zone résidentielle, et non au centre de l'agglomération.

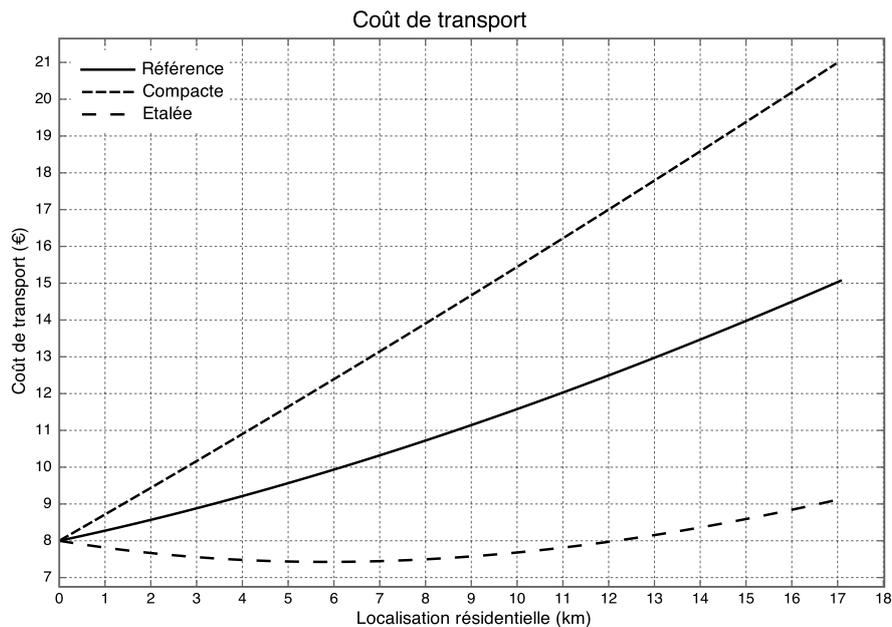


FIGURE 8.11.: Coût de transport en situation de référence et pour les deux situations équivalentes.

trouvé, dans le chapitre 4, qu'en Île-de-France le coût temporel unitaire moyen était élevé pour les ménages résidant en bordure immédiate de Paris et que ce coût diminuait ensuite. Ce résultat trouve, avec notre modèle, une explication naturelle. Par ailleurs, la décroissance est beaucoup plus marquée pour la configuration à zone d'emplois étalée qu'en configuration de référence ou pour une configuration compacte.

La modification de la courbe des distances domicile-travail avec l'étalement des emplois s'accompagne d'une transformation du profil d'usage des transports, comme nous l'avons mentionné au 8.5.1.2. Ainsi, la Figure 8.13 présente les profils d'intensité d'usage du réseau de transport, pour la situation de référence et les deux situations contrastées équivalentes, en présence de congestion dans le centre.

On y constate l'écrasement de la pointe d'intensité lorsque l'emploi se disperse. Celle-ci a une double origine : d'une part, la baisse des distances domicile-travail des ménages a une influence directe sur la quantité de véhicule-kilomètres parcourus ; d'autre part, l'extension de la zone d'emplois réduit la concentration de ces véhicule-kilomètres. La décentralisation de l'emploi induit une décentralisation des déplacements, et donc une diminution potentielle de la congestion, en particulier près du centre. La pointe d'intensité se déplace vers la périphérie.

Bien que nous ayons choisi de centrer nos investigations sur la question des dis-

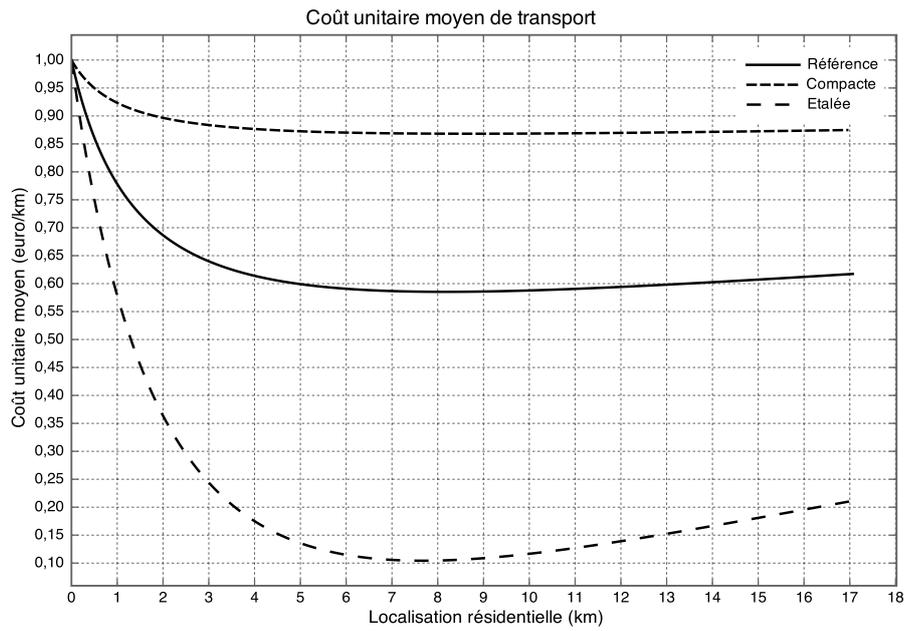


FIGURE 8.12.: Coût de transport unitaire moyen en situation de référence et pour les deux situations équivalentes.

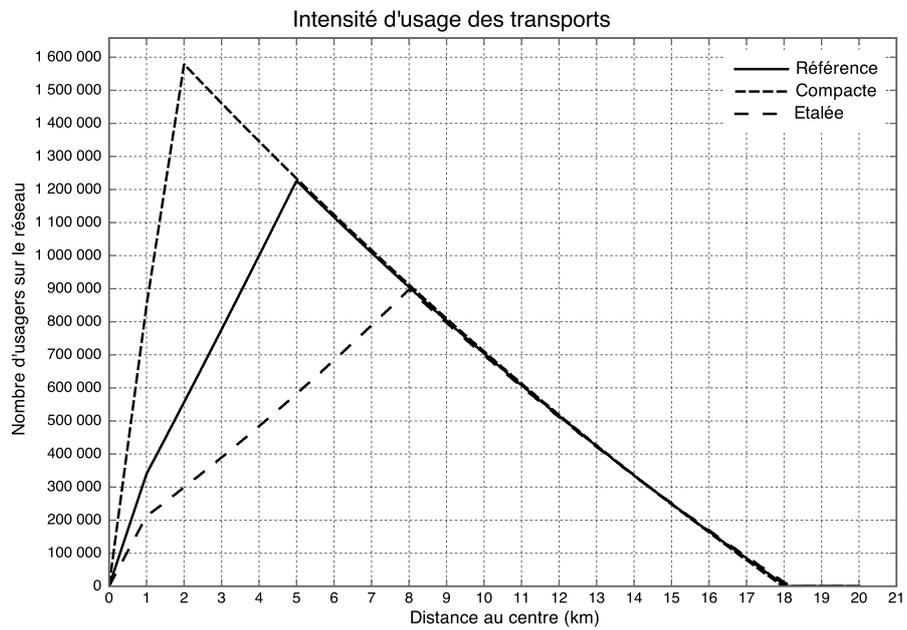


FIGURE 8.13.: Profil d'usage des transports en situation de référence et pour les deux situations équivalentes.

tances parcourues et de l'usage des transports, l'analyse de la rente foncière et de la consommation d'espace par les ménages revêt un intérêt, car ces dernières permettent d'éclairer les constats établis précédemment. Ainsi, une taille plus faible des logements permet un resserrement des ménages, et donc des distances à parcourir réduites. Les Figures 8.14 et 8.15 présentent, toujours dans le cas avec congestion centrale, les courbes de rente foncière et de surface des logements.

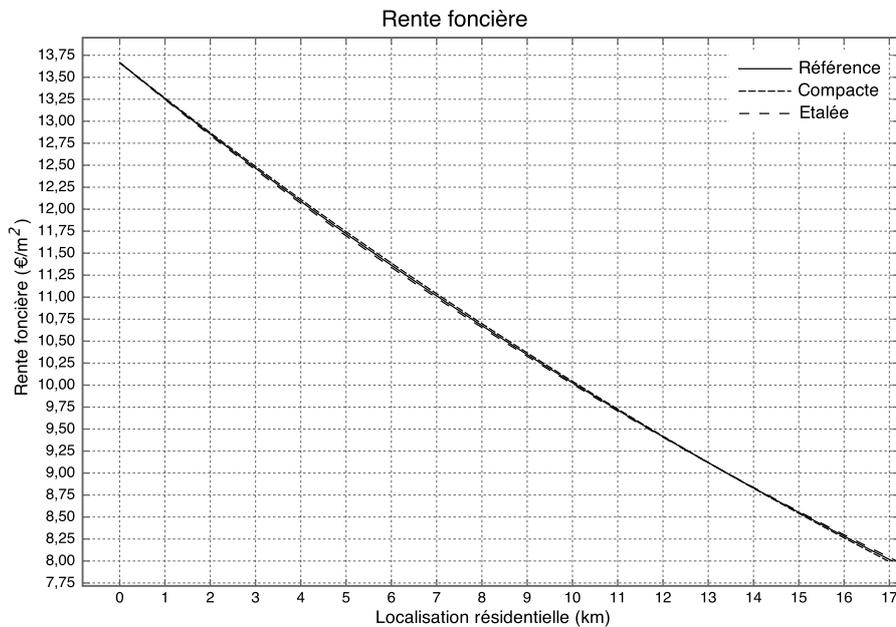


FIGURE 8.14.: Rente foncière en situation de référence et pour les deux situations équivalentes.

On constate d'une part que les localisations intermédiaires perdent de la valeur, au profit des localisations périphériques, lorsque l'emploi d'étale, et d'autre part que la pente de la courbe des surfaces de logement s'accroît avec l'étalement de l'emploi. Autrement dit, les localisations centrales sont toujours aussi attractives pour les ménages qui travaillent au centre, mais ces derniers sont contraints de limiter leur consommation d'espace, du fait de la baisse de revenu qu'ils subissent. Au contraire les travailleurs ayant un emploi plus distant du centre bénéficient à plein de l'étalement des emplois : ils peuvent bénéficier de logements plus grands en périphérie, rendant la zone intermédiaire moins attirante, et donc moins onéreuse¹⁵. L'étalement des emplois engendre donc une forme de décongestion du marché im-

15. L'étalement urbain qui résulte de l'étalement des emplois dans notre modèle a ceci de particulier qu'il ne se traduit pas par une homogénéisation des surfaces des logements, mais par une augmentation de la surface moyenne des logements, induisant un accroissement de l'emprise urbaine sans accroissement de la population. La nature de cet étalement diffère par conséquent de celle présentée à la Figure 8.1.

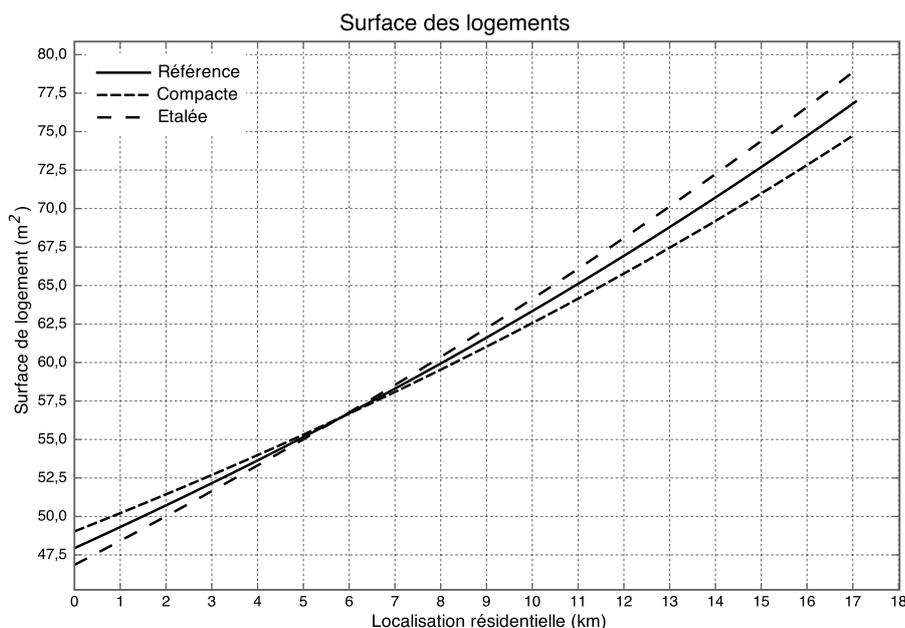


FIGURE 8.15.: Surface des logements en situation de référence et pour les deux situations équivalentes.

mobilier de la zone intermédiaire : la pression foncière diminue.

8.5.2.3. Équité spatiale et étalement

Le paragraphe précédent le laisse pressentir : la localisation des emplois a une incidence sur la distribution des niveaux d'utilité atteints par les ménages. En effet, dans notre modèle, en plus du revenu brut Y , la localisation de l'emploi constitue un facteur d'utilité, par le biais d'une dotation en nature. De manière à mesurer, en termes monétaires, les différents niveaux d'utilité, nous emploierons la notion de revenu équivalent, appliquée ici non pas à une transformation dans le temps, mais à une transformation dans l'espace.

La notion de revenu brut équivalent En effet, le revenu net des ménages ne convient pas pour monétiser l'utilité des ménages, car il n'est qu'une indication indirecte, dans laquelle la localisation résidentielle joue un rôle, par le biais du coût de transport. Il est donc préférable d'évaluer le revenu brut équivalent qu'il faudrait allouer à un ménage de référence pour lui faire atteindre le niveau d'utilité \tilde{u}_ρ d'un autre ménage. Le ménage de référence peut être fixé de manière arbitraire ; nous pouvons par exemple choisir le ménage initial pour lequel $\rho = 0$, ou le ménage « médian » employé en $\rho = \rho_f/2$.

Le revenu équivalent s'écrit donc :

$$\tilde{Y}_{\rho_{ref}}(\rho) = T_{\omega}(\rho_{ref}) + E(R_{ref}, \tilde{u}_{\rho}) \quad (8.25)$$

en notant $E(R, u)$ la fonction de dépense (*expenditure function*) qui fournit le revenu net permettant au ménage d'atteindre le niveau d'utilité u lorsque le prix du sol est de R . Le revenu équivalent en ρ , pour un ménage de référence en ρ_{ref} , est donc la somme du coût de transport de ce ménage de référence et du revenu net lui permettant, compte tenu de la rente foncière en ρ_{ref} , d'atteindre le niveau d'utilité \tilde{u}_{ρ} que les ménages en ρ atteignent.

Pour la fonction d'utilité de type Cobb-Douglas retenue ici, la fonction de dépense est :

$$E(R, u) = R^{\beta'} \left(\frac{u}{U_0} \right)^{1/(\alpha+\beta)} \quad (8.26)$$

Si bien que relativement au ménage initial employé en 0 et établi en r_0 , le revenu équivalent du ménage employé en ρ et établi en $r_{\omega}(\rho)$ est :

$$\tilde{Y}_0(\rho) = Y + \frac{a'}{\alpha'} \tilde{\rho} \left[1 - \left(\frac{R_{\omega}(\rho)}{R_0} \right)^{\alpha'} \right] \quad (8.27)$$

De la même manière, relativement au ménage médian employé en $\rho_f/2$, le revenu équivalent du ménage employé en ρ est :

$$\tilde{Y}_{\rho_f/2} = Y + \frac{a'}{\alpha'} \tilde{\rho} \left(1 - \frac{\rho_f}{2\tilde{\rho}} \right) \left[1 - \left(\frac{R_{\omega}(\rho)}{R_{1/2}} \right)^{\alpha'} \right] \quad (8.28)$$

où $R_{1/2}$ est la rente foncière payée par le ménage médian. Les calculs figurent en Annexe 8.A.2.

On remarque que conformément au constat fait au chapitre 6 de la croissance de l'utilité des ménages avec la distance au centre, on a $\tilde{Y}_0(\rho) \geq Y$, et $\tilde{Y}_{\rho_f/2}(\rho) \leq Y$ selon que $\rho \leq \rho_f/2$.

Plus précisément, la Figure 8.16 présente les courbes de revenu brut équivalent, en situation de référence, et en situations compacte et étalée. Le ménage de référence retenu est, dans les trois cas, le ménage résidant en r_0 .

La figure amène plusieurs commentaires. Tout d'abord, on vérifie que, comme nous l'avons montré dans le chapitre 6, l'utilité des ménages à l'équilibre — représentée ici sous forme monétaire — augmente avec la distance au centre. De plus, la pente de la courbe augmente avec l'étalement des emplois.

Par ailleurs, on note que pour ces transformations à surplus agrégé nul, certains ménages — ceux qui habitent en périphérie — bénéficient, dans une situation où

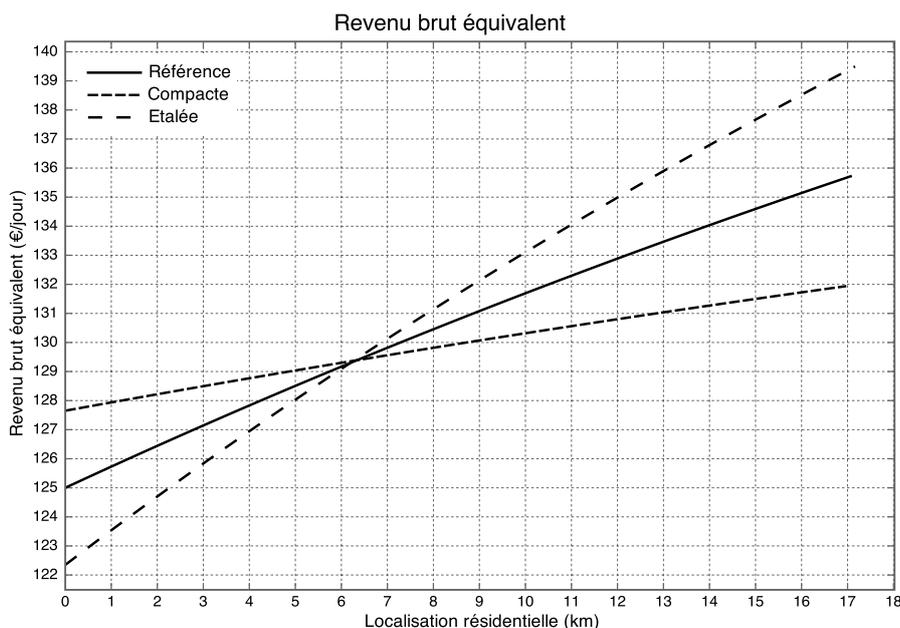


FIGURE 8.16.: Revenu brut équivalent, en situation de référence et pour les deux situations équivalentes.

l'emploi est étalé, d'un niveau d'utilité supérieur à celui qu'ils atteignent dans une configuration compacte; l'inverse se produit pour les ménages résidant au centre. Autrement dit, comme nous l'avions évoqué plus tôt, si la transformation se fait à surplus nul, cela ne garantit en rien l'équité de celle-ci. Ce résultat suggère que l'avantage procuré par la localisation de l'emploi est très inégalement réparti, et ce d'autant plus que l'emploi s'étale.

On peut essayer de tirer de ce résultat quelques idées portant sur un éventuel scénario d'évolution de la structure urbaine. En effet, les travailleurs du centre semblent être ceux qui ont le plus à perdre à un étalement des emplois, tandis que les ménages périphériques peuvent y trouver leur compte, au moins à moyen terme. Par conséquent, on peut imaginer que travailleurs du centre et de la périphérie ne favoriseront pas les mêmes politiques de développement de l'emploi. Il est probable que les travailleurs centraux valoriseront la densité de l'emploi, voire sa densification, tandis que les travailleurs de la périphérie verront vraisemblablement d'un bon œil un étalement des emplois.

S'il est clair que de nombreuses causes, liées aux préférences intrinsèques des individus, peuvent expliquer que les habitants du centre soient ceux qui valorisent la densité, tandis que les habitants de la périphérie privilégient l'espace — un phénomène d'autosélection est clairement en jeu — l'analyse que nous avons menée à partir de notre modèle fournit toutefois une explication alternative.

On peut par ailleurs noter, faisant ici écho à la remarque de Fujita (1989, p. 128), que cet état de fait est potentiellement à la source d'un conflit entre travailleurs du centre et de la périphérie, sur la question de l'évolution urbaine souhaitable. Or, les décideurs politiques et leurs conseillers font très souvent partie de ces travailleurs centraux, qui ont tout intérêt à une densification des villes. Autrement dit, et même s'il n'est pas question de faire preuve ici de cynisme, il ne serait pas étonnant que la majorité d'entre eux soit partisan de la compacité...

8.5.2.4. Décentralisation des emplois et économies d'agglomération

Nous avons comparé, dans les paragraphes précédents, des situations pour lesquelles les ménages sont indifférents car *leur salaire a été ajusté en conséquence*. Or, les mécanismes liés aux économies d'agglomération peuvent expliquer cette différence de salaire. Nous l'avons vu au chapitre 1, les économies d'agglomération, à l'origine des villes, peuvent émerger de mécanismes faisant intervenir l'amélioration de la productivité des entreprises lorsqu'elles se regroupent, par l'intermédiaire des rendements d'échelle croissants (Fujita et Thisse, 1996). On peut, plus précisément, distinguer deux types de mécanismes : le premier fait intervenir la taille du marché, ou le nombre d'entreprises de la zone d'emploi ; le second rend compte du rôle joué par la concentration spatiale de ces entreprises, autrement dit leur densité. Plus précisément, il est légitime de penser que si les firmes se dispersent dans l'espace, elles perdent une partie des bénéfices liés à leur concentration.

Nous proposons ici, en nous appuyant sur un modèle de diffusion de l'information au sein d'une ville, une justification à l'existence d'une différence de salaire entre des configurations urbaines différentes. Plus précisément, nous adaptons ici le modèle développé par Ogawa et Fujita (1980) et Imai (1982), basé sur le rôle des retombées informationnelles, avec une portée géographique linéairement décroissante en distance¹⁶. Autrement dit, on suppose que le profit, dû aux retombées informationnelles, d'une firme située en ρ est :

$$\pi(\rho) = \int_0^{\rho_f} p(\rho, \rho') f(\rho') d\rho' \quad (8.29)$$

où $p(\rho, \rho') = \pi_0 - \ell|\rho - \rho'|$ est le profit unitaire que retire la firme en ρ de sa proximité avec une firme en ρ' , et $f(\rho') = b = N/\rho_f$ ¹⁷. ℓ peut être interprété comme

16. Ces auteurs ont montré que dans ce cadre linéaire, les configurations urbaines possibles sont monocentrique, partiellement mixte, ou complètement mixte. Avec des bénéfices à décroissance exponentielle en distance, les configurations urbaines possibles sont les précédentes, auxquelles s'ajoutent des configurations à plusieurs centres.

17. On notera qu'on considère ici que chaque firme n'emploie qu'un seul travailleur ; en effet, on peut sans perte de généralité faire cette hypothèse qui revient simplement à modifier les valeurs de π_0 et ℓ .

une mesure de l'influence de la distance sur l'importance des retombées informationnelles. En particulier, on peut supposer que le développement des technologies de l'information et de la communication réduisent la valeur de ℓ .

On obtient ensuite une expression analytique du profit d'une firme en ρ :

$$\pi(\rho) = \frac{N}{\rho_f} \left(\pi_0 \rho_f - \frac{1}{2} \ell \rho_f^2 + \ell \rho_f \rho - \ell \rho^2 \right) \quad (8.30)$$

Enfin, la somme des profits générés par l'ensemble des firmes grâce à ces échanges informels d'information s'écrit :

$$\begin{aligned} \Pi &= \int_0^{\rho_f} \pi(\rho) f(\rho) d\rho \\ &= N^2 \left(\pi_0 - \frac{1}{3} \ell \rho_f \right) \end{aligned} \quad (8.31)$$

On remarque que le profit global croît de manière quadratique avec le nombre de firmes (et de travailleurs).

De manière à rester dans le cadre de notre modèle, où tous les ménages touchent le même salaire, on suppose que ce profit global est intégralement redistribué aux ménages, et qu'il constitue leur revenu. Autrement dit :

$$Y = \frac{\Pi}{N} = N\pi_0 - \frac{N}{3} \ell \rho_f \quad (8.32)$$

Dans cette expression, $N\pi_0$, qu'on notera Y_M , représente le salaire que recevraient les ménages dans une ville purement monocentrique, avec l'emploi concentré en un point¹⁸. Le terme $\frac{N}{3} \ell \rho_f$ constitue la perte de revenu liée à la déconcentration de l'emploi dans un rayon ρ_f . On la notera $D(\rho_f)$. De manière symétrique, si ρ_f est supposé fixé, on notera $\tilde{D}(b) = \frac{N^2 \ell}{3b}$, fonction décroissante en b .

On pourra remarquer aussi que le revenu des ménages devant rester positif, et même supérieur à a_0 , le coût fixe de transport, on doit avoir :

$$\rho_f < \frac{3N}{\ell} (\pi_0 - a_0) \quad (8.33)$$

Autrement dit, le rayon maximal de dispersion des firmes est d'autant plus grand que des technologies de communication efficaces à distance sont disponibles.

Le modèle précédent montre que selon l'intensité de la force d'agglomération — qui dépend de la population de la ville et des technologies de communication

18. Notons que ce revenu est croissant en N , fournissant donc une force d'agglomération au niveau global de la ville. Autrement dit, plus la ville comprend d'emplois, plus les salaires versés sont élevés. Toutefois, la pression foncière et la congestion augmentant dans le même temps, les coûts du logement et du transport limitent l'afflux de population.

disponibles — le différentiel de salaire sera supérieur ou inférieur à la variation compensatrice de revenu globale mise en évidence précédemment, si bien que l'une ou l'autre des configurations urbaines sera vraisemblablement plus avantageuse pour les ménages, pris dans leur ensemble.

Autrement dit, la configuration compacte peut se révéler meilleure que la configuration étalée pour certaines valeurs de la population et certaines technologies de diffusion de l'information, mais l'inverse peut se produire pour d'autres valeurs. Les phénomènes d'inertie et d'hystérésis rendent néanmoins le passage brutal d'une configuration à l'autre inenvisageable dans la réalité, mais une évolution plus lente est possible. De plus, nous avons montré que l'équité du passage d'une configuration à l'autre pouvait poser problème.

8.6. Conclusion

La question de l'influence des localisations respectives des emplois et des résidences sur les déplacements est complexe, tant par le caractère fortement endogène des choix de localisation de ces deux types d'acteurs que sont les ménages et les entreprises, que du fait des nombreux effets externes qu'elle implique (congestion, pression foncière, etc.). Dans ce contexte scientifique, le cadre de modélisation que nous avons développé et mis en œuvre ici permet de rendre compte, de manière théorique, de certains phénomènes mis en évidence de manière empirique. Ainsi, nous avons constaté la dissociation partielle entre la distance parcourue et le temps de transport pour certaines valeurs des paramètres urbains. Ce phénomène est de nature à fournir un moteur à l'étalement urbain.

Par ailleurs, nous avons mis en évidence une diminution des niveaux de congestion dans le centre des agglomérations directement liée à l'étalement des emplois, ainsi qu'une baisse des distances parcourues par les ménages. Dans le même temps, l'extension urbaine d'une ville dans laquelle l'emploi est étalé est supérieure à celle d'une ville plus compacte. La consommation d'espace « naturel » y est donc plus importante, ce qui peut être porté au crédit de la compacité. Enfin, nous avons exploré succinctement la question de la variation des économies d'agglomération selon la forme urbaine et montré que la détermination de la meilleure configuration de la distribution des emplois dépend de paramètres tels que la population de la ville et l'intensité spatiale des forces d'agglomération en jeu.

Finalement, dans le cadre restreint d'un modèle monocentrique à distribution exogène des emplois, nous avons tenté de nuancer le discours, fréquent, tendant à donner la compacité comme la forme urbaine à atteindre par excellence. Une forme d'étalement des emplois peut améliorer le bien-être des ménages, tout en diminuant les distances qu'ils parcourent pour se rendre sur leur lieu de travail.

Annexe 8.A Annexes

8.A.1 Ménage obtenant le surplus maximal

À partir de l'équation (8.20), on calcule la dérivée de $\sigma(n)$:

$$\frac{d\sigma}{dn} = \frac{\rho_{f2} - \rho_{f1}}{N} \frac{a'}{\alpha'} \left[1 - \beta' \left(\frac{\lambda R_0}{\lambda R_0 - an} \right)^{\alpha'} \right]$$

Si bien que la condition $\frac{d\sigma}{dn} \geq 0$ est obtenue si et seulement si :

$$n \leq \frac{\lambda R_0}{a} (1 - \beta'^{1/\alpha'})$$

Pour que la valeur n^* soit atteinte dans la ville, il faut que $n^* < N$, ce qui signifie :

$$\left(\frac{\lambda R_A}{a} + N \right) (1 - \beta'^{1/\alpha'}) < N$$

Et donc :

$$N > \frac{\lambda R_A}{a} \left(\frac{1}{\beta'^{1/\alpha'}} - 1 \right)$$

8.A.2 Calcul du revenu brut équivalent

Le revenu équivalent, en ρ , pour le ménage de référence en 0 s'écrit, compte tenu de l'expression de $\tilde{u}_\rho = V(R_\omega(\rho), I_\omega(\rho))$, et de l'expression de $I_\omega(\rho)$ obtenue au chapitre précédent :

$$\begin{aligned} \tilde{Y}_0(\rho) &= T_\omega(0) + E(R_0, \tilde{u}_\rho) \\ &= a_0 + ar_0 + R_0^{\beta'} R_\omega(\rho)^{-\beta'} I_\omega(\rho) \\ &= a_0 + ar_0 + \left[Y_0 + \frac{a'}{\alpha'} \tilde{\rho} - ar_0 - \frac{a'}{\alpha'} \tilde{\rho} \left(\frac{R_\omega(\rho)}{R_0} \right)^{\alpha'} \right] \\ &= Y + \frac{a'}{\alpha'} \tilde{\rho} \left[1 - \left(\frac{R_\omega(\rho)}{R_0} \right)^{\alpha'} \right] \end{aligned}$$

De la même manière, on calcule le revenu brut équivalent, en ρ , pour le ménage de référence en $\rho_f/2$. On utilise pour cela l'expression de $r_\omega(\rho)$ obtenue au chapitre

précédent. On obtient donc, en notant $R_{1/2} = R_\omega(\rho_f/2)$:

$$\begin{aligned}
 \tilde{Y}_{\rho_f/2}(\rho) &= T_\omega(\rho_f/2) + E(R_{1/2}, \tilde{u}_\rho) \\
 &= a_0 + ar_\omega(\rho_f/2) - \frac{a'\rho_f}{2} + \left(\frac{R_{1/2}}{R_\omega(\rho)}\right)^{\beta'} I_\omega(\rho) \\
 &= a_0 + a \left(A + B \frac{R_{1/2}}{R_0} - C \left(\frac{R_{1/2}}{R_0}\right)^{\beta'} \right) - \frac{a'\rho_f}{2} \dots \\
 &\quad \dots + aC \left(\frac{R_{1/2}}{R_0}\right)^{\beta'} - \frac{a'}{\alpha' \tilde{\rho}} \frac{R_{1/2}}{R_0} \left(\frac{R_{1/2}}{R_\omega(\rho)}\right)^{-\alpha'} \\
 &= Y + a' \left(\tilde{\rho} - \frac{\rho_f}{2} \right) + \frac{a'}{\alpha' \tilde{\rho}} \frac{R_{1/2}}{R_0} \left[\beta' - \left(\frac{R_\omega(\rho)}{R_{1/2}}\right)^{\alpha'} \right]
 \end{aligned}$$

Or, $R_{1/2}/R_0 = 1 - \frac{\rho_f}{2\tilde{\rho}}$, si bien que finalement :

$$\tilde{Y}_{\rho_f/2}(\rho) = Y + \frac{a'}{\alpha' \tilde{\rho}} \left(1 - \frac{\rho_f}{2\tilde{\rho}} \right) \left[1 - \left(\frac{R_\omega(\rho)}{R_{1/2}}\right)^{\alpha'} \right]$$

Conclusion générale

Cette thèse avait un double objectif : mieux comprendre les liens entre congestion des systèmes de transport et localisation des ménages et des emplois, d'une part, et d'autre part, apporter une contribution théorique à l'analyse des conséquences de la localisation des emplois sur les choix de localisation des ménages et leur usage des transports. Le caractère éminemment spatial des questions que nous avons abordées tout au long de cette thèse (congestion des transports, étalement urbain) nous a conduit à prendre en compte l'espace comme dimension principale d'analyse. Cette prise en compte s'est traduite notamment par une analyse des cheminements effectués par les usagers sur le réseau routier, ou sous la forme d'une modélisation de la décentralisation des emplois. Nous avons ainsi montré qu'une telle prise en compte permet d'affiner le diagnostic réalisé, que celui-ci porte sur la congestion des transports ou sur les questions de rivalité pour l'usage du sol.

Structure urbaine et congestion sont intimement liées

Nous avons, dans la première partie de la thèse, rappelé comment structure urbaine et processus d'agglomération sont liés. Comme nous l'avons souligné, la structure émerge de la rencontre, pour les différents agents, entre forces d'agglomération et forces de dispersion (chapitre 1). Ces forces, et les mécanismes qui les soutiennent, sont variées et se renforcent généralement l'une l'autre. Parmi les forces de dispersion, le coût de transport, et en particulier la congestion, joue un rôle important. Or, il s'agit d'une force de dispersion qui résulte, de manière indirecte, de la structure urbaine puisqu'elle dépend de la manière dont les flux se concentrent sur certains itinéraires et autour de certains pôles. Autrement dit, si la structure est le résultat de la confrontation de forces antagonistes, l'intensité de ces forces dépend elle-même de la structure urbaine. Nous avons voulu montrer à quel point les mécanismes par lesquels la congestion se manifeste sont nombreux et enchevêtrés (chapitre 2). En mettant en évidence cette complexité, nous avons également pu en tirer des enseignements sur la manière de mesurer la congestion, et de faire des choix de modélisation adaptés en restant conscient des limites de cette démarche de modélisation.

L'approche retenue dans la deuxième partie de la thèse visait à mettre en relation la congestion avec les choix d'itinéraires des usagers en Île-de-France. La mise en œuvre d'une méthodologie s'appuyant sur une modélisation des flux et la définition d'un certain nombre d'indicateurs complémentaires (chapitre 3) nous a permis, d'une part, de mettre en évidence que la variabilité, tant spatiale que temporelle de la congestion routière était très élevée, et d'autre part que la prise en compte de la logique sous-jacente des déplacements (les origines et destinations des déplacements) permet de faire émerger des structures spatiales claires (chapitre 4). Autrement dit, nous avons montré que la compréhension du phénomène de congestion dans sa dimension spatiale nécessite de faire appel aux caractéristiques des déplacements : l'étude des flux sur les arcs ne suffit pas. Un des enjeux de cette thèse résidait dans la compréhension des interactions à double sens entre congestion et localisation des emplois et des ménages. Nos analyses ont souligné à quel point l'étude de la congestion faisait d'elle-même émerger la structure sous-jacente des localisations. De plus, en ce qui concerne l'équité spatiale, qui constituait un autre enjeu de notre travail, nous avons constaté que les usagers subissant les plus hauts niveaux de congestion étaient également ceux qui en produisent le plus. Par ailleurs, nous avons pu exploiter certains aspects de notre analyse de la variabilité spatiale dans une méthode d'évaluation opérationnelle de projets de transport. Nous avons ainsi montré qu'une segmentation du trafic liant segmentation de l'offre et segmentation de la demande permet d'améliorer sensiblement l'évaluation des gains de temps de décongestion routière par rapport aux méthodes traditionnellement utilisées en Île-de-France notamment (chapitre 5).

L'étalement des emplois modifie en profondeur l'usage des transports

Dans la troisième partie de la thèse, nous avons retenu une approche qui visait à rendre compte, de manière théorique, du rôle joué par la localisation des emplois et des ménages sur l'usage des transports dans un cadre monocentrique. Nous avons donc proposé de prendre en compte les origines et destinations des déplacements domicile-travail dans un modèle théorique d'équilibre urbain de moyen terme, de type monocentrique, incluant une distribution exogène des emplois (chapitre 6). Le caractère exogène de la distribution des emplois nous a permis de disposer d'un modèle soluble de manière analytique, nous autorisant à modifier à volonté cette distribution et ainsi à en observer les conséquences sur les distances domicile-travail. La résolution de ce modèle donne lieu à un équilibre où tous les ménages n'atteignent pas le même niveau d'utilité, démontrant le rôle de l'espace dans la différenciation des ménages. L'analyse en statique comparative dans un cas particulier de l'influence des différents paramètres sur l'équilibre a montré, notamment, qu'une amélioration de la qualité de service de transport dans le centre d'une aggloméra-

tion peut limiter l'étalement urbain (chapitre 7). Nous avons également pu mettre en évidence, dans le cas d'une ville que nous avons appelée « déconcentrée », une dissociation entre la distance domicile-travail et le coût de transport, fournissant un fondement théorique à de nombreuses observations empiriques : alors que les distances parcourues augmentent avec la distance au centre du logement, le coût subi par les ménages peut rester stable voire diminuer. Autrement dit, la congestion qui caractérise généralement le centre des grandes agglomérations, associée à une relative dispersion de l'emploi, favorise potentiellement l'étalement urbain. Nous avons par ailleurs montré que l'étalement urbain qui résulte d'un étalement maîtrisé des emplois peut conduire à une diminution des distances domicile-travail (chapitre 8). Toutefois, à moyen terme, un tel étalement est susceptible d'engendrer des problèmes d'équité spatiale, certains ménages y trouvant bénéfice tandis que d'autres y perdent.

Perspectives de recherche et pistes de réflexion

Au-delà des résultats proprement dits que nous venons de rappeler, cette thèse apporte, selon nous, des éléments de réflexion sur les sujets que nous avons abordés, et ouvre, dans son prolongement, des perspectives de recherche.

Ainsi, il s'avère que l'action publique dans le domaine de la maîtrise de la congestion se base généralement assez peu sur un diagnostic technique de la congestion, mais plutôt sur des analyses portant sur les questions d'accessibilité, et sur des visions politiques dans lesquelles l'éclairage technique n'a qu'un poids très limité. Il y a vraisemblablement plusieurs raisons à cela. Tout d'abord, les outils généralement employés pour mesurer la congestion ne correspondent pas véritablement au ressenti des usagers. En particulier, les questions de fiabilité ne sont pas véritablement prises en compte. Il nous semble également que les indicateurs retenus ne rendent pas compte de la structure spatiale de la congestion : ils restent généralement au niveau du tronçon routier, sans se soucier de l'intégralité du parcours de l'utilisateur, au cours duquel il peut faire face à d'autres points de ralentissement. S'il existe bien des indicateurs qui fournissent ce type d'information, par exemple le temps de parcours moyen supplémentaire à l'heure de pointe, ce sont des valeurs moyennes à l'échelle d'une agglomération, et ils ne permettent donc pas d'entrer dans le détail spatial de l'agglomération. Au contraire, la méthodologie que nous avons proposée s'appuie sur des outils communément employés par les ingénieurs chargés d'études sur la congestion, mais les exploite de manière suffisamment différente pour produire des résultats qui nous semblent plus parlants, et susceptibles d'être employés pour fournir une aide à la décision en matière de maîtrise de la congestion.

Toutefois, l'utilisation d'un modèle dynamique de trafic permettrait de compléter la méthode en y ajoutant la possibilité d'introduire des variations aléatoires de la

capacité (survenue d'un accident par exemple) ou de la demande. Ceci constituerait selon nous une voie à explorer afin d'améliorer encore le diagnostic économique de la congestion que nous avons proposé. Par ailleurs, afin d'affiner cette méthode et de confirmer sa capacité à évaluer de manière rapide et satisfaisante les gains de décongestion, il serait utile de mettre en œuvre notre méthodologie d'évaluation des gains de décongestion sur d'autres projets, sur d'autres territoires, franciliens notamment.

Concernant l'étalement urbain, les discours sur cette question sont bien souvent empreints d'une forte idéologie : on est pour ou contre la ville compacte, pour ou contre l'étalement. De nombreux arguments scientifiques ont été avancés pour appuyer l'une ou l'autre des positions, à partir de modèles théoriques ou de données empiriques. Toutefois, malgré les riches enseignements qu'ils constituent, les résultats obtenus par ces recherches ou ces études sont souvent délicats à généraliser, du fait de la complexité et de la diversité des objets et des situations analysés. Il est donc difficile de trancher entre ces deux modes de développement. Nous avons également souhaité alimenter ce débat sur des bases scientifiques en apportant des éléments d'ordre théorique, à travers l'étude du rôle que peut jouer l'étalement des emplois dans l'étalement urbain. En somme, nous avons dégagé deux résultats principaux : l'existence de hauts niveaux de congestion au centre des agglomérations est de nature à créer les conditions favorables à un éloignement des ménages des centres-villes ; l'étalement urbain résultant de l'étalement des emplois est plutôt économe en distances parcourues. Ces deux résultats suggèrent d'une part qu'une grande vigilance est de mise quant à des politiques de transport conduisant à une diminution de la capacité, notamment routière, sans réelle compensation par les transports collectifs, et d'autre part que chercher à restreindre la décentralisation des emplois pour limiter l'étalement urbain peut aboutir à une augmentation des distances parcourues pour les déplacements domicile-travail.

De plus, notre travail sur le modèle d'équilibre urbain ouvre selon nous un certain nombre de pistes de recherche à court terme. En tout premier lieu, la technologie de transport que nous avons retenue (réseau de transport dense et isotrope) correspond à la configuration des réseaux routiers urbains, si bien que notre modèle rend principalement compte de l'usage de la voiture particulière. Or, une prise en compte d'un autre mode de transport, collectif, est possible dans le cadre monocentrique (Kilani *et al.*, 2010). Étudier, dans un tel cadre, comment la décentralisation des emplois influence le partage modal et donc la consommation d'énergie dans les transports de manière plus fine serait intéressant. De même, concernant toujours la technologie de transport, le modèle, dans son état actuel, ne prend pas en compte la congestion des transports de manière endogène. Par l'intermédiaire d'une résolution numérique du modèle, il est tout à fait possible d'introduire une telle endogénéité,

ce qui permettrait d'analyser de manière plus fine les liens entre étalement et niveau de congestion, et d'interroger un peu plus encore la capacité de l'étalement des emplois à produire de l'étalement urbain, et la forme que peut prendre cet étalement. Par ailleurs, nous n'avons représenté, dans notre modèle, que le marché immobilier résidentiel, sans s'intéresser directement au marché immobilier d'entreprise, en nous fondant sur l'hypothèse qu'ils sont, au moins en partie, distincts, pour des raisons de rigidité du type d'usage des sols ou des locaux. Toutefois, ces deux marchés ne sont, en réalité, pas totalement hermétiques, et un certain recouvrement entre les deux a lieu. Autrement dit, une concurrence est susceptible de s'établir entre les usages résidentiels et professionnels. Une extension de notre modèle à l'aide d'une représentation à la manière de Wheaton (2004) de la concurrence entre marchés immobiliers résidentiel et d'entreprises serait intéressante, car elle renseignerait sur les conditions de l'étalement des emplois, et de la mixité des usages du sol.

Enfin, conçu comme une extension du modèle monocentrique standard, notre modèle ne prend en compte que les déplacements domicile-travail. Si nous avons rappelé, à partir d'une analyse bibliographique, que ces déplacements restaient très importants tant en termes de volume que parce qu'ils structurent une grande partie des autres déplacements des ménages, il n'en demeure pas moins que d'autres types de déplacements tendent à prendre de l'importance (achats, loisirs). Or, ces déplacements, dont les logiques peuvent être très différentes de celles qui sous-tendent les déplacements domicile-travail, peuvent influencer le choix résidentiel, ou à l'inverse avoir un fort impact sur la mobilité de ménages dont le choix résidentiel a été établi sur la base de leurs déplacements domicile-travail. Dans tous les cas, la prise en compte de ces autres déplacements dans le cadre d'un modèle théorique aussi simple que le modèle monocentrique (ou l'extension que nous en avons proposée) reste une piste à explorer, afin de couvrir un champ plus vaste de la mobilité des ménages.

Bibliographie

- ABDEL-RAHMAN, H. et FUJITA, M. (1990). Product variety, marshallian externalities, and city size. *Journal of Regional Science*, 30:165–183.
- AGUILÉRA, A. et MIGNOT, D. (2002). Structure des localisations intra-urbaines et mobilité domicile-travail. *Recherche Transports Sécurité*, 77:311–325.
- AGUILÉRA, A. et MIGNOT, D. (2007). Formes urbaines et migrations alternantes – Les enseignements d’une comparaison des aires urbaines de Lille, Lyon et Marseille. In *43ème colloque de l’ASRDLF*, Grenoble-Chambéry. ASRDLF.
- AGUILÉRA, A., WENGLANSKI, S. et PROULHAC, L. (2009). Employment suburbanisation, reverse commuting and travel behaviour by residents of the central city in the Paris metropolitan area. *Transportation Research Part A*, 43(7):685–691.
- AGUILÉRA, V. et LEURENT, F. (2010). The marginal congestion delay and its external social cost : A dynamic, network-based analysis. In *89th Annual Meeting of the Transportation Research Board CD-Rom*, Washington, DC. TRB.
- AKÇELİK, R. (1991). Travel time functions for transport planning purposes : Davidson’s function, its time-dependent form and an alternative travel time function. *Australian Road Research*, 21:49–59.
- ALONSO, W. (1964). *Location and Land Use; Toward a General Theory of Land Rent*. Harvard University Press, Cambridge, Massachussets.
- ANAS, A., ARNOTT, R. et SMALL, K. A. (1998). Urban spatial structure. *Journal of Economic Literature*, 36(3):1426–1464.
- ANAS, A. et KIM, I. (1996). General equilibrium models of polycentric urban land use with endogenous congestion and job agglomeration. *Journal of Urban Economics*, 40(2):232–256.
- ANDERSON, S. P. et DE PALMA, A. (2004). The economics of pricing parking. *Journal of Urban Economics*, 55(1):1–20.

- ARNOTT, R. (2007). Congestion tolling with agglomeration externalities. *Journal of Urban Economics*, 62(2):187–203.
- ARNOTT, R., DE PALMA, A. et LINDSEY, R. (1993). A structural model of peak-period congestion : A traffic bottleneck with elastic demand. *The American Economic Review*, 83(1):161–179.
- ARNOTT, R. et INCI, E. (2006). An integrated model of downtown parking and traffic congestion. *Journal of Urban Economics*, 60:418–442.
- ARNOTT, R. et ROWSE, J. (1999). Modeling parking. *Journal of Urban Economics*, 45:97–124.
- ARNOTT, R. J. et KRAUS, M. (1995). Self-financing of congestible facilities in a growing economy. Working Paper.
- ARNOTT, R. J. et KRAUS, M. (1998). When are anonymous congestion charges consistent with marginal cost pricing? *Journal of Public Economics*, 67(1):45–64.
- ARNOTT, R. J. et MACKINNON, J. G. (1978). Market and shadow land rents with congestion. *The American Economic Review*, 68(4):588–600.
- ARNOTT, R. J. et STIGLITZ, J. E. (1981). Aggregate land rents and aggregate transport costs. *The Economic Journal*, 91(362):331–347.
- BAIROCH, P. (1985). *De Jéricho à Mexico. Villes et économie dans l'histoire*. Gallimard, Paris.
- BAR-GERA, H., NIE, Y. M., BOYCE, D., HU, Y. et LIU, Y. (2010). Consistent route flows and the condition of proportionality. In *89th Annual Meeting of the Transportation Research Board CD-Rom*, Washington, DC. TRB.
- BATES, J. J., POLAK, J., JONES, P. et COOK, A. (2001). The valuation of reliability for personal travel. *Transportation Research Part E*, 37:191–229.
- BAUMSTARK, L. (2007). La mesure de l'utilité sociale des investissements : l'enjeu du processus de production des valeurs tutélaires. In MAURICE, J. et CROZET, Y., éditeurs : *Le calcul économique dans le processus de choix collectif des investissements de transport*, chapitre 5, pages 165–190. Economica, Paris.
- BECKER, G. S. (1965). A theory of the allocation of time. *The Economic Journal*, 75(299):493–517.
- BECKMANN, M. J., MCGUIRE, C. B. et WINSTEN, C. B. (1956). *Studies in the Economics of Transportation*. Yale University Press.

- BERLIANT, M., REED, R. R. et WANG, P. (2006). Knowledge exchange, matching, and agglomeration. *Journal of Urban Economics*, 60:69–95.
- BERNARD, A. (1998). Comment faire du calcul économique en milieu urbain? Document de travail.
- BLAUDIN DE THÉ, C. (2010). Sprawl and fuel consumption. Mémoire de master, Paris School of Economics.
- BOARNET, M. G. (1994). The monocentric model and employment location. *Journal of Urban Economics*, 36(1):79–97.
- BOITEUX, C. et HURIOT, J.-M. (2000). Services supérieurs et recomposition urbaine. Document de travail n°2000-02.
- BOITEUX, M. et BAUMSTARK, L. (2001). Transport : choix des investissements et coût des nuisances. Rapport technique, Commissariat Général au Plan, La Documentation Française, Paris.
- BOITEUX, M., BROSSIER, C., BUREAU, D., HUART, Y., QUINET, E., MATHEU, M., HALAUNBRENNER, G., LAPEYRE, J. et LAVILLE, P. (1994). Transport : pour un meilleur choix des investissements. Rapport technique, Commissariat Général au Plan, La Documentation Française, Paris.
- BOULAHBAL, M. (2001). Effet polarisant du lieu de travail sur le territoire de la vie quotidienne des actifs. *Recherche Transports Sécurité*, 73:43–63.
- BOUSSAUW, K., NEUTENS, T. et WITLOX, F. (2010). Spatial variations of the minimum home-to-work distance in the North of Belgium. *In 89th Annual Meeting of the Transportation Research Board CD-Rom*, Washington, DC. TRB.
- BOVY, P. H. L. et JANSEN, G. R. M. (1983). Network aggregation effects upon equilibrium outcomes : An empirical investigation. *Transportation Science*, 17(3): 240–262.
- BPR (1964). *Traffic Assignment Manual*. US Department of Commerce, Urban Planning Division, Bureau of Public Roads, Washington, D.C.
- BRANSTON, D. (1976). Link capacity functions : A review. *Transportation Research*, 10(4):223–236.
- BRETEAU, V. (2010). Déplacements domicile-travail dans un modèle monocentrique étendu, avec localisation des ménages relativement aux firmes. *In Les Premières Journées du Pôle Ville*, Marne-la-Vallée, France. PRES Université Paris-Est.

- BRETEAU, V. et LEURENT, F. (2010). Housing and commuting in an extended monocentric city. *In Conference "Public Policies and Industrial Organization in the City"*, Lille. Equippe. Hal-00505490, <http://hal.archives-ouvertes.fr/hal-00505490/fr/>.
- BROWNSTONE, D. et SMALL, K. A. (2005). Valuing time and reliability : Assessing the evidence from road pricing demonstrations. *Transportation Research Part A*, 39:279–293.
- BRUECKNER, J. K. (2000). Urban sprawl : Diagnosis and remedies. *International Regional Science Review*, 23(2):160–171.
- BRUECKNER, J. K., THISSE, J.-F. et ZENOU, Y. (1999). Why is central Paris rich and downtown Detroit poor ? An amenity-based theory. *European Economic Review*, 43:91–107.
- BRUEGMAN, R. (2005). *Sprawl : A Compact History*. University of Chicago Press, Chicago, Illinois.
- BUCHANAN, J. M. (1965). An economic theory of club. *Economica, New Series*, 32(125):1–14.
- BUISSON, C. et LADIER, C. (2009). Exploring the impact of the homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. *In 88th Annual Meeting of the Transportation Research Board CD-Rom*, Washington, DC. TRB.
- CALFEE, J. et WINSTON, C. (1998). The value of automobile travel time : Implications for congestion policy. *Journal of Public Economics*, 69:83–102.
- CALIPER (2007). *Travel Demand Modeling with TransCAD User's Guide*. Caliper Corporation.
- CASTEL, J.-C. (2005). La marché favorise-t-il la densification ? peut-il produire de l'habitat alternatif à la maison individuelle ? *In Colloque ADEF*, Lyon. CERTU.
- CASTEL, J.-C. (2006). Les coûts de la ville dense ou étalée. *Études Foncières*, 119:18–21.
- CCTN (2009). Les comptes des transports 2008 – tome 2 : Les dossiers d'analyse économique des politiques publiques des transports. 46^{ème} rapport, Commission des Comptes des Transports de la Nation.

- CEC (1995). Toward fair and efficient pricing in transport – policy options for internalising the external costs of transport in the European Union. Green paper, Commission of the European Communities, Brussels.
- CHARRON, M. (2007). *La relation entre la forme urbaine et la distance de navettage : les apports du concept de « possibilité de navettage »*. Thèse de doctorat, Université du Québec à Montréal, Montréal.
- CHEVASSON, G. (2007). L'influence relative des différentes valeurs tutélaires : une étude par la sensibilité des indicateurs socio-économiques. In MAURICE, J. et CROZET, Y., éditeurs : *Le calcul économique dans le processus de choix collectif des investissements de transport*, chapitre 6, pages 191–220. Economica, Paris.
- CLARK, C. (1968). *Population Growth and Land Use*. Macmillan and Co., Ltd, London.
- COHEN, S. (2005). *Cours d'ingénierie du trafic*. Ecole des Ponts ParisTech, Marne-la-Vallée, France.
- COHEN, S., ZHANG, M. et P., G. (2001). Gestion du trafic - les modèles de prévision : les relations temps de parcours-débit sur le réseau routier d'Île-de-France, un outil pour la planification. *Revue Générale des routes*, 792:78–84.
- COLES, M. G. et SMITH, E. (1998). Marketplaces and matching. *International Economic Review*, 39(1):239–255.
- COMBES, P.-P. (2000a). Economic structure and local growth : France, 1984-1993. *Journal of Urban Economics*, 47(3):329–355.
- COMBES, P.-P. (2000b). Marshall-Arrow-Romer externalities and city growth. CERAS Working Paper n°99-06.
- COOKE, T. W. (1983). Testing a model of intraurban firm relocation. *Journal of Urban Economics*, 13(3):257–282.
- COSTES, N. (2008). *Choix de localisation des entreprises, intervention publique et efficacité urbaine*. Thèse de doctorat, Université Paris 1, Paris.
- COULOMBEL, N., LEURENT, F. et DESCHAMPS, M. (2007). Residential choice and households strategies in the Greater Paris region. In *European Transport Conference Proceedings*, Noordwijkerhout, Netherlands. Association for European Transport.
- CROPPER, M. L. et GORDON, P. L. (1991). Wasteful commuting : A re-examination. *Journal of Urban Economics*, 29(1):2–13.

- DAGANZO, C. F. (1994). The cell transmission model, part I : A simple dynamic representation of highway traffic. *Transportation Research Part B*, 28(4):269–287.
- DAGANZO, C. F. (1995). The cell transmission model, part II : Network traffic. *Transportation Research Part B*, 29B(2):79–93.
- DAGANZO, C. F. et SHEFFI, Y. (1977). On stochastic models of traffic assignment. *Transportation Science*, 11(3):253–274.
- DAVEZIES, L. (2008). *La République et ses territoires*. La République des Idées. Seuil, Paris.
- DAVIDSON, K. B. (1966). A flow-travel time relationship for use in transportation planning. *Australian Road Research Bulletin*, 8(1):32–35.
- DE PALMA, A. et FONTAN, C. (2000). Enquête MADDIF : Multimotif adaptée à la dynamique des comportements de déplacement en Île-de-France. Rapport de recherche, Ministère des Transports, Paris.
- DE PALMA, A., KILANI, M., DE LARA, M. et PIPERNO, S. (2008). Cordon pricing and long term urban form : Application to Île-de-France. Hal-00348437.
- DE PALMA, A. et PICARD, N. (2006). Equilibria and information provision in risky networks with risk-adverse drivers. *Transportation Science*, 40(4):393–408.
- DE SOLÈRE, R., HUREZ, C. et MERMOUD, F. (2010). Les déplacements vers le travail : neuf vérités bonnes à dire. *Certu Le Point Sur – Mobilités et transports*, 14.
- DEAKIN, E. (2007). Les conséquences environnementales de l'étalement urbain. In *Transport, Formes urbaines et Croissance économique*, volume 137 de *CEMT – Table Ronde d'Économie des Transports*, pages 57–92, Paris. OCDE.
- DEAR, M. (1992). Understanding and overcoming the NIMBY syndrome. *Journal of the American Planning Association*, 58(3):288–300.
- DEBRINCAT, L., GOLDBERG, J., DUCHATEAU, H., KROES, E. et KOUWENHOVEN, M. (2006). Valorisation de la régularité des radiales ferrées en Île-de-France. In *Proceedings of the ATEC Congress, CD-Rom edition*, Paris. ATEC.
- DECORLA-SOUZA, P., COHEN, H., HALING, D. et HUNT, J. (1998). Using STEAM for benefit-cost analysis of transportation alternatives. *Transportation Research Record*, 1649:63–71.

- DELONS, J., COULOMBEL, N. et LEURENT, F. (2008). Pirandello, an integrated transport and land-use model for the Paris area. Paper submitted for the 88th TRB Meeting, 2009 – hal-00319087.
- DENANT-BOËMONT, L. et GOFFETTE-NAGOT, F. (2002). Choix de localisation résidentielle en agglomération et externalités de congestion : une approche par l'économie expérimentale. Rapport de recherche, ACI Villes.
- DENO, K. T. (1988). The effect of public capital on U.S. manufacturing activity : 1970 to 1978. *Southern Economic Journal*, 55:400–411.
- DESERPA, A. C. (1971). A theory of the economics of time. *The Economic Journal*, 81(324):828–846.
- DFT (2001). Perceptions of congestion. Report on qualitative research findings, Department for Transport, London.
- DIDIER, M. et PRUD'HOMME, R. (2007). Infrastructures de transport, mobilité et croissance. Rapport technique, Conseil d'Analyse Économique, La Documentation Française, Paris.
- DIXIT, A. (1973). The optimum factory town. *The Bell Journal of Economics and Management Science*, 4(2):637–651.
- DOWNS, A. (1962). The law of peak-hour expressway congestion. *Traffic Quarterly*, 16(3):393–409.
- DREIF (2004a). les déplacements des Franciliens. Enquête globale de transport, Direction Régionale de l'Équipement d'Île-de-France, Paris.
- DREIF (2004b). Modélisation du réseau routier d'Île-de-France. Actualisation des réseaux VP de la Dreif, Direction Régionale de l'Équipement d'Île-de-France, Paris.
- DREIF (2008). MODUS v2.1. Documentation détaillée du modèle de déplacements de la Dreif, Direction Régionale de l'Équipement d'Île-de-France, Paris.
- DREYFUS, J. (2005). Le profil des déplacements journaliers en transports en commun et voiture particulière. *Les Cahiers de l'Enquête Globale Transport*, (2).
- DUPUY, G. (1995). *Les territoires de l'automobile*. Collection Villes. Anthropos, Paris.
- DURANTON, G. (1998). Labor specialization, transport costs, and city size. *Journal of Regional Science*, 38(4):553–573.

- DURANTON, G. et PUGA, D. (2000). Diversity and specialisation in cities : Why, where and when does it matter ? *Urban Studies*, 37(3):533–555.
- DURANTON, G. et PUGA, D. (2001). Nursery cities : Urban diversity, process innovation, and the life cycle of products. *American Economic Review*, 91(5): 1454–1477.
- DURANTON, G. et PUGA, D. (2004). Micro-foundations of urban agglomeration economies. In HENDERSON, V. J. et THISSE, J.-F., éditeurs : *Handbook of Regional and Urban Economics*, volume 4, chapitre 48, pages 2063–2117. Elsevier North-Holland, Amsterdam.
- DURANTON, G. et TURNER, M. (2008a). Urban growth and transportation. CEPR Discussion Paper No. DP6633.
- DURANTON, G. et TURNER, M. A. (2008b). The fundamental law of highway congestion : Evidence from the US. Draft.
- EATON, J. et ECKSTEIN, Z. (1997). Cities and growth : Theory and evidence from france and japan. *Regional Science and Urban Economics*, 27(4–5):443–474.
- EBERTS, R. W. (1981). An empirical investigation of intraurban wage gradients. *Journal of Urban Economics*, 10(1):50–60.
- EBERTS, R. W. et MCMILLEN, D. P. (1999). Agglomeration economies and urban public infrastructure. In MILLS, E. S. et CHESHIRE, P., éditeurs : *Handbook of Regional and Urban Economics*, volume 3, chapitre 38, pages 1455–1495. Elsevier North-Holland, Amsterdam.
- ECMT (2007). Managing urban traffic congestion. Report of the working group on “Managing congestion in large urban areas”, European Conference of Ministers of Transport, Organisation for Economic Co-operation and Development, Joint Transport Research Centre.
- ECOFFEY, F. et PFLIEGER, G. (2010). Évaluation des coûts et des modalités de financement de l'étalement urbain pour les services d'eau potable – le cas de Lausanne. *Flux*, 79/80:16–33.
- FLORIAN, M. et NGUYEN, S. (1978). A combined trip distribution modal split and trip assignment model. *Transportation Research*, 12(4):241–246.
- FOUCHIER, V. (1997). *Les densités urbaines et le développement durable – Le cas de l'Île-de-France et des villes nouvelles*. Edition du SGVN, Paris.

- FOUCHIER, V. (1998). Influence de la densité urbaine sur les déplacements en Île-de-France. *Transports Urbains*, (99):21–24.
- FTA (2000). Transit benefits 2000 working papers, a public choice policy analysis. Rapport technique, Federal Transit Administration, US Department of Transportation, Washington, DC.
- FTA (2009). Reporting instructions for the section 5309 New Starts Criteria. Rapport technique, Federal Transit Administration, US Department of Transportation, Washington, DC.
- FUJITA, M. (1982). Spatial patterns of residential development. *Journal of Urban Economics*, 12(1):22–52.
- FUJITA, M. (1985). Existence and uniqueness of equilibrium and optimal land use : Boundary rent curve approach. *Regional Science and Urban Economics*, 15(2):295–324.
- FUJITA, M. (1989). *Urban Economic Theory : Land Use and City Size*. Cambridge University Press, New York, New York.
- FUJITA, M., KRUGMAN, P. et VENABLES, A. J. (2001). *The Spatial Economy — Cities, Regions, and International Trade*. The MIT Press, Cambridge, Massachusetts.
- FUJITA, M. et OGAWA, H. (1982). Multiple equilibria and structural transition of non-monocentric urban configurations. *Regional Science and Urban Economics*, 12(2):161–196.
- FUJITA, M. et THISSE, J.-F. (1996). Economics of agglomeration. *Journal of the Japanese and International Economies*, 10:339–378.
- FUJITA, M. et THISSE, J.-F. (2002). *Economics of Agglomeration - Cities, Industrial Location, and Regional Growth*. Cambridge University Press, Cambridge.
- FUJITA, M. et THISSE, J.-F. (2009). New economic geography : An appraisal on the occasion of Paul Krugman’s 2008 Nobel Prize in economics. *Regional Science and Urban Economics*, 39(2):109–119.
- GAUDRY, M. (2007). Structure de la modélisation du trafic et théorie économique. In MAURICE, J. et CROZET, Y., éditeurs : *Le calcul économique dans le processus de choix collectif des investissements de transport*, chapitre 1, pages 6–97. Economica, Paris.

- GENRE-GRANDPIERRE, C. (2007). Des « réseaux lents » contre la dépendance automobile? concept et implications en milieu urbain. *L'Espace Géographique*, 2007(1):27–39.
- GEROLIMINIS, N. et DAGANZO, C. F. (2007). Macroscopic modeling of traffic in cities. In *86th Annual Meeting of the Transportation Research Board CD-Rom*, Washington, DC. TRB.
- GEROLIMINIS, N. et DAGANZO, C. F. (2008). Existence of urban-scale macroscopic fundamental diagrams : Some experimental findings. *Transportation Research Part B*, 42:759–770.
- GIULIANO, G. et SMALL, K. A. (1993). Is the journey to work explained by urban structure? *Urban Studies*, 30(9):1485–1501.
- GLAESER, E. L. et KAHN, M. E. (2001). Decentralized employment and the transformation of the American city. *Brookings-Wharton Papers on Urban Affairs*, pages 1–63.
- GLAESER, E. L. et KAHN, M. E. (2008). The greenness of cities. Policy briefs, Harvard University – Rappaport Institute for Greater Boston / Taubman Center for State and Local Government.
- GOFFETTE-NAGOT, F. (2000). Urban spread beyond the city edge. In HURIOT, J.-M. et THISSE, J.-F., éditeurs : *Economics of Cities : Theoretical Perspectives*, chapitre 9, pages 318–340. Cambridge University Press.
- GRANGEON, D. (2010). Monétarisation des externalités environnementales. Rapport d'études, Sétra, Bagneux, France.
- GUILLAIN, R., LE GALLO, J. et BOITEUX-ORAIN, C. (2006). Changes in spatial and sectoral patterns of employment in Île-de-France, 1978-97. *Urban Studies*, 43(11):2075–2098.
- HAIGHT, F. (1963). *Mathematical Theories of Traffic Flow*. Academic Press, New York.
- HAMILTON, B. W. (1982). Wasteful commuting. *Journal of Political Economy*, 90(5):1035–1053.
- HAMILTON, B. W. (1989). Wasteful commuting again. *The Journal of Political Economy*, 97(6):1497–1504.
- HARDIN, G. (1968). The tragedy of the commons. *Science*, 162(13):1243–1248.

- HAUTREUX, J. (1969). Sur les coûts et la tarification des transports urbains. Rapport technique, Commission d'étude des coûts d'infrastructure de transport, Paris.
- HAYWOOD, L. et KONING, M. (2010). Modal shifts and pushy Parisian elbows – Evidence of taste for comfort in travel based on the contingent valuation methodology. In *Conference "Public Policies and Industrial Organization in the City"*, Lille. Equippe.
- HEATCO (2006). Developing harmonized European approaches for transport costing and project assessment. Deliverable 5 – proposal for harmonized guidelines, IER, Germany.
- HELSLEY, R. W. et STRANGE, W. C. (1990). Matching and agglomeration economies in a system of cities. *Regional Science and Urban Economics*, 20(2):189–212.
- HENDERSON, V. J. (1974). The sizes and types of cities. *The American Economic Review*, 64(4):640–656.
- HENDERSON, V. J., KUNCORO, A. et TURNER, M. A. (1995). Industrial development in cities. *The Journal of Political Economy*, 103(5):1067–1090.
- HOCHMAN, O. (2010). Efficient agglomeration of spatial clubs. *Journal of Urban Economics*.
- HOLDEN, D. J. (1989). Wardrop's third principle - Urban traffic congestion and traffic policy. *Journal of Transport Economics and Policy*, 23(3):239–262.
- HORNER, M. W. (2002). Extensions of the concept of excess commuting. *Environment and Planning A*, 34:543–566.
- HOTELLING, H. (1929). Stability in competition. *Economic Journal*, 39:41–57.
- HURIOT, J.-M. et BOURDEAU-LEPAGE, L. (2009). *Économie des villes contemporaines*. Economica, Paris.
- HURIOT, J.-M. et THISSE, J.-F., éditeurs (2000). *Economics of Cities : Theoretical Perspectives*. Cambridge University Press, Cambridge.
- IAU-IDF (2009). Un portrait par les chiffres. <http://www.iau-idf.fr/lile-de-france/un-portrait-par-les-chiffres/>.
- IAURIF-INSEE (2003). *Atlas des Franciliens, Tome 4*. IAURIF – INSEE, Paris.
- IMAI, H. (1982). CBD hypothesis and economies of agglomeration. *Journal of Economic Theory*, 28:275–299.

- JARA-DIAZ, S. (1986). On the relations between user's benefits and the economics of transportation activities. *Journal of Regional Science*, 26(2):379–391.
- JARA-DIAZ, S. (2003). On the goods-activities technical relations in the time allocation theory. *Transportation*, 30(3):245–260.
- JARA-DIAZ, S. (2007). *Transport Economic Theory*. Elsevier Science.
- JIWATTANAKULPAISARN, P. (2008). *The Impact of Transport Infrastructure Investment on Regional Employment : An Empirical Investigation*. Thèse de doctorat, Imperial College, London.
- JOBERT, A. (1998). L'aménagement en politique. Ou ce que le syndrome NIMBY nous dit de l'intérêt général. *Politix*, 11(42):67–92.
- KAHN, M. E. (2007). La qualité de la vie et la productivité dans les villes étalées par opposition aux villes denses aux États-Unis. In *Transport, Formes urbaines et Croissance économique*, volume 137 de *CEMT – Table Ronde d'Économie des Transports*, pages 93–120, Paris. OCDE.
- KEELER, T. E. et SMALL, K. A. (1977). Optimal peak-load pricing, investment, and service levels on urban expressways. *The Journal of Political Economy*, 85(1):1–25.
- KILANI, M., LEURENT, F. et DE PALMA, A. (2010). A monocentric city with discrete transit stations. In *89th Annual Meeting of the Transportation Research Board CD-Rom*, Washington, DC. TRB.
- KIM, S. (1990). Labor heterogeneity, wage bargaining, and agglomeration economies. *Journal of Urban Economics*, 28(2):160–177.
- KIM, S. (1991). Heterogeneity of labor markets and city size in an open spatial economy. *Regional Science and Urban Economics*, 21(1):109–126.
- KOGUT, B. et ZANDER, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science*, 3(3):383–397.
- KONING, M. (2009). La congestion du Boulevard Périphérique parisien. Estimations, évolutions 2000-2007, discussions. CES Working Papers, Hal SHS-00363389.
- KORSU, E. et MASSOT, M.-H. (2006). Rapprocher les ménages de leurs lieux de travail : les enjeux pour la régulation de l'usage de la voiture en Île-de-France. *Les Cahiers Scientifiques du Transport*, 50:61–90.

- KRUGMAN, P. (1991a). *Geography and Trade*. MIT Press, Cambridge, Massachusetts.
- KRUGMAN, P. (1991b). Increasing returns and economic geography. *Journal of Political Economy*, 99:483–499.
- KRUGMAN, P. (1998). *L'économie auto-organisatrice*. De Boeck, Bruxelles. French Traduction by F. Leloup.
- KWAN, M.-P. (1999). Gender, the home-work link, and space-time patterns of nonemployment activities. *Economic Geography*, 75(4):370–394.
- LECLERCQ, L. (2002). *Modélisation dynamique du trafic et applications à l'estimation du bruit routier*. Thèse de doctorat, INSA, Lyon.
- LEURENT, F. (1999). Accessibility to vacant activities : a novel model of destination choice. In *European Transport Conference Proceedings*, Cambridge, UK. Association for European Transport.
- LEURENT, F. (2001). Modèles désagrégés du trafic. Rapport Outils et Méthodes n°10, INRETS, Arcueil, France.
- LEURENT, F. (2003). On network assignment and supply-demand equilibrium : An analysis framework and a simple dynamic model. In *European Transport Conference Proceedings*, Strasbourg, France. Association for European Transport.
- LEURENT, F. (2005). La capacité d'écoulement du trafic. Rapport n° 265, INRETS, Arcueil, France.
- LEURENT, F. (2006). *Structures de réseau et modèles de cheminement*. Tec & Doc. Lavoisier, Paris.
- LEURENT, F. (2007). Économie du système routier. Rapport introductif du comité technique sur l'économie du système routier, Association Mondiale de la Route (AIPCR), Paris.
- LEURENT, F. (2009). *Méthodes d'Analyse des Systèmes Territoriaux – Document de cours*. Ecole des Ponts ParisTech, Marne-la-Vallée, France.
- LEURENT, F. (2010). Les contraintes de capacité en transport public de voyageurs : une analyse systémique. Document de travail. Version anglaise soumise à European Transport Research Review.
- LEURENT, F. et AGUILÉRA, V. (2009). Large problems of dynamic network assignment and traffic equilibrium : Computational principles and application to Paris road network. *Transportation Research Record*, 2132:122–132.

- LEURENT, F. et BRETEAU, V. (2009). On the marginal cost of road congestion : An evaluation method with application to the Paris region. *In European Transport Conference Proceedings*, Noordwijkerhout, Netherlands. Association for European Transport.
- LEURENT, F. et BRETEAU, V. (2010). Is a quick method relevant to evaluate the benefits of a transit scheme to the road users? *In 89th Annual Meeting of the Transportation Research Board CD-Rom*, Washington, DC. TRB.
- LEURENT, F., BRETEAU, V. et WAGNER, N. (2009). Coût social marginal de la congestion routière — Actualisation et critique de l'approche Hautreux. Rapport technique, DGITM - Ministère de l'Ecologie, de l'Energie, du Développement Durable et de l'Aménagement du Territoire, Paris.
- LEURENT, F. et LIU, K. (2009). On seat congestion, passenger comfort and route choice in urban transit : A network equilibrium model with application to Paris. *In 88th Annual Meeting of the Transportation Research Board CD-Rom*, Washington, DC. TRB.
- LEURENT, F. et NGUYEN, T.-P. (2010). Dynamic information and its value to individual users and to traffic : Probabilistic model with economic analysis. *In 89th Annual Meeting of the Transportation Research Board CD-Rom*, Washington, DC. TRB.
- LIGHTHILL, M. et WHITHAM, G. (1955). On kinetic waves, II – a theory of traffic flow on long crowded roads. *In Proceedings of the Royal Society*, volume 229A, pages 317–345, London.
- LINDSAY, C., GAUNT, G. et ELSTON, I. (2007). Derivation of link-based marginal social cost based charges in very large assignment models. *In European Transport Conference Proceedings*, Noordwijkerhout, Netherlands. Association for European Transport.
- LINDSEY, R. et VERHOEF, E. T. (2000). Congestion modelling. *In HENSHER, D. A. et BUTTON, K. J., éditeurs : Handbook of Transport Modelling*, chapitre 21, pages 353–373. Pergamon, Oxford.
- LU, S. et NIE, Y. M. (2010). Stability of user-equilibrium route flow solutions for the traffic assignment problem. *In 89th Annual Meeting of the Transportation Research Board CD-Rom*, Washington, DC. TRB.
- LUCAS, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, 22(1):3–42.

- MA, K.-R. et BANISTER, D. (2006). Excess commuting : A critical review. *Transport Reviews*, 26(6):749–767.
- MACKIE, P. J., WARDMAN, A., FOWKES, A. S., WHELAN, G., NELLTHORP, J. et BATES, J. (2003). Values of travel time savings in the uk. Report to department for transport, Institute of Transport Studies, University of Leeds.
- MAIBACH, M., SCHREYER, C., SUTTER, D., van HESSEN, H., BOON, B., SMOKERS, A., DOLL, C., PAWLOWSKA, B. et BAK, M. (2008). Handbook on estimation of external costs in the transport sector. Internalisation measures and policies for all external cost of transport (IMPACT) report, CE Delft, Delft.
- MARSHALL, A. (1890). *Principles of Economics*. Macmillan and Co., Ltd, London, 1st édition. Trad. franç. : F. Sauvaire-Jourdan, 1906.
- MARSHALL, A. (1920). *Industry and Trade*. Macmillan and Co., Ltd, London, 3rd édition.
- MAURICE, J. et CROZET, Y., éditeurs (2007). *Le calcul économique dans le processus de choix collectif des investissements de transport*. Méthodes et Approches. Economica, Paris.
- MCDONALD, J. F. (2009). Calibration of a monocentric city model with mixed land use and congestion. *Regional Science and Urban Economics*, 39(1):90–96.
- MCMILLEN, D. P. et MCDONALD, J. F. (1998). Suburban subcenters and employment density in metropolitan chicago. *Journal of Urban Economics*, 43(2):157–180.
- MCMILLEN, D. P. et SINGELL, L. J. (1992). Work location, residence location, and the intraurban wage gradient. *Journal of Urban Economics*, 32(2):195–213.
- MELIA, S., PARKHURST, G. et BARTON, H. (2011). The paradox of intensification. *Transport Policy*, 18(1):46–52.
- MEYERE, A., COUREL, J. et NGUYEN-LUONG, D. (2005). L'impact des modes de vie sur les déplacements. *Les Cahiers de l'Enquête Globale Transport*, (4).
- MIESZKOWSKI, P. et MILLS, E. S. (1993). The causes of metropolitan suburbanization. *The Journal of Economic Perspectives*, 7(3):135–147.
- MILLS, E. S. (1967). An aggregative model of resource allocation in a metropolitan area. *The American Economic Review*, 57(2):197–210.

- MILLS, E. S. (1972). *Studies in the Structure of the Urban Economy*. Johns Hopkins Press (for Resources for the Future), Baltimore.
- MILLS, E. S. et HAMILTON, B. W. (1994). *Urban Economics*. Harper Collins, New York, 5th édition.
- MIRRLEES, J. A. (1972). The optimum town. *The Swedish Journal of Economics*, 74(1):114–135.
- MOHRING, H. (1999). Congestion. In GÓMEZ-IBÁÑEZ, J. A., TYE, W. B. et WINSTON, C., éditeurs : *Essays in Transportation Economics and Policy*, chapitre 6, pages 181–221. Brookings Institution Press, Washington, DC.
- MOHRING, H. et HARWITZ, M. (1962). *Highway Benefits : An Analytical Framework*. Northwestern University Press, Evanston, IL.
- MUTH, R. F. (1969). *Cities and Housing*. University of Chicago Press, Chicago.
- NEWBERY, D. M. (1990). Pricing and congestion : Economic principles relevant to pricing roads. *Oxford Review of Economic Policy*, 6(2):22–38.
- NEWELL, G. F. (1965). Approximation methods for queues with application to the fixed-cycle traffic light. *SIAM Review*, 7(2):223–240.
- NEWMAN, P. et KENWORTHY, J. (1989). Gasoline consumption and cities. *Journal of the American Planning Association*, 55(1):24–37.
- NOLAND, R. et SMALL, K. A. (1995). Travel-time uncertainty, departure time choice, and the cost of morning commutes. *Transportation Research Record*, 1493:150–158.
- NOUAILLE, P. (2009). Coûts collectifs d'urbanisation, densité et formes urbaines. Rapport à la DDEA du Maine et Loire, CETE Ouest, MEEDM, Nantes, France.
- OGAWA, H. et FUJITA, M. (1980). Equilibrium land use patterns in a non-monocentric city. *Journal of Regional Science*, 20(4):445–475.
- ORFEUIL, J.-P. (1995). Les déplacements domicile-travail dans l'enquête transports INSEE 1993-1994. Rapport de recherche, INRETS.
- ORTÚZAR, J. d. D. et WILLUMSEN, L. G. (2001). *Modelling Transport*. John Wiley & Sons, Ltd, Chichester, England, 3rd édition.
- OTA, M. et FUJITA, M. (1993). Communication technologies and spatial organization of multi-unit firms in metropolitan areas. *Regional Science and Urban Economics*, 23:695–729.

- PALIVOS, T. et WANG, P. (1996). Spatial agglomeration and endogeneous growth. *Regional Science and Urban Economics*, 26:645–669.
- PICARD, P. (2007). *Éléments de microéconomie*, volume 1. Théorie et applications. Montchrestien, Paris, 7ème édition.
- PIGOU, A. C. (1920). *The Economics of Welfare*. Macmillan and Co., Ltd, London.
- POLACCHINI, A. et ORFEUIL, J.-P. (1999). Les dépenses des ménages franciliens pour le logement et les transports. *Recherche Transports Sécurité*, 63:31–46.
- POTTIER, P., SALEMBIER, L. et FRANCASTEL, S. (2007). Pôles d’emploi franciliens : quatre emplois sur dix dans les services à la production. *INSEE Ile-de-France – À la Page*, 287.
- POUYANNE, G. (2004). *Forme urbaine et mobilité quotidienne*. Thèse de doctorat, Université Montesquieu Bordeaux IV, Bordeaux.
- PRUD’HOMME, R. (1999a). La congestion et ses coûts. Mimeo.
- PRUD’HOMME, R. (1999b). Les coûts de la congestion dans la région parisienne. *Revue d’Economie Politique*, 109(4):425–441.
- PRUD’HOMME, R. et SUN, Y.-M. (1999). Trois essais sur la congestion et son coût. L’OEIL/IUP. Polygr. 49 p., Créteil.
- PRUD’HOMME, R. et SUN, Y.-M. (2000). Le coût économique de la congestion du périphérique parisien : une approche désagrégée. *Les Cahiers Scientifiques du Transport*, 37:59–73.
- QIAN, Z. et ZHANG, H. M. (2010). On centroids connectors in static traffic assignment : Their effects on flow patterns and how to optimize their selection. In *89th Annual Meeting of the Transportation Research Board CD-Rom*, Washington, DC. TRB.
- QUINET, E. (1998). *Principes d’Économie des Transports*. Economica, Paris.
- RAUX, C. (1993). Centralité, polynucléarité et étalement urbain : application au cas de l’agglomération lyonnaise. In BUSSIÈRE, Y. et BONNAFOUS, A., éditeurs : *Transport et étalement urbain : les enjeux*, Collection Les Chemins de la Recherche, pages 75–98. Programme Rhône-Alpes – Recherche en Sciences Humaines, Lyon.
- RICHARDS, P. (1956). Shock waves on the highway. *Operations Research*, 4(1):42–51.

- RIFF BREMS, C., KRISTENSEN, N. B. et SLOTH, B. (2002). Congestion costs. *In European Transport Conference Proceedings*, Cambridge, UK. Association for European Transport.
- RIGUELLE, F., THOMAS, I. et VERHETSEL, A. (2007). Measuring urban polycentrism : A european case study and its implications. *Journal of Economic Geography*, 7:193–215.
- ROSENTHAL, S. S. et STRANGE, W. C. (2004). Evidence on the nature and sources of agglomeration economies. *In HENDERSON, V. J. et THISSE, J.-F., éditeurs : Handbook of Regional and Urban Economics*, volume 4, chapitre 49, pages 2119–2171. Elsevier North-Holland, Amsterdam.
- ROUCHAUD, d. et SAUVANT, A. (2004). Prix des logements et coûts de transports : un modèle global d'équilibre en Île-de-France. *Notes de synthèse du SES*, 154.
- ROVOLIS, A. et SPENCE, N. (2002). Duality theory and cost function analysis in a regional context : The impact of public infrastructure capital in the Greek regions. *The Annals of Regional Science*, 36(1):55–78.
- SALANIÉ, B. (1998). *Microéconomie : les défaillances du marché*. Economica, Paris.
- SALOP, S. C. (1979). Monopolistic competition with outside goods. *Bell Journal of Economics*, 10(1):141–156.
- SCHELLING, T. C. (1969). Models of segregation. *The American Economic Review*, 59(2):488–493.
- SCHWANEN, T., DIELEMAN, F. M. et DIJST, M. (2004). The impact of metropolitan structure on commute behavior in the netherlands : A multilevel approach. *Growth and Change*, 35(3):304–333.
- SCITOVSKY, T. (1954). Two concepts of external economies. *Journal of Political Economy*, 62:143–151.
- SDRIF (2008). *Structure et fonctionnement du réseau routier*, chapitre Mobilité et Transport en Île-de-France. Région Île-de-France, Paris.
- SEITZ, H. (1995). The productivity and supply of urban infrastructures. *The Annals of Regional Science*, 29(2):121–141.
- SESP (2007). Simulations à partir d'un modèle de ville circulaire et monocentrique – Application à l'Île-de-France. Document de travail.

- SHEARMUR, R. (2006). Travel from home : An economic geography of commuting distances in Montreal. *Urban Geography*, 27(4):330–359.
- SHOUP, D. C. (2006). Cruising for parking. *Transport Policy*, 13:479–486.
- SMALL, K. A. (1999). Project evaluation. In GÓMEZ-IBÁÑEZ, J. A., TYE, W. B. et WINSTON, C., éditeurs : *Essays in Transportation Economics and Policy*, chapitre 5, pages 137–177. Brookings Institution Press, Washington, DC.
- SMALL, K. A., NOLAND, R., CHU, X. et LEWIS, D. (1999). Valuation of travel-time savings and predictability in congested conditions for highway user-cost estimation. Report 431, National Cooperative Highway Research Program, Transportation Research Board, Washington, D.C.
- SMALL, K. A. et SONG, S. (1992). “Wasteful” commuting : A resolution. *The Journal of Political Economy*, 100(4):888–898.
- SMALL, K. A. et SONG, S. (1994). Population and employment densities : Structure and change. *Journal of Urban Economics*, 36(3):292–313.
- SMALL, K. A. et VERHOEF, E. T. (2007). *The Economics of Urban Transportation*. Routledge, New York.
- SMITH, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. W. Strahan and T. Cadell, London.
- SNCF (2008). Le projet tram-train Massy-Évry. Dossier d’objectifs et de caractéristiques principales, Stif, RFF, SNCF, Paris.
- SOLOW, R. M. (1972). Congestion, density and the use of land in transportation. *Swedish Journal of Economics*, 74(1):161–173.
- SOLOW, R. M. (1973a). Congestion cost and the use of land for streets. *The Bell Journal of Economics and Management Science*, 4(2):602–618.
- SOLOW, R. M. (1973b). On equilibrium models of urban location. In PARKIN, J. M., éditeur : *Essays in Modern Economics*, pages 2–16. Longman, London.
- STAHL, K. et WALZ, U. (2001). Will there be a concentration of alike? : The impact of labor market structure on industry mix in the presence of product market shocks. HWWA Discussion Paper 140.
- STANCANELLI, E. G. F. (2006). Les couples sur le marché de l’emploi. une analyse exploratoire des années récentes. *Revue de l’OFCE*, (99):235–272.

- STARRS, M. M. et STARKIE, D. N. M. (1986). An integrated road pricing and investment model : A South Australian application. *Australian Road Research*, 16:1–9.
- STEINNES, D. (1982). Do “people follow jobs” or “jobs follow people” : A causality issue in urban economics. *Urban Studies*, 19:187–192.
- STIF (2007). Tangentielle sud – étude de trafic. Note interne.
- SULLIVAN, A. M. (1983a). The general equilibrium effects of congestion externalities. *Journal of Urban Economics*, 14(1):80–104.
- SULLIVAN, A. M. (1983b). A general equilibrium model with external scale economies in production. *Journal of Urban Economics*, 13(2):235–255.
- SULLIVAN, A. M. (1983c). Second-best policies for congestion externalities. *Journal of Urban Economics*, 14(1):105–123.
- TALBOT, J. (2001). Les déplacements domicile-travail – de plus en plus d’actifs travaillent loin de chez eux. *Insee Première*, 767.
- THISSE, J.-F. (2007). Équité, efficacité et acceptabilité dans la localisation des équipements collectifs. In MAURICE, J. et CROZET, Y., éditeurs : *Le calcul économique dans le processus de choix collectif des investissements de transport*, chapitre 12, pages 361–401. Economica, Paris.
- TIMOTHY, D. et WHEATON, W. C. (2001). Intra-urban wage variations, employment location, and commuting times. *Journal of Urban Economics*, 50(2):338–366.
- TROUTBECK, R. (2000). Modeling signalized and unsignalized junctions. In HENSHER, D. A. et BUTTON, K. J., éditeurs : *Handbook of Transport Modelling*, chapitre 22, pages 375–391. Pergamon, Oxford.
- UNITED NATIONS (2010). World urbanization prospects : The 2009 revision. Rapport technique, United Nations, Department of Economics and Social Affairs, Population Division, New York.
- VAN OMMEREN, J. et GUTIÉRREZ-I-PUIGARNAU, E. (2010). Are workers with a long commute less productive? an empirical analysis of absenteeism. *Regional Science and Urban Economics*. À paraître.
- VENABLES, A. J. (1996). Equilibrium locations of vertically linked industries. *International Economic Review*, 37(2):341–359.

- VERHETSEL, A., THOMAS, I. et BEELEN, M. (2010). Commuting in Belgian metropolitan areas – the power of the Alonso-Muth model. *Journal of Transport and Land Use*, 2(3/4):109–131.
- VERHOEF, E. T. (1994). External effects and social costs of road transport. *Transportation Research Part A*, 28A(4):273–287.
- VERHOEF, E. T. (1999). Time, speeds, flows and densities in static models of road traffic congestion and congestion pricing. *Regional Science and Urban Economics*, 29(3):341–369.
- VERHOEF, E. T. (2001). An integrated dynamic model of road traffic congestion based on simple car-following theory : Exploring hypercongestion. *Journal of Urban Economics*, 49(3):505–542.
- VICKREY, W. S. (1969). Congestion theory and transport investment. *The American Economic Review*, 59(2):251–260.
- VON THÜNEN, J. H. (1826). *Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*. Perthes, Hamburg. English Translation : *The Isolated State*, Pergammon Press, Oxford 1966.
- WALTERS, A. A. (1961). The theory and measurement of private and social cost of highway congestion. *Econometrica*, 29(4):676–699.
- WARDROP, J. G. (1952). Some theoretical aspects of road traffic research. In *Proceedings of the Institution of Civil Engineers*, pages 325–362.
- WHEATON, W. C. (1974). A comparative static analysis of urban spatial structure. *Journal of Economic Theory*, 9(2):223–237.
- WHEATON, W. C. (1977). Income and urban residence : An analysis of consumer demand for location. *The American Economic Review*, 67(4):620–631.
- WHEATON, W. C. (2004). Commuting, congestion and employment dispersal in cities with mixed land-use. *Journal of Urban Economics*, 58:417–438.
- WHITE, M. J. (1976). Firm suburbanisation and urban subcenters. *Journal of Urban Economics*, 24(129–152).
- WHITE, M. J. (1988a). Location choice and commuting behavior in cities with decentralized employment. *Journal of Urban Economics*, 24:129–152.
- WHITE, M. J. (1988b). Urban commuting journeys are not “wasteful”. *The Journal of Political Economy*, 96(5):1097–1110.

- WHITE, M. J. (1999). Urban areas with decentralized employment : Theory and empirical work. In MILLS, E. S. et CHESHIRE, P., éditeurs : *Handbook of Regional and Urban Economics*, volume 3, chapitre 36, pages 1375–1412. Elsevier North-Holland, Amsterdam.
- WIEL, M. (1999). *La transition urbaine ou le passage de la ville pédestre à la ville motorisée*. Collection Architecture et Recherche. Mardaga, Paris.
- YOUNG, W. (2000). Modeling parking. In HENSHER, D. A. et BUTTON, K. J., éditeurs : *Handbook of Transport Modelling*, chapitre 24, pages 409–420. Pergamon, Oxford.

A

Appendices

A.1. Cartes complémentaires de congestion du réseau

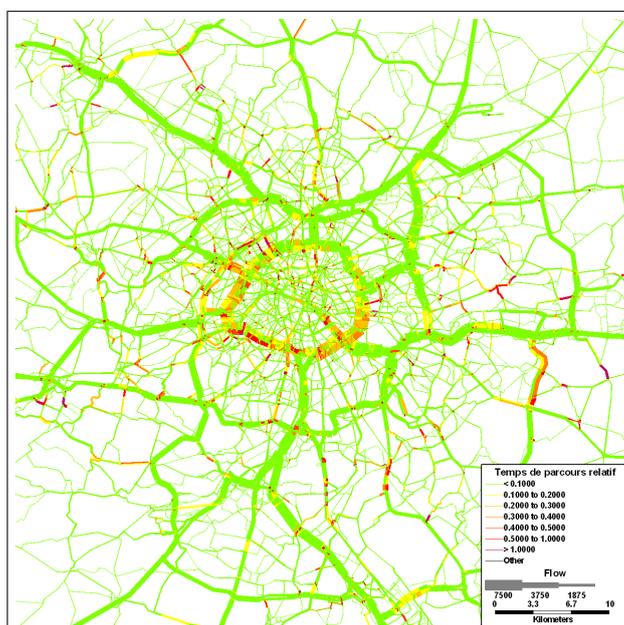


FIGURE A.1.: Les écarts de temps relatifs en heure creuse

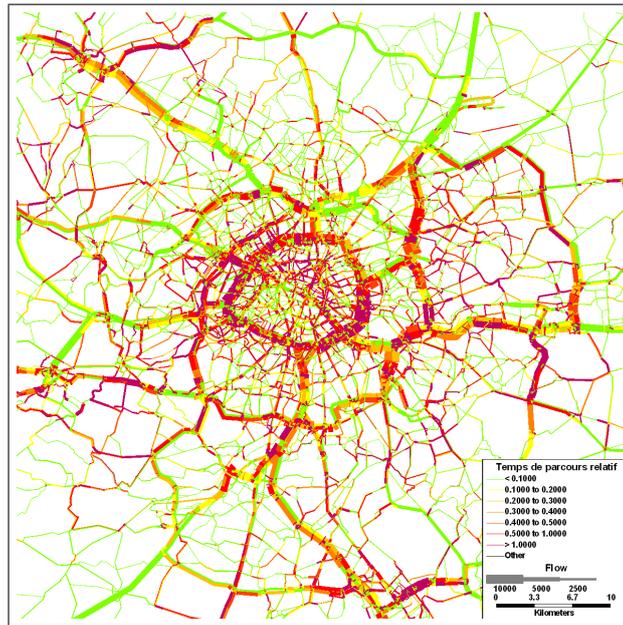


FIGURE A.2.: Les écarts de temps relatifs à l'heure de pointe du soir

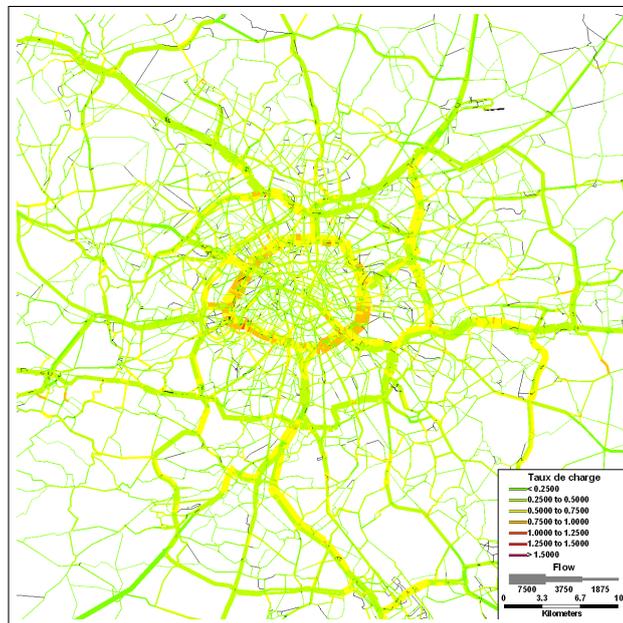


FIGURE A.3.: Les taux de charge en heure creuse

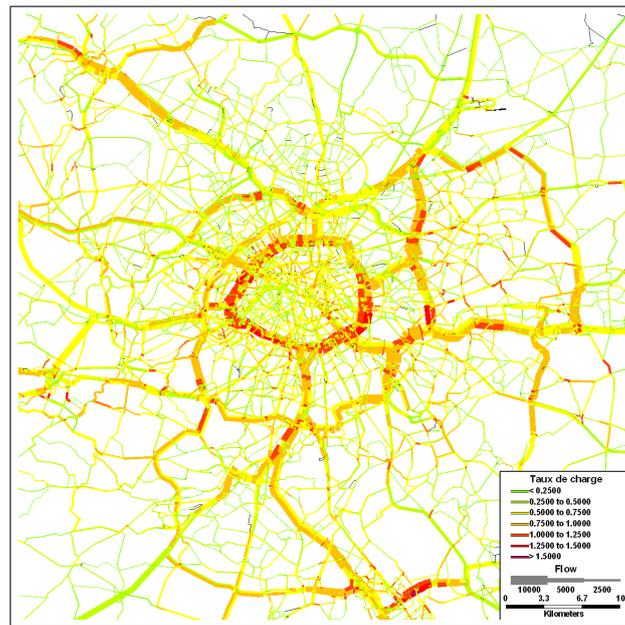


FIGURE A.4.: Les taux de charge à l'heure de pointe du soir

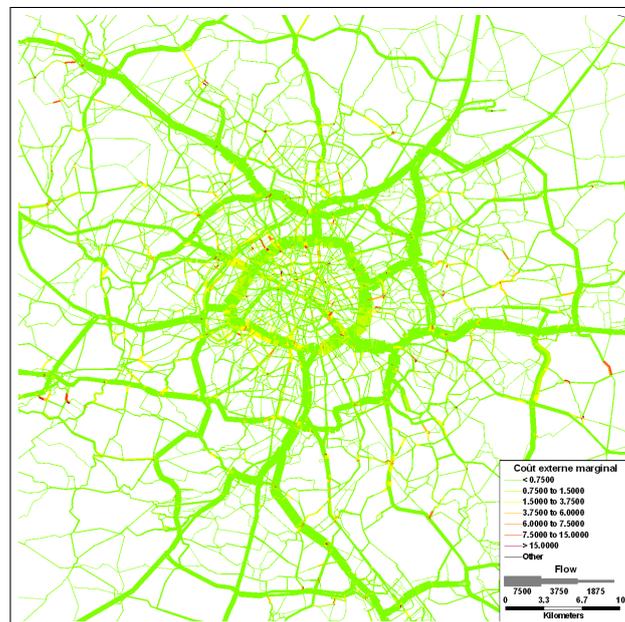


FIGURE A.5.: Le coût externe marginal en heure creuse

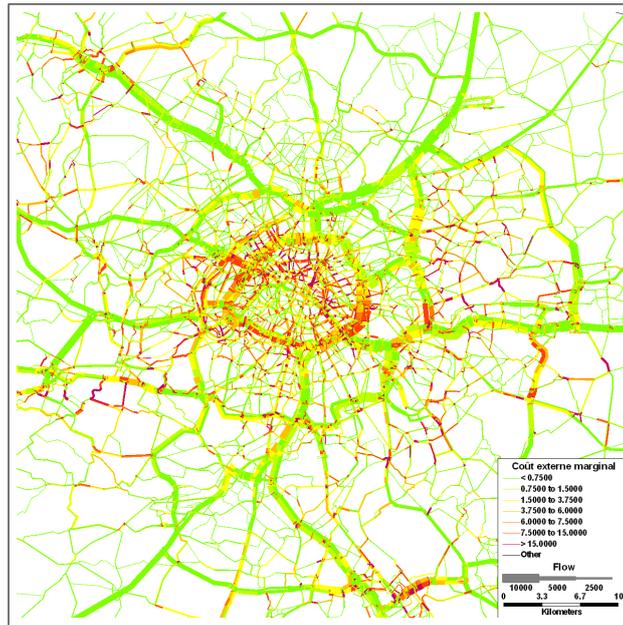


FIGURE A.6.: Le coût externe marginal à l'heure de pointe du soir

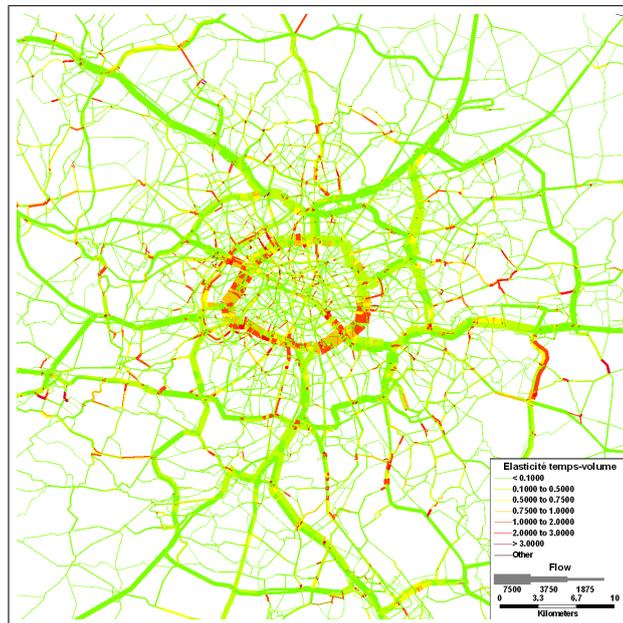


FIGURE A.7.: L'élasticité temps-volume en heure creuse

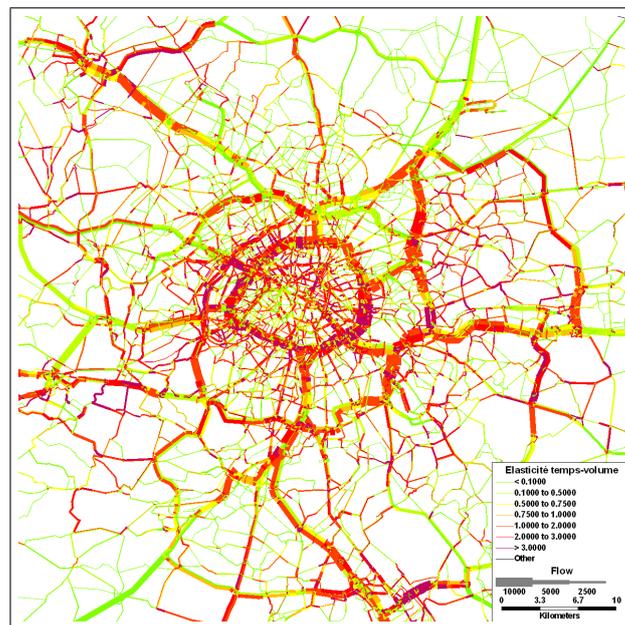


FIGURE A.8.: L'élasticité temps-volume à l'heure de pointe du soir

A.2. Cartes globales de l'Île-de-France

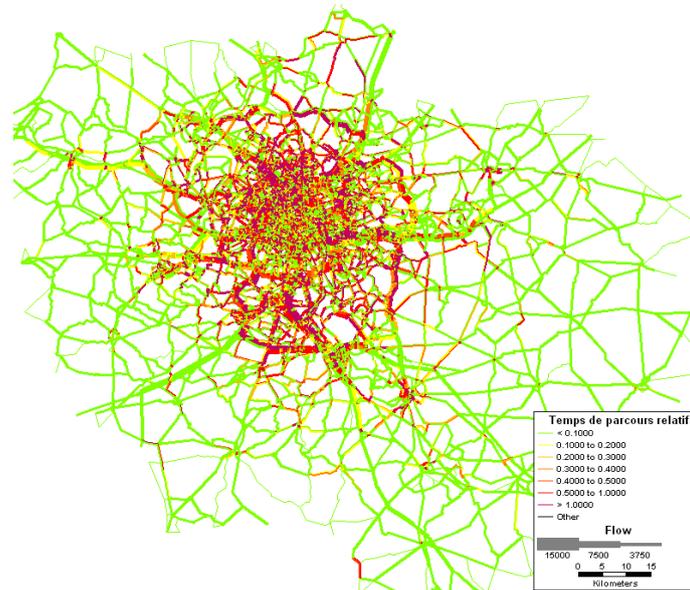


FIGURE A.9.: Le temps de parcours relatif à l'heure de pointe du matin.

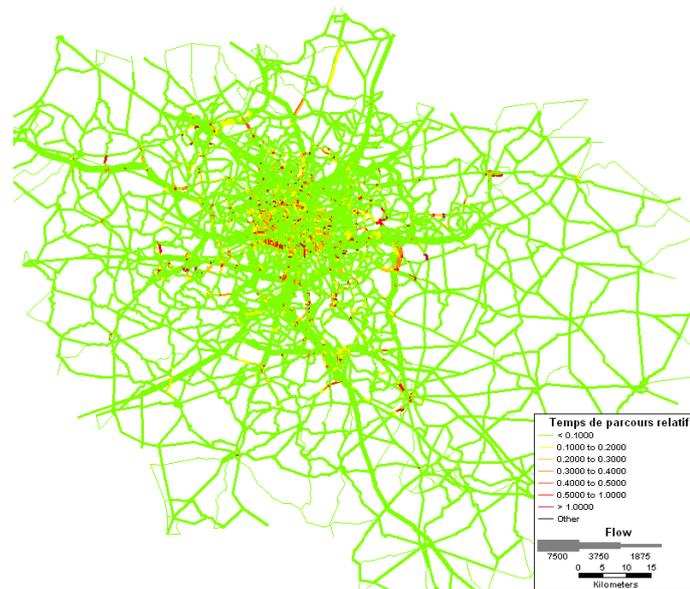


FIGURE A.10.: Le temps de parcours relatif en heure creuse.

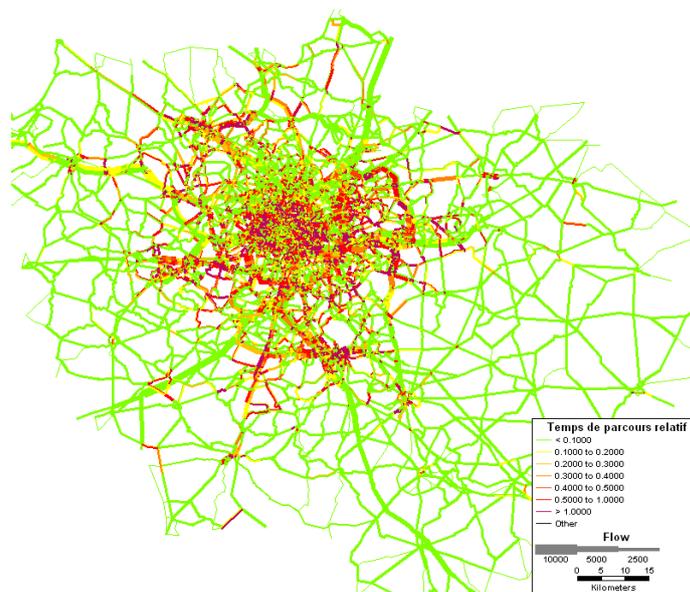


FIGURE A.11.: Le temps de parcours relatif à l'heure de pointe du soir.

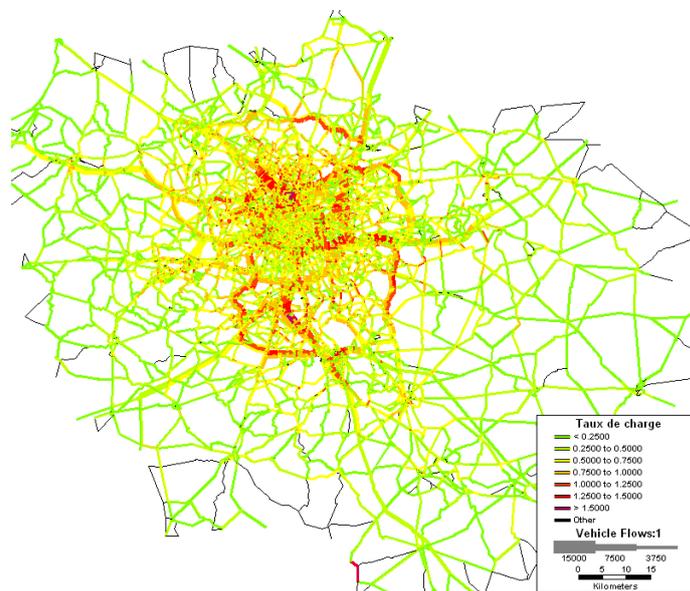


FIGURE A.12.: Le taux de charge à l'heure de pointe du matin.

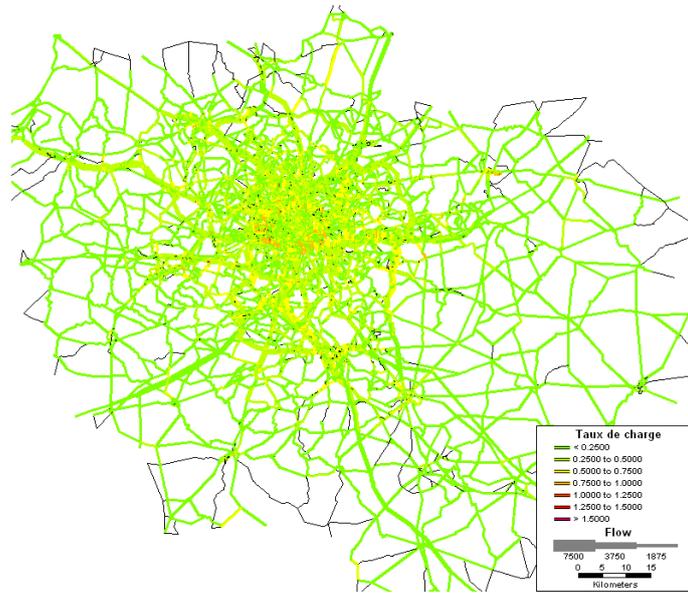


FIGURE A.13.: Le taux de charge en heure creuse.

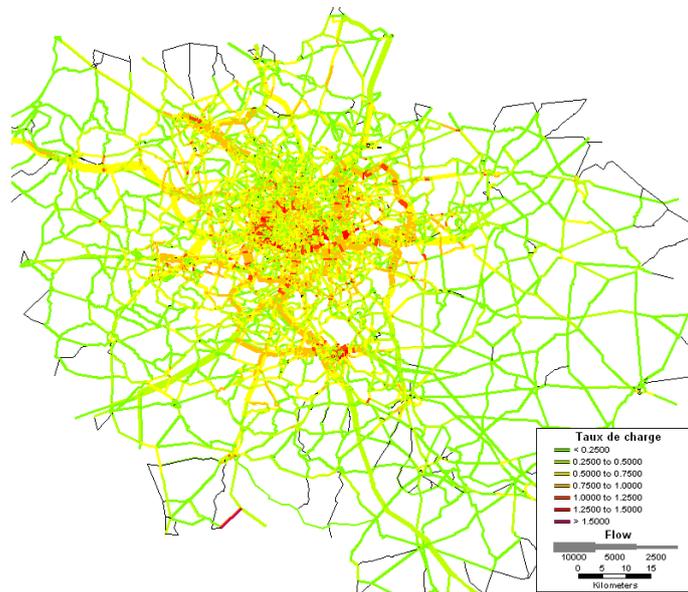


FIGURE A.14.: Le taux de charge à l'heure de pointe du soir.

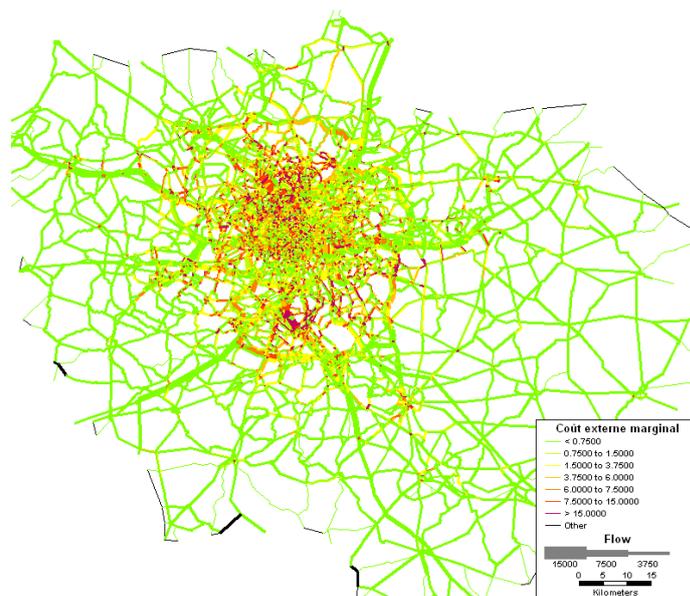


FIGURE A.15.: Le coût externe marginal à l'heure de pointe du matin.

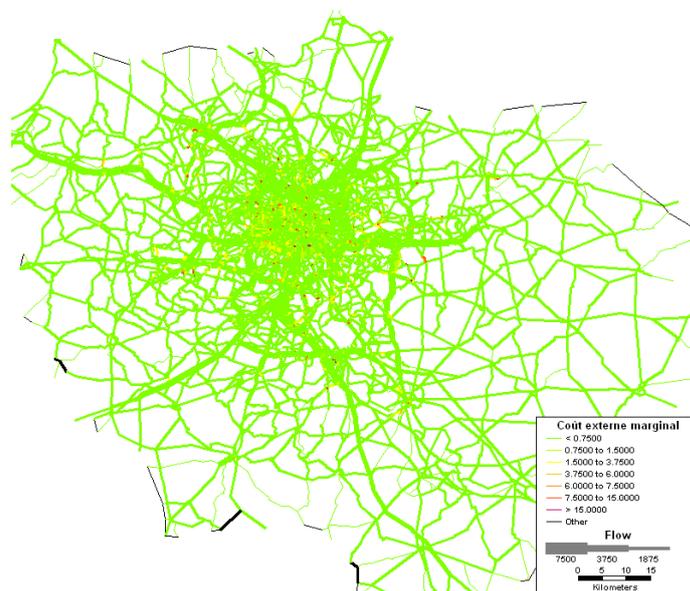


FIGURE A.16.: Le coût externe marginal en heure creuse.

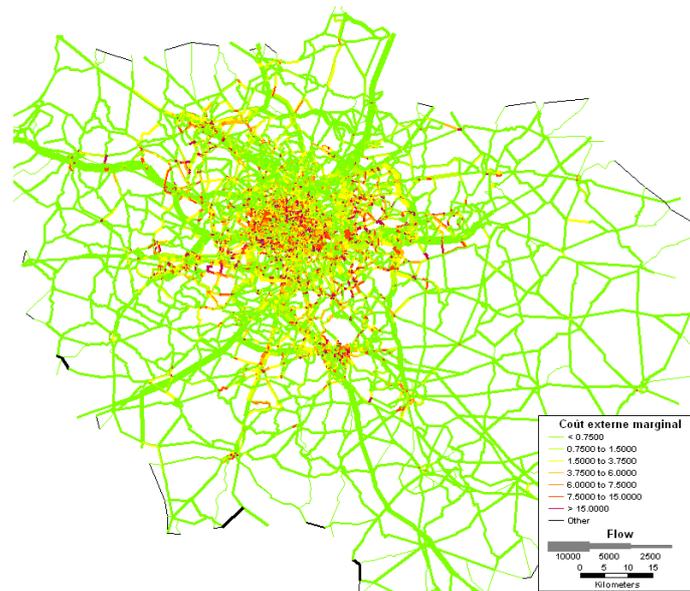


FIGURE A.17.: Le coût externe marginal à l'heure de pointe du soir.

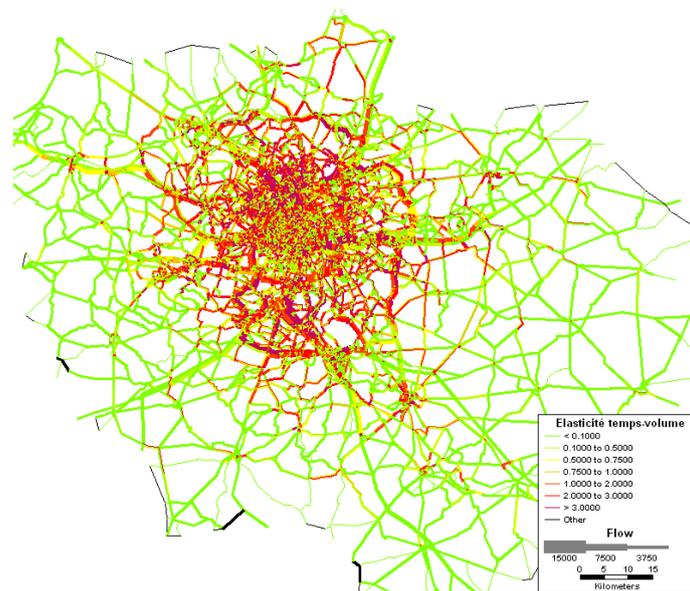


FIGURE A.18.: L'élasticité temps-volume à l'heure de pointe du matin.

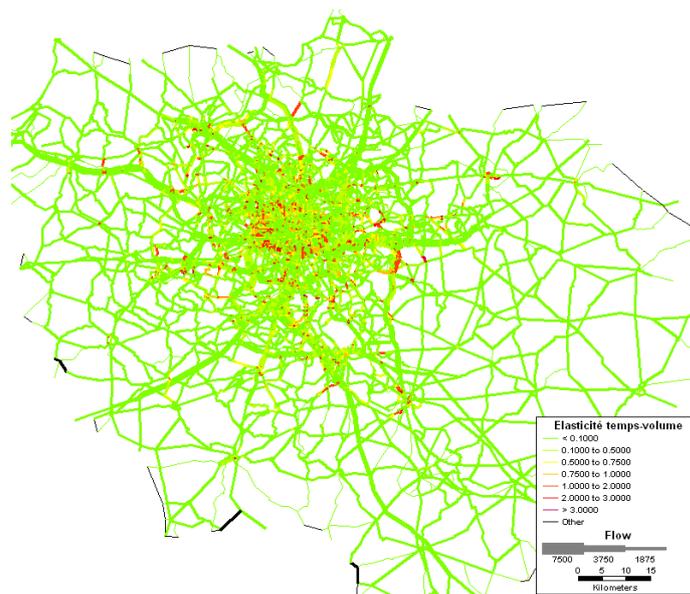


FIGURE A.19.: L'élasticité temps-volume en heure creuse.

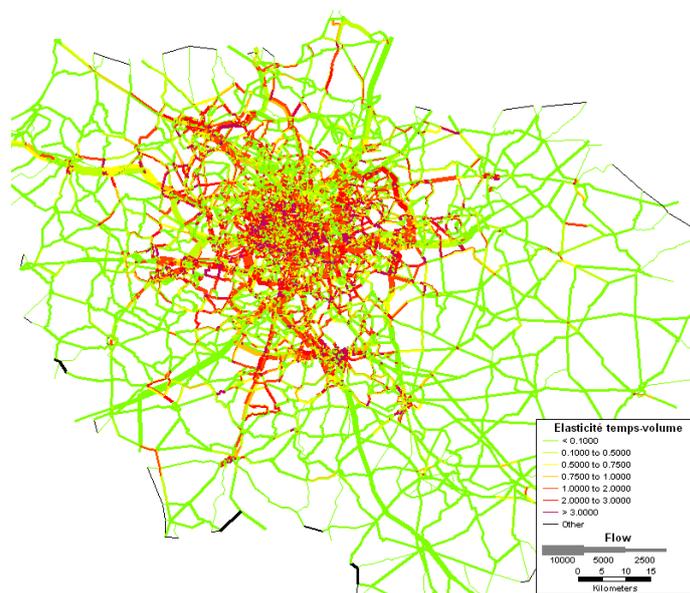


FIGURE A.20.: L'élasticité temps-volume à l'heure de pointe du soir.

A.3. Cartes complémentaires des indicateurs zonaux

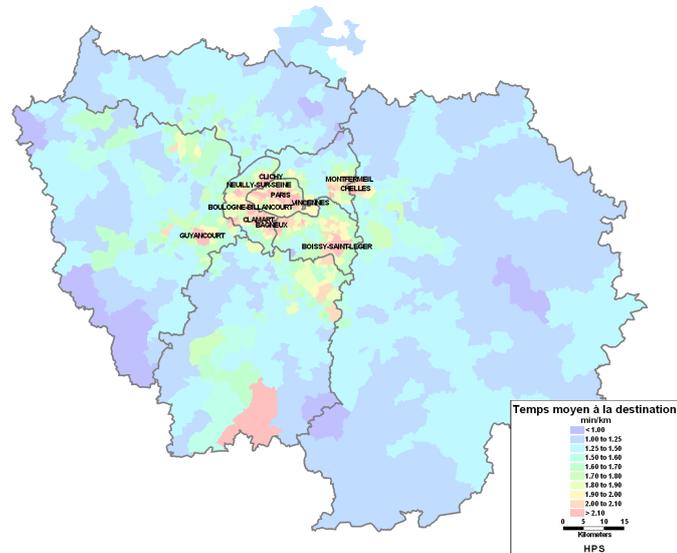


FIGURE A.21.: Temps de parcours moyen par zone de destination des déplacements, à l'HPS

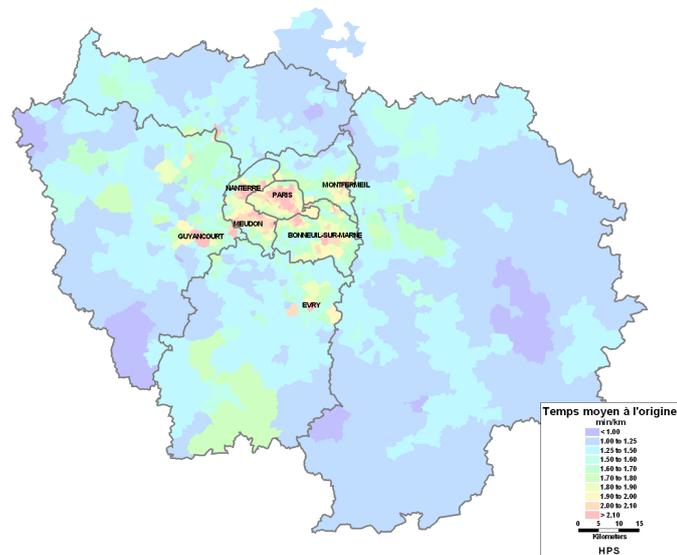


FIGURE A.22.: Temps de parcours moyen par zone d'origine des déplacements, à l'HPS

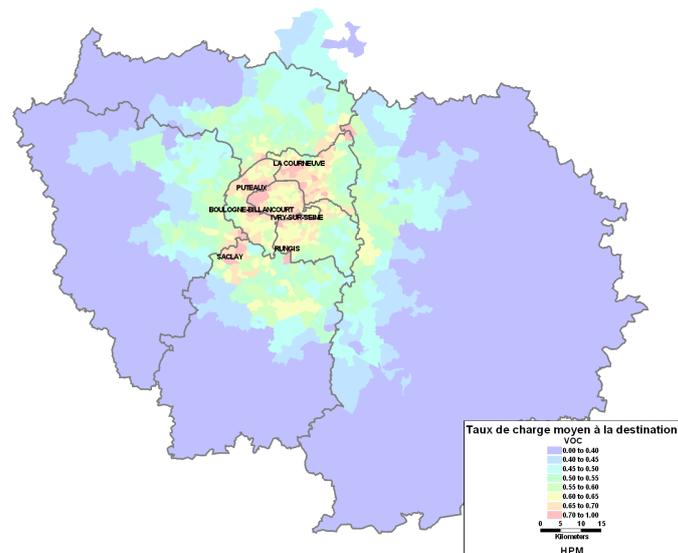


FIGURE A.23.: Taux de charge moyen par zone de destination des déplacements, à l'HPM

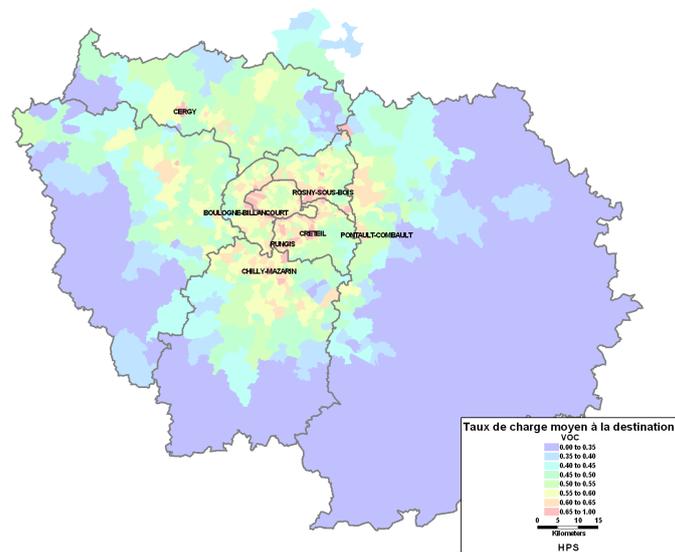


FIGURE A.24.: Taux de charge moyen par zone de destination des déplacements, à l'HPS

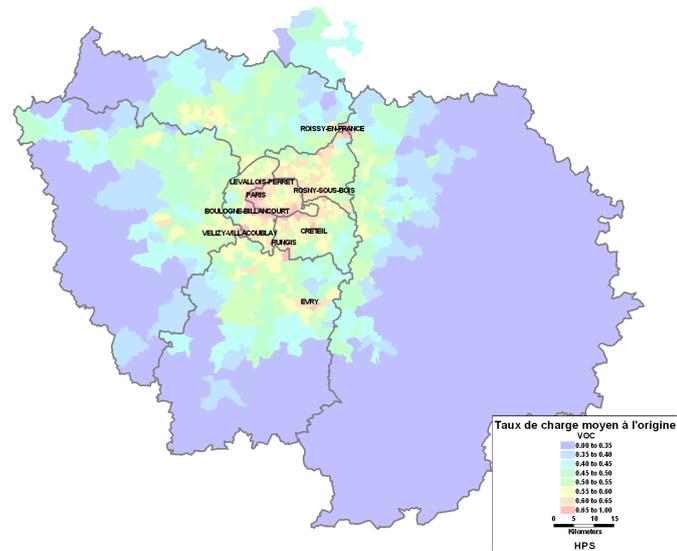


FIGURE A.25.: Taux de charge moyen par zone d'origine des déplacements, à l'HPS

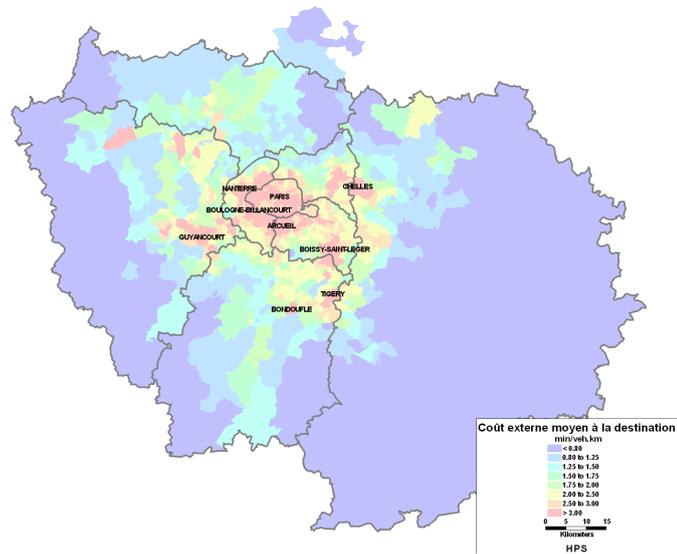


FIGURE A.26.: Coût externe moyen kilométrique par zone de destination des déplacements, à l'HPS

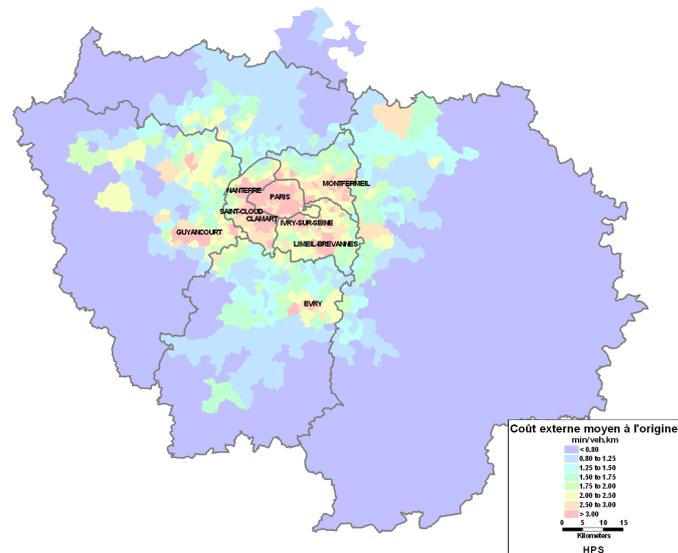


FIGURE A.27.: Coût externe moyen kilométrique par zone d'origine des déplacements, à l'HPS

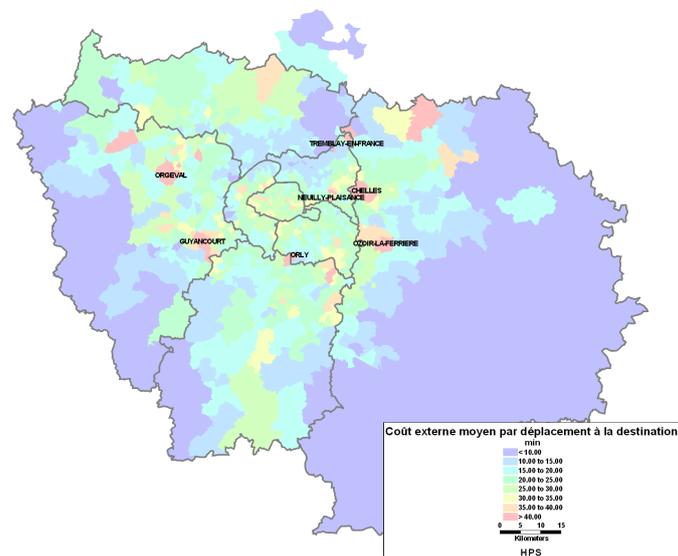


FIGURE A.28.: Coût externe moyen par déplacement par zone de destination des déplacements, à l'HPS

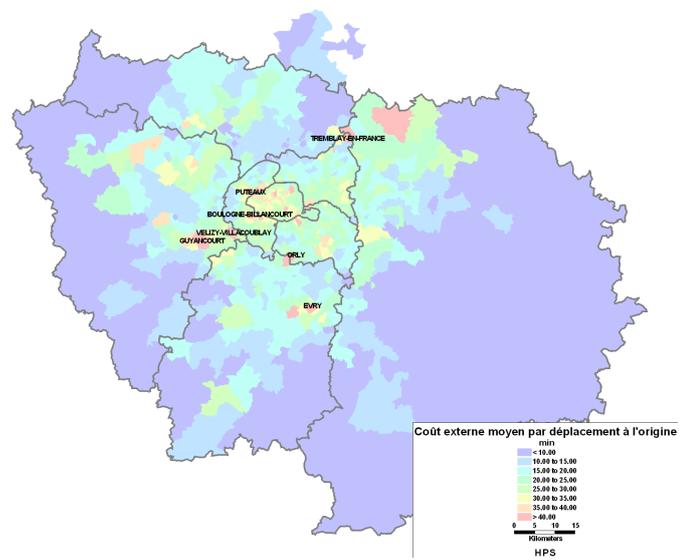


FIGURE A.29.: Coût externe moyen par déplacement par zone d'origine des déplacements, à l'HPS