



N° d'ordre NNT : 2019LYSE2014

THESE de DOCTORAT DE L'UNIVERSITÉ DE LYON

Opérée au sein de

L'UNIVERSITÉ LUMIÈRE LYON 2

École Doctorale : ED 486 Sciences Économique et de Gestion

Discipline : Sciences économiques

Soutenue publiquement le 8 avril 2019, par :

Ouassim Manout

Spatial aggregation issues in traffic assignment models.

Devant le jury composé de :

Aura REGGIANI, Professeure, Université de Bologne, Présidente

Julie LE GALLO, Professeure, AgroSup Dijon, Rapporteur

Thierry BLAYAC, Professeur des universités, Université de Montpellier, Rapporteur

Louafi BOUZOUINA, Chargé de recherche, École Nationale des Travaux Publics de l'État, Examineur

Frédéric DERKX, Directeur de modélisation, ForCity (Financeur CIFRE), Examineur

Patrick BONNEL, Ingénieur de Recherche, École Nationale des Travaux Publics de l'État, Directeur de thèse

Contrat de diffusion

Ce document est diffusé sous le contrat *Creative Commons* « [Paternité – pas de modification](#) » : vous êtes libre de le reproduire, de le distribuer et de le communiquer au public à condition d'en mentionner le nom de l'auteur et de ne pas le modifier, le transformer ni l'adapter.

Université de Lyon

Biais d'agrégation spatiale dans les modèles d'affectation des déplacements

Discipline : Sciences économiques

Ouassim MANOUT

Directeur de thèse : Patrick BONNEL (LAET)

École doctorale des sciences économiques et gestion

Université de Lyon

Avril 2019

Résumé

Les villes sont des systèmes complexes que les modèles urbains peuvent aider à comprendre. Des modèles les plus simplistes aux modèles les plus sophistiqués, la modélisation urbaine a permis de mieux comprendre la question urbaine et ses implications sociétales. Dans ce contexte, les modèles peuvent avoir une valeur-ajoutée appréciable dans le processus de décision publique. Encore faut-il que ces modèles deviennent pratiques et répondent aux contraintes opérationnelles de la chaîne de décision. Dans ce sens, peu de recherches s'est intéressée à la question de praticité des modèles urbains et leur utilisation en situation opérationnelle. À ce jour, les modèles urbains standard qui reposent sur une description agrégée de l'espace sont parmi les approches de modélisation les plus opérationnelles et aussi les plus répandues. De par sa relative praticité, cette approche standard est attractive et simple à mettre en oeuvre. Toutefois, l'agrégation spatiale peut aussi être une source de biais statistiques préjudiciables à la qualité de la modélisation. C'est en particulier, le cas des modèles intégrés Transport-Urbanisme ou des modèles de transport à quatre étapes.

La présente thèse a pour objectif d'étudier la question de l'agrégation spatiale dans les modèles transport et plus particulièrement dans les modèles d'affectation des déplacements. Les modèles d'affectation servent à calculer les temps de parcours et les conditions de déplacement sous congestion, présents et futurs, des personnes et des marchandises. Ils servent aussi à calculer les accessibilités nécessaires aux modèles d'usage des sols dont les modèles de choix de localisation des ménages et des entreprises. Toute erreur ou biais dans l'affectation des déplacements peut compromettre la validité et la qualité globales de la modélisation. Dans ce cadre, une attention particulière doit être allouée au problème d'agrégation spatiale dans les modèles d'affectation. Dans ces modèles, l'agrégation spatiale consiste à regrouper les observations individuelles en utilisant une description agrégée de l'espace, i.e. des zones. Par nature, l'utilisation d'une description agrégée à la place d'une représentation continue engendre une omission de l'information et de sa variabilité et donc un biais statistique dans la modélisation. C'est le cas par exemple avec l'utilisation des connecteurs de zones ou avec l'omission des trafics intrazones dans les modèles d'affectation.

En reposant sur les zones comme unité spatiale de base, les modèles de transport recourent à l'utilisation des connecteurs de zones pour relier les centroïdes de zones au réseau de transport. Les connecteurs sont des liens fictifs qui modélisent les conditions moyennes d'entrée et de sortie du réseau de transport. Pour ce faire, la majorité des modèles de transport reposent sur une méthode simpliste sujette au problème d'agrégation spatiale. La présente thèse examine en détail l'impact de cette description simpliste sur les résultats et la qualité d'un modèle d'affectation des déplacements en transports en commun. Cette thèse propose aussi une nouvelle méthode de modélisation des connecteurs de zones afin de s'affranchir partiellement du biais d'agrégation spatiale dans la modélisation des conditions d'accès au réseau des transports en commun.

L'utilisation des zones comme unité spatiale de base a aussi pour conséquence l'omission des trafics intrazones de l'affectation des déplacements. Les trafics intrazones ont pour origine et pour destination la même zone et de ce fait ne sont pas pris en compte par les modèles standard d'affectation. Cette omission a souvent été ignorée et son impact sur la qualité de la modélisation demeure non évalué. Cette thèse développe une méthode stochastique pour l'évaluation de cet impact. Les résultats fournis par cette étude ont permis de concevoir de nouvelles méthodes

d'affectation qui se veulent plus robustes vis-à-vis de l'omission des trafics intrazones en particulier et du biais d'agrégation en général.

1 Contexte du projet

1.1 Contexte général

L'espace urbain est complexe. Schématiquement, il est le lieu de rencontre de plusieurs acteurs accomplissant différentes activités et consommant une variété de ressources. C'est un espace *géographique d'échange*. Cette notion d'échange est un concept clé dans la vie urbaine. Elle amène les différents agents à entretenir des relations d'ordre social, économique, administratif, etc. D'où leur besoin vital d'interconnexion. Ainsi, à la notion d'espace géographique se superpose la notion des réseaux. Ces derniers peuvent être des réseaux d'énergie, de transport, d'eau, d'information, de communication, de collecte de déchets... De par les flux qu'ils transportent, ces réseaux nourrissent le territoire et participent à son attractivité et à sa compétitivité locale, régionale et internationale. En France, presque 80% de la population et des emplois sont concentrés au niveau des aires urbaines. Ce chiffre était de l'ordre de 74% en 1999 (Insee, 2011). L'aire urbaine de Lyon a, en particulier, vu croître sa population de 40%¹ et doubler de superficie entre ces deux dates (Insee, 2011). Cette croissance reflète, certes, une attractivité de l'espace urbain, mais elle met aussi au jour la complexité progressive de cette organisation. Au fil du temps, elle a abouti à un renforcement et à une complexification des relations entre acteurs et à une multiplication, à la fois, des flux échangés et des coûts et des externalités générées. Pour mieux accompagner cette situation et répondre le plus convenablement possible aux besoins des agents dans un cadre de développement durable, les acteurs publics s'intéressent de plus en plus aux mesures de régulation, de gestion et d'investissement à mettre en place. L'implémentation de ces mesures s'inscrit dans un territoire et une temporalité donnés. Leurs répercussions, quant à elles, peuvent déborder de ces limites spécifiques pour intéresser d'autres espaces à différents horizons temporels. Ce faisant, la portée et les fruits de ces actions ne sont pas complètement maîtrisés et peuvent parfois être décevants, contre-intuitifs ou même contre-productifs. Cette incertitude est plus vraisemblable dans le cas des grands projets. En effet, outre les ressources financières mobilisées, ces derniers ont souvent une large portée spatiale et par conséquent impactent plusieurs agents et systèmes urbains et ont un horizon d'implémentation à moyen ou long terme.

Les systèmes de transport et d'usage du sol font incontestablement partie de cette complexité urbaine (Anthony D. May et al, 2003; Wilson, 1998). L'usage du sol offre des opportunités de localisation -ou impose des contraintes de localisation- aux agents et façonne et organise, ainsi, l'espace en fonction des activités recherchées. À travers la notion d'accessibilité, les transports relient les différents éléments de cette organisation spatiale en lui affectant une cadence temporelle qui en quelque sorte organise aussi l'espace. Aujourd'hui, l'interaction entre ces deux systèmes est évidente (Oryani and Harris, 1997; Stead et al., 2000; Wegener, 2004). Les choix de localisation sont intimement liés aux choix de déplacement, et vice versa. Dans le passé, ces deux systèmes étaient appréhendés séparément ou du moins sans interaction et plusieurs investissements ont été mis en place avec cette logique. Ainsi, plusieurs projets de transport ou d'urbanisme ont été évalués, décidés et implémentés sans prendre en compte la dynamique urbaine qui allait les accompagner (relocalisation des ménages, relocalisation des entreprises, impacts sur le marché foncier, gentrification, densification, étalement urbain, congestion). Cette situation a abouti dans certains cas à des projets contestés comme le cas de l'extension des métros de Lille ou de Toulouse qui, au lieu de maintenir une diversité de la population des quartiers desservis, ont favorisé l'exode des ménages aisés (Ministère du Développement durable, 2014). Dans un autre cas, une évaluation ex-post de plusieurs systèmes urbains de transport ferroviaire aux États-Unis a démontré une tendance

¹ Cette croissance moyenne cache des disparités entre la croissance du pôle urbain (13%) et de la couronne urbaine (158%).

à la gentrification de certains quartiers récemment raccordés aux transports collectifs et à une baisse d'usage de ces derniers. Pour cause, l'amélioration des conditions d'accessibilité a provoqué une augmentation des prix fonciers aboutissant à une fuite des populations pauvres, les plus susceptibles d'utiliser ces transports, et par conséquent à une baisse de leur usage (Pollack et al., 2010). Ces exemples laissent dire que certains mécanismes de rétroaction entre les systèmes de transport et d'urbanisme existent et qu'ils sont déterminants mais souvent négligés dans l'évaluation des politiques d'aménagement.

1.2 Genèse des modèles LUTI

Compte tenu de cette complexité, les décideurs publics sont demandeurs d'outils d'aide à la décision (Commissariat Général à la Stratégie et à la Prospective, 2013). À présent, peu d'outils permettent d'intégrer cette complexité. En France, la majorité des projets de transport sont évalués à travers les sorties des modèles de prévision de trafic à quatre étapes. Ces modèles se proposent de prédire les flux des déplacements de personnes ou de marchandises en suivant un schéma séquentiel [Figure 1]. Dans ce cadre, l'usage du sol est uniquement pris en compte au niveau des données d'entrée exogènes [Figure 1]. Ainsi, de par leur conception, ces modèles ne permettent pas d'intégrer les mécanismes de rétroaction transport-urbanisme, ce qui entache leurs résultats d'incertitude et remet en question, par conséquent, les évaluations qui se basent dessus (Oryani and Harris, 1997). Conscients de cette limite des outils et de la nécessité de coupler transport et urbanisme, les acteurs publics intègrent de plus en plus les projets de transport dans des programmes globaux avec une logique de gestion de portefeuille de projets. C'est le cas par exemple des contrats d'axe en France (Centre d'études sur les réseaux, les transports, l'urbanisme et les constructions publiques, 2010). Pour mieux accompagner ces projets, les contrats d'axe associent autorité organisatrice des transports et collectivité locale pour convenir et optimiser les investissements à entreprendre et assurer une coordination et une performance globales. Encore faut-il que l'on sache mesurer l'impact de ces actions sur une base quantitative et identifier leur combinaison optimale. Dans ce cas, les modèles classiques à schéma séquentiel semblent être inadaptés.

Pour pallier à ce manque d'intégration dans la prise de décision, une nouvelle famille de modèles dite LUTI (Land Use and Transport Interaction) a été conceptualisée et mise en application. Ces modèles intègrent -ou permettent d'intégrer- deux ou plusieurs modules dont un relatif aux transports et un autre relatif à l'usage du sol. Ces deux modules sont imbriqués de sorte que les résultats d'un module alimentent l'entrée de l'autre dans un schéma en boucle. Cette configuration garantit une interaction permanente entre les deux systèmes tout au long des années de simulation [Figure 2]. Grâce à cette logique interactive, ces modèles semblent enfin être capables d'évaluer les projets de transport et leurs externalités diffuses dans le tissu urbain. D'ailleurs, une des recommandations du dernier rapport sur l'évaluation socioéconomique des investissements publics présidé par Émile Quinet est l'adoption des LUTI pour enrichir ces évaluations et pour répondre à certaines questions. Cependant, le rapport souligne la difficulté de la mise en place de tels modèles (Commissariat Général à la Stratégie et à la Prospective, 2013).

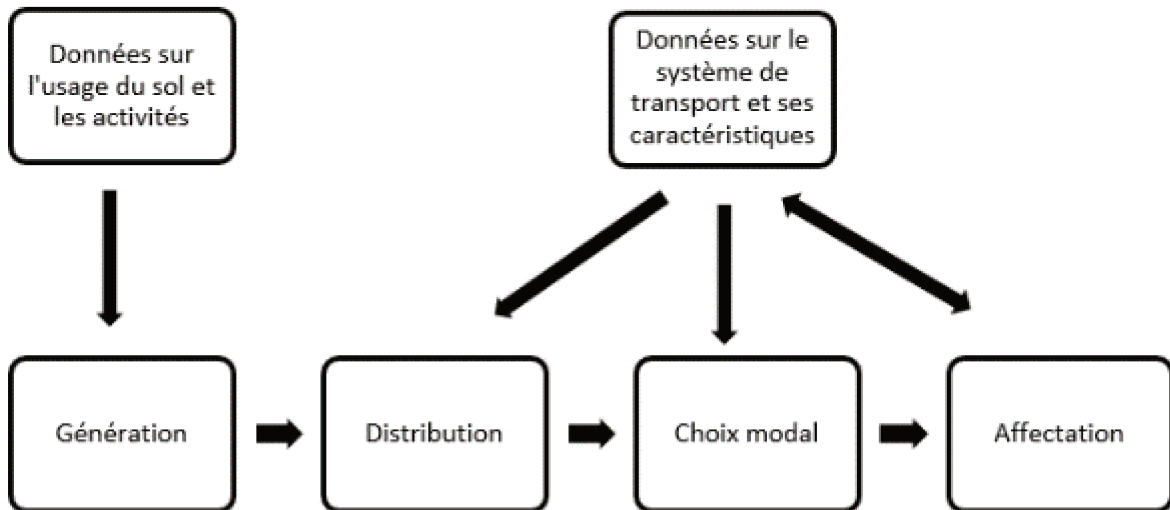


FIGURE 1 : SCHEMA SEQUENTIEL DES MODELES A QUATRE ETAPES (ADAPTE DE LA THESE DE SOPHIE MASSON)

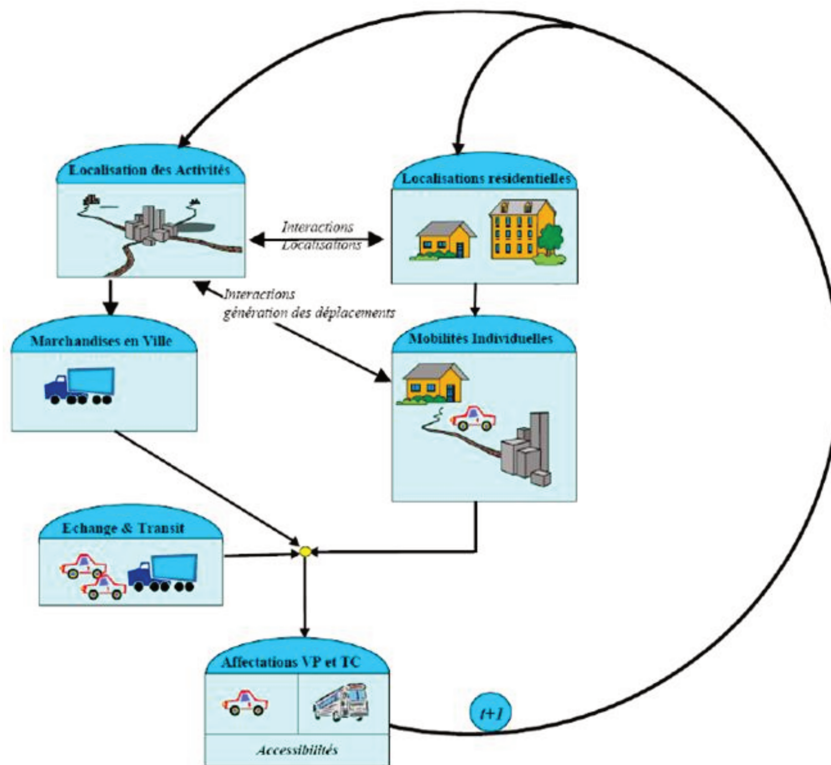


FIGURE 2 : SCHEMA EN BOUCLE ADOPTE PAR LES MODELES LUTI (NICOLAS ET AL., 2009)

C'est dans ce contexte que ForCity a été créée et que SIMBAD a vu le jour. Consciente de la demande des acteurs publics et de l'inadéquation des outils commerciaux actuels, ForCity propose un produit novateur intégrant les attentes des clients.

2 Objectifs et enjeux de la thèse : une problématique commune

Aujourd'hui, de plus en plus d'intérêt est accordé aux modèles LUTI, notamment de la part des chercheurs. Dans ce sens, plusieurs modèles ont été développés. UrbanSim, PECAS, DELTA, MARS, TRANUS et CUBE Land en sont des exemples. Toutefois et à l'inverse des États-Unis et de quelques pays européens où ces modèles ont gagné en maturité, en France, ces derniers sont encore en expérimentation. Ils sont souvent portés par des organismes de recherche. MOBISIM, SIMAURIF, SIMBAD ou encore Pirandello[®] sont autant d'exemples de modèles -ou de plates-formes de modélisation- LUTI nés dans les laboratoires de recherche français². De par leur aspect académique, hormis Pirandello[®], ces modèles sont encore complexes, difficiles à maîtriser et à mettre en place. Ce qui rend, par conséquent, leur implémentation à la fois complexe, lourde et chronophage. De ce fait, les différentes expériences françaises sont restées cantonnées à leur aire d'étude initiale qui est souvent l'aire urbaine d'implantation du laboratoire de recherche en charge de son développement et donc peu transférables spatialement. C'est le cas, aujourd'hui, de la plate-forme SIMBAD.

Cette situation affecte l'essor de la modélisation LUTI en France. En effet, si les acteurs publics conviennent du bien-fondé de l'approche LUTI, ils restent dubitatifs devant leur adoption comme outil d'aide à la décision. Exception faite du projet du Grand Paris dont le maître d'ouvrage a fait appel à trois modèles LUTI : UrbanSim-Metropolis³, Pirandello[®] et Relu-Tran pour évaluer les questions posées par ce projet. Ce recours à plusieurs modèles montre à la fois l'intérêt porté à la modélisation LUTI et l'incertitude des mêmes acteurs vis-à-vis des sorties des modèles et leur capacité à répondre aux questionnements posés. Ainsi, en parallèle aux travaux de développement des modèles LUTI, un grand effort doit être fourni à leur simplification en prenant en compte les attentes des utilisateurs finaux. Notre thèse s'inscrit dans cet effort de simplification des modèles LUTI. Pour ce faire, elle se base sur le cas de la ville de Lyon et de son aire urbaine modélisée par la plate-forme SIMBAD (SIMuler les MoBilités pour une Agglomération Durable). Cette thèse a pour objectif global d'appréhender la complexité de la modélisation LUTI à travers la question des échelles de modélisation et de proposer en conséquence des pistes de simplification à mettre en application dans la plate-forme SIMBAD. À terme, cette simplification profitera à ForCity pour proposer un produit commercial issu du monde de la recherche, ergonomique et orienté client.

Il est de plus important de souligner que les dernières décennies ont vu la naissance et l'obsolescence de plusieurs modèles LUTI. Dans certains cas, ces modèles étaient ingénieux et novateurs mais malheureusement trop complexes pour être appropriés par un public autre que leurs développeurs. Ce constat renforce notre volonté de focaliser notre effort plus sur la simplification que le développement. Un effort qui, à terme, peut être également bénéfique à la diffusion de l'usage des LUTI et leur adoption par un public profane à la modélisation mais intéressé par des questions auxquelles ces modèles essayent d'apporter des réponses.

² Pirandello[®] est un LUTI porté par un acteur privé (VINCI Concessions) en collaboration avec des organismes de recherche.

³ UrbanSim-Metropolis est une plate-forme de modélisation LUTI intégrant UrbanSim et Metropolis.

3 Description des travaux à mener

Pour appréhender et modéliser la complexité urbaine, les modèles LUTI recourent à plusieurs sources de données. Ces données servent au calage des modèles et à la simulation des scénarios. La qualité de ces simulations et leurs résultats sont donc fortement tributaires de la propre qualité des données. L'implémentation de la plupart de ces modèles nécessite des données sur la population, les emplois, l'usage du sol, les déplacements, les marchandises, les réseaux de transport, la conjoncture économique, les tendances démographiques, les prix fonciers, etc. Ces données d'entrée sont pour la plupart spatiales (par opposition aux données non spatiales). Elles proviennent de plusieurs sources disparates et à différentes échelles de résolution. En outre, dans le cadre de ses missions, ForCity travaille avec plusieurs acteurs publics et privés qui n'ont pas nécessairement une même échelle spatiale d'analyse et d'intervention sur l'espace. De ce fait, toute utilisation éventuelle de ces données spatiales nécessite une harmonisation des échelles de résolution par agrégation préalable des données à un niveau spatial commun. Si cette opération ne pose pas beaucoup de difficultés méthodologiques, notamment avec l'essor des outils de sciences de l'information géographique (SIG), elle peut introduire un biais statistique très préjudiciable à la qualité de la modélisation (Miller, 1999; Openshaw, 1977; Viegas et al., 2009). Il s'agit du problème de MAUP (pour Modifiable Areal Unit Problem ou Unités Zonales Modifiables). Le problème de MAUP est inhérent à l'agrégation des données spatiales ponctuelles. Dans le présent sujet de thèse, nous nous proposons d'étudier le problème d'agrégation zonale et ses impacts sur la qualité des ajustements des modèles et leurs sorties et de construire une méthodologie d'agrégation minimisant ces impacts.

Il est évident que le niveau d'agrégation des données spatiales et la finesse de représentation du réseau de transport sont intimement liés (Bonnell, 2004; Bovy and Jansen, 1983; Hensher et al., 2004; Ortúzar S, Juan D and Willumsen, 2011). Le choix d'un niveau de zonage dépend, en principe, du choix du niveau de description du réseau et vice versa. Il serait, par exemple, inutile de coupler un zonage très fin à un réseau grossier ou un zonage grossier à un réseau très fin et illusoire d'analyser finement les résultats de sortie dans les deux cas (Bonnell, 2004). De ce fait, une cohérence entre ces deux échelles (des données spatiales et du réseau de transport) s'impose. Ainsi, le recours à l'agrégation zonale fait appel à l'agrégation du réseau et inversement. En outre, la modélisation des transports peut recourir à une représentation agrégée du réseau pour réduire le coût de codification et/ou d'implémentation. Détailler finement un réseau serait chronophage et parfois inutile. Car, si aujourd'hui nous sommes en mesure de décrire fidèlement le réseau, demain, en situation de prévision, cette précision serait hypothétique (Bonnell, 2004; Ortúzar S, Juan D and Willumsen, 2011). Alors, comme pour l'agrégation zonale, nous allons étudier le problème d'agrégation du réseau et ses impacts sur les sorties du modèle et notamment sur l'estimation des accessibilités qui sont d'une importance majeure dans la modélisation intégrée. Par la suite, nous allons proposer une méthodologie d'agrégation du réseau minimisant ces impacts.

Dans cette vision de simplification par agrégation, les résultats de sortie de la modélisation intégrée seront davantage entachés du biais d'agrégation. Ce biais reprend l'impact de l'agrégation des données spatiales au niveau des zones et de concentrer les différentes informations au niveau des centroïdes de zones. Ainsi, le prix de la simplification par l'agrégation est une perte d'information sur les conditions locales d'accessibilité. Ce biais est davantage prépondérant que le niveau de détail du zonage est agrégé ou grossier. Dans le cadre de cette thèse, nous serons donc mené à traiter ce problème. Pour réduire son extension, nous allons recourir à la notion de connecteurs. Ces derniers relient les zones au réseau modélisé. Ils traduisent, normalement, les conditions locales d'accès et de sortie du réseau modélisé. Toutefois, les méthodologies actuelles de codification des connecteurs ne permettent pas de refléter ces conditions (Bonnell, 2004; Manout, 2014; Ortúzar S, Juan D and

Willumsen, 2011). Il est nécessaire alors de réfléchir à une nouvelle méthode de définition des connecteurs de zones.

Notre thèse est organisée en trois axes :

1. Les effets de l'agrégation zonale : impact sur la qualité des ajustements des modèles et les estimations en sortie de SIMBAD. Vers une méthodologie automatique d'agrégation zonale minimisant ces impacts ;
2. Les effets de l'agrégation du réseau de transport : impact sur la qualité des ajustements des modèles et les estimations en sortie de SIMBAD, et en particulier sur les accessibilités. Vers une méthodologie automatique d'agrégation du réseau minimisant ces impacts ;
3. Réduction du biais d'agrégation par une nouvelle procédure de codification des connecteurs de zones. Cas pratique des transports en commun de l'aire urbaine de Lyon.

Enfin, il va sans dire que ces méthodologies automatiques de changement d'échelles basées sur l'étude des effets de l'agrégation seraient un apport bénéfique à la simplification de la modélisation intégrée. En effet, outre la réduction des temps d'implémentation, les changements des échelles de modélisation apportent une flexibilité à la modélisation intégrée : adaptation au niveau de détail des données disponibles, adaptation aux différentes attentes des différents clients, adaptation aux différents projets, amélioration de la comparabilité, etc. Une flexibilité qui, à terme, pourrait être bénéfique à la transférabilité spatiale du modèle porté par ForCity.

3.1 Développement des axes de la thèse

3.1.1 L'agrégation zonale et ses effets sur la modélisation intégrée

Dans la modélisation intégrée, la majorité des données d'entrée sont spatiales. Leur disponibilité et utilisation sont souvent sujettes à une ou plusieurs agrégations préalables. En effet, la plupart des protocoles de diffusion des données spatiales agrègent ces dernières pour garantir leur anonymat. Par conséquent, les données correspondantes ne sont disponibles qu'à partir d'une certaine échelle d'agrégation (En France, l'IRIS ou la commune, par exemple). Il s'agit là d'une agrégation inhérente aux données et imposée au modélisateur. Ainsi, les différentes sources de données d'entrée sont diffusées à différentes échelles. Leur utilisation dans le cadre de la modélisation intégrée nécessite alors une harmonisation de ces échelles de disponibilité. D'où le besoin d'une deuxième opération d'agrégation sur la base d'un niveau commun de résolution. Dans les deux cas, ces changements d'échelles ne se font pas sans impact sur la modélisation (Openshaw, 1977). Dans le cadre de cette thèse, nous nous intéressons à l'incidence de la procédure d'harmonisation spatiale des niveaux de résolution sur la modélisation LUTI. Cette opération relève d'un choix du modélisateur et est donc souvent arbitraire à l'inverse de l'agrégation imposée par le protocole de diffusion.

Si les effets de l'agrégation zonale sont largement connus et débattus dans les sciences géographiques et l'analyse spatiale, ces derniers sont souvent négligés ou ignorés dans la modélisation intégrée et en particulier dans la modélisation des transports (Miller, 1999; Páez and Scott, 2005). La mise en évidence de l'incidence de l'agrégation des données spatiales ponctuelles remonte aux travaux de Gehlke et Biehl qui ont relevé, dans le modèle de régression linéaire, une dépendance de l'estimation du coefficient de corrélation au niveau d'agrégation des données utilisées (Gehlke and Biehl, 1934). Plus on agrège les données, plus le coefficient de corrélation s'approche de 1. Suite à ce constat, d'autres auteurs se sont intéressés à ce problème et ont démontré l'existence de l'effet d'agrégation dans d'autres modèles statistiques : modèles gravitaires, régression logistique (logit) et autres modèles de choix discrets (Amrhein, 1995; Amrhein and

Flowerdew, 1992; Batty and Sikdar, 1982a, 1982b, 1982c, 1982d; Clark and Avery, 1975; Fotheringham and Wong, 1991; Guo and Bhat, 2004; Openshaw, 1977). Ces différents travaux ont montré que le calage des modèles utilisant des données spatiales est largement tributaire du zonage choisi. Un choix arbitraire du zonage amènerait à un calage arbitraire et donc à des résultats arbitraires. Avec Openshaw, ce sujet a connu un renouveau (Openshaw, 1983, 1978, 1977). Il a décomposé le problème d'agrégation des données dit MAUP (Modifiable Areal Unit Problem) en deux composantes : l'effet d'échelle ('Scale effect') et l'effet de zonage ('aggregation effect'). Le premier reprend l'effet du choix d'un niveau de résolution. Le second traduit l'incidence du choix d'un zonage donné à un niveau de détail fixé. Ces deux effets sont très préjudiciables à la qualité d'estimation des modèles spatiaux. Dans une étude des modèles gravitaires, Openshaw a montré l'instabilité des estimations des paramètres des fonctions de résistance qui, dans certains cas, changeaient même de signe (Openshaw, 1977). Dans ce sens, Fotheringham et Rogerson ont déclaré que les conclusions tirées de l'étude du problème de MAUP sont, partiellement, décourageantes, car elles rendent compte de la non fiabilité de toute analyse spatiale agrégée (Stewart Fotheringham and Rogerson, 1993). Un manque de fiabilité qui peut corrompre la qualité des sorties de la modélisation.

Pour s'abstraire des biais d'agrégation introduits par le zonage, certains auteurs ont travaillé avec une approche totalement désagrégée ou individus-centrés (Chapleau and De Cea, 1983). D'autres auteurs ont tenté de garder l'agrégation zonale et de réduire son impact et plus particulièrement l'impact des connecteurs en introduisant des fonctions probabilistes d'accès au réseau (Leurent et al., 2011). Ces approches restent cependant peu répandues en raison notamment de la complexité des modèles développés. De plus, les enquêtes continuent de collecter les origines et les destinations des déplacements à travers un zonage spatial.

Dans SIMBAD, plusieurs modules utilisent des données spatiales agrégées au niveau des zones (caractéristiques socioéconomiques de la population, données d'emplois, prix fonciers...) et des modèles statistiques spatiaux (régression linéaire, régression logistique, modèles de choix discrets, etc.). Par exemple, la localisation des ménages est estimée par un modèle de régression logistique calibré sur des données zonales. À la lumière des effets démontrés de la MAUP, nous proposons de mesurer l'incidence de ce problème sur la modélisation LUTI, avec comme cas pratique la plate-forme SIMBAD. Nous allons tester l'impact des effets d'échelle et de zonage sur, à la fois, l'estimation des paramètres des modèles, leur significativité statistique et les indicateurs en sortie de la plate-forme. L'étude de ces effets et leur quantification par des indicateurs adéquats, nous amènera à la définition d'une méthodologie automatique d'agrégation ayant comme critère de validation la minimisation d'une fonction-objectif à définir en fonction des données de référence à disposition (données de charges de lignes de transports, données de comptage, données carroyées de la population fiscale, etc.).

Cet axe est organisé comme suit :

1. Mise en évidence de l'incidence de la procédure d'agrégation des données (MAUP) sur la qualité d'ajustement des modèles internes à SIMBAD et en particulier sur le modèle de distribution (modèle gravitaire) et choix modal (modèle de régression logistique). Nous allons étudier la dépendance des estimations des paramètres des modèles et leur significativité statique aux variations du zonage.
2. Évaluation des effets sur les sorties des modèles impactés. Cette évaluation prendra la forme d'une étude de sensibilité. Dans ce sens, nous allons tester plusieurs zonages et relever les variations des sorties de la plate-forme (indicateurs d'accessibilité, indicateurs de charges du réseau viaire et des transports en commun, etc.). Cette partie débouchera sur l'identification des indicateurs pertinents pour mesurer l'impact de la MAUP.

3. Définition d'une fonction-objectif sur la base des indicateurs identifiés dans le 2^{ème} point. La validation de la méthodologie d'agrégation aura comme critère la minimisation (ou la maximisation) de cette fonction-objectif. Ainsi, le zonage résultant de cette méthodologie ne sera pas arbitraire et aura un moindre impact sur la modélisation.

3.1.2 L'agrégation du réseau et ses effets sur la modélisation intégrée

L'accessibilité est un concept central dans la modélisation intégrée Transport-Usage du sol. Elle relie les différents modules. Elle est prise en compte dans la localisation des ménages et des activités et dans leurs interactions. Son calcul est dérivé des estimations de la fonction de temps généralisé. Cette dernière fonction intègre les temps de parcours entre les différentes origines et destinations de l'aire d'étude. Ce temps de parcours est le résultat de la procédure d'affectation sur le réseau de transport. Ainsi, l'estimation de l'accessibilité dépend du réseau de transport et de sa représentation.

Dans la pratique de modélisation de l'offre de transport, la plupart des modèles recourent à une représentation basée sur la théorie des graphes (Bonnell, 2004). Le réseau modélisé est formé de liens et de nœuds représentant, plus ou moins, les routes et les carrefours (Hensher and Button, 2000; Ortúzar S, Juan D and Willumsen, 2011; Sheffi, 1984). Ainsi, en fonction des objectifs du projet et des ressources disponibles (données, budget, compétences), cette représentation peut être proche ou éloignée de la réalité du réseau. Dans les deux cas, des simplifications, marginales ou profondes, sont souvent apportées au réseau modélisé. Cette simplification est souvent nécessaire voire obligatoire. Car, modéliser fidèlement un réseau de transport se ferait à des coûts de codification et d'implémentation élevés pour une amélioration non assurée des résultats de sortie (Bonnell, 2004; Bovy and Jansen, 1983; Chang et al., 2002; Ortúzar S, Juan D and Willumsen, 2011). En outre, cette fidélité de description serait difficile à maintenir au fil des années de simulation par manque de données. Il s'agit là d'un exercice difficile d'arbitrage entre la fidélité de représentation et le coût correspondant. Ainsi, les modèles de transport utilisent souvent une représentation simplifiée du réseau, en supprimant, par exemple, certains tronçons ou en ne gardant que certains axes structurants. Il s'agit de l'agrégation du réseau de transport.

Pour agréger un réseau de transport, deux méthodes existent : l'extraction et l'abstraction (Chan, 1976). L'extraction consiste en la suppression d'un niveau hiérarchique du réseau (le réseau non structurant par exemple). L'abstraction consiste à agréger dans un lien fictif les caractéristiques des liens réels qu'il remplace. Dans les deux cas, il est démontré que les résultats de l'affectation sont sensibles au niveau d'agrégation (Bovy and Jansen, 1983; Chan, 1976; Friesz, 1985). Un réseau modélisé grossièrement, par exemple, pourrait surestimer les taux de charge et les temps de parcours des liens en introduisant de la congestion fictive (Bovy and Jansen, 1983; Jeon et al., 2012). De même, Bovy et jansen ont conclu que l'apport d'une description très fine du réseau serait marginal à partir d'un certain seuil de détail (Bovy and Jansen, 1983). Par conséquent, l'agrégation aléatoire du réseau peut produire des estimations erronées de l'accessibilité et impacter tout le processus de la modélisation intégrée. Pour s'affranchir de ce problème de représentation du réseau, certains chercheurs ont développé des modèles continus du réseau ('Continuum models'), décrivant la densité de trafic à chaque point de l'espace (Dafermos, 1980; Daganzo, 1980; Daniele et al., 2003; Wong, 1998). Toutefois, ces modèles théoriques n'ont pas trouvé d'applications pratiques dans la modélisation des transports. Enfin, il est important de signaler qu'en dehors de la recherche opérationnelle, peu de travaux se sont intéressés à l'agrégation du réseau dans le cadre de la

modélisation des transports et moins dans la modélisation LUTI. Ces mêmes travaux sont souvent de nature empirique et ne débouchent pas sur des méthodes systématiques d'agrégation⁴.

La codification du réseau de transport, dans SIMBAD, est finement détaillée. Ce choix était surtout dicté par une volonté d'automatisation (c'est-à-dire sans recourir à l'agrégation) de l'opération de codification que par une recherche de finesse dans la modélisation (Nicolas et al., 2009). Cette codification repose sur la base de données 'Navteq 1999' qui décrit finement le réseau de l'aire urbaine de Lyon. Cette base comporte presque 110 000 arcs et 45 000 nœuds. Ces arcs sont hiérarchisés en 5 niveaux : des voies rapides et autoroutes (niveau 1) aux rues résidentielles (niveau 5). Toutefois, cette finesse de représentation n'est pas accompagnée d'une finesse de description des attributs du réseau. De ce fait, la précision résultant de cette représentation détaillée serait illusoire (Nicolas et al., 2009). Toutefois, les temps de calcul pour l'affectation sur ce réseau et les efforts à déployer pour le faire évoluer sont proportionnels à son niveau de détail. Dans ce sens, nous nous sommes intéressés à l'agrégation du réseau de transport et à l'automatisation de cette procédure dans la plate-forme intégrée SIMBAD. Ce travail d'automatisation sera éclairé par une analyse préalable de la sensibilité des résultats de l'affectation, d'une part, et de la modélisation intégrée, d'autre part au niveau d'agrégation. Une première approche de simplification serait d'agréger (ou de supprimer) le niveau hiérarchique le plus bas du modèle de réseau. Ce dernier niveau représente les voies locales ou résidentielles et n'est donc pas structurant. Il est attendu que la suppression de ce niveau impacterait les indicateurs de charges et d'accessibilité en introduisant de la congestion fictive due au report de charge du réseau supprimé au reste du réseau (Bovy and Jansen, 1983). De même, il est attendu que la suppression des tronçons modifie la topologie du réseau en provoquant des coupures de réseau. Nous allons quantifier cet impact par des indicateurs adéquats (écarts entre les charges prédites et les données de charges observées, écarts entre les temps de parcours modélisés et observés, etc.). Sur la base de ces indicateurs identifiés, une méthode automatique d'agrégation du réseau est conçue et appliquée à SIMBAD.

Cet axe est traité en trois points :

1. Mise en évidence de l'impact de l'agrégation du réseau sur les résultats de l'affectation de la demande de transport et le reste de la modélisation intégrée (localisation des ménages et des établissements, etc.). Une étude de sensibilité est appliquée au cas pratique du réseau de l'aire urbaine de Lyon. Différentes formes d'agrégation (par extraction et par abstraction) seront testées et évaluées. Cette étape permettra, à terme, d'identifier les indicateurs pertinents pour évaluer l'impact de l'agrégation.
2. Sur la base des indicateurs identifiés, nous allons concevoir une méthodologie d'agrégation du réseau. Cette méthodologie aura comme principe le respect d'un certain nombre d'invariants, c'est-à-dire, des valeurs constantes entre le réseau détaillé et le réseau agrégé. Par exemple, conserver le temps total de parcours et/ou la distance totale parcourue entre les deux représentations du réseau. Viendra, ensuite, l'identification des paramètres d'ajustement du réseau pour faire respecter ce principe de conservation (modification de la capacité des liens, des courbes débit-vitesse, etc.).
3. La validation de cette méthodologie reposera, d'une part, sur le respect du principe de conservation émis au 2^{ème} point, et sur la minimisation d'une fonction-objectif

⁴ *Hormis les travaux de Richard Connors et David Watling (Connors and Watling, 2008) et de Chan (Chan, 1976)*

formée par les indicateurs pertinents à construire sur la base des données simulées et observées.

3.1.3 Réduction du biais d'agrégation par la codification des connecteurs de zones

L'accessibilité joue un rôle central dans les études de projets de transport tant pour la modélisation des comportements que pour l'évaluation socio-économique de ces projets. Son calcul est dérivé des temps de parcours entre les différentes zones de l'aire d'étude. Toutefois, le recours au zonage introduit un biais dans l'estimation des temps de déplacement. Toutes les origines et les destinations sont concentrées au niveau des centroïdes de zones et connectées au réseau modélisé par des connecteurs de centroïde (Chang et al., 2002; Sean Qian and Zhang, 2012). Cette approximation est d'autant plus préjudiciable à la qualité de la modélisation que le zonage est agrégé. Dans le cadre de cette thèse, où nous allons développer une méthodologie d'agrégation zonale, l'intérêt de ce sujet (biais d'agrégation) est considérable.

Aujourd'hui, la disponibilité de données fines sur le plan spatial présente une opportunité pour réduire l'impact de ce biais en prenant en compte l'accessibilité locale. Certaines enquêtes permettent de disposer des origines et des destinations exactes et de connaître, de manière précise, l'accessibilité au réseau ou aux aménités du territoire.

Dans les modèles de transport basés sur un zonage spatial, tous les déplacements empruntant un connecteur de centroïde sont supposés avoir pour origine ou pour destination ce même centroïde et ce quel que soit la taille de cette zone. En réalité, ces déplacements sont spatialement disséminés à l'intérieur de la zone. Par conséquent, les méthodologies actuelles basées sur la position des centroïdes ne permettent pas de traduire les conditions locales de connexion au réseau notamment pour les zones agrégées. Cette approximation dans le calcul du temps de parcours et donc des accessibilités pourrait avoir des incidences non négligeables sur les principales sorties des modèles de transport, et en particulier sur le résultat de l'affectation (Chang et al., 2002; Khatib et al., 2000; Sean Qian and Zhang, 2012).

C'est tout particulièrement le cas des transports collectifs pour lesquels le temps d'accès en marche à pied au réseau représente une part relativement importante du temps de déplacement. De plus, dans la plupart des modèles, ce temps de marche est souvent pénalisé car il est considéré comme étant pénible et mal perçu par les usagers. La codification des connecteurs des transports collectifs représente donc un enjeu important dans la réduction du biais d'agrégation et dans l'estimation des temps généralisés et donc des accessibilités. C'est la raison pour laquelle nous proposons une nouvelle méthode de codification automatique de ces connecteurs qui s'appuie sur la disponibilité de plus en plus fréquente des bases de données spatiales détaillées. Cette méthode a été développée dans le cadre du mémoire du candidat (Manout, 2014). Elle se base sur deux hypothèses :

1. Les résultats de sortie des modèles de prévision de trafic, et en particulier les charges des lignes et la matrice des temps de parcours, ne sont pas indépendants de la méthodologie de définition des connecteurs de zones ;
2. Le recours à une définition basée sur des données fines et localisées permet d'améliorer la qualité d'ajustement de ces modèles.

Cette nouvelle méthodologie a pour idée principale l'implémentation de la définition originelle des connecteurs, à savoir, *représenter les conditions locales d'accès et de sortie du réseau* modélisé. Schématiquement, elle se décline en trois parties :

1. Pour chaque zone, la sélection des lignes accessibles depuis cette zone et les arrêts correspondants ;

2. Pour chaque zone et chaque ligne sélectionnée, le calcul de la distribution des distances de marche à pied entre les différents bâtiments situés dans les zones de chalandise des arrêts et ces derniers ;
3. Pour chaque zone et chaque ligne, le calcul d'une distance moyenne pondérée de la distribution des distances calculée à l'étape 2. Le connecteur de zone a comme longueur cette distance moyenne.

Pour ce faire, nous avons construit par recoupement de quatre bases publiques : recensement de la population, recensement des entreprises (SIRENE[®]), données fiscales localisées (Données carroyées) et données du bâti (BD Topo[®]), une nouvelle base de la population et des emplois localisés aux bâtiments sur l'aire urbaine de Lyon. Cette base sert à calculer une distance précise entre la population et les emplois localisés au niveau de chaque bâtiment et les arrêts de transports collectifs accessibles depuis ces bâtiments. Comme les individus (population ou emploi) localisés à proximité d'un arrêt de transport en commun sont plus susceptibles d'utiliser ce mode. Nous avons retenu une fonction de pondération décroissante en fonction de la distance à l'arrêt. Le calcul résultant de l'agrégation permet ainsi de traduire la distance moyenne d'accès des usagers potentiels des transports collectifs.

Dans le cadre de cette thèse, nous allons automatiser et valider cette nouvelle méthode de codification afin de réduire l'impact de l'agrégation zonale. Un cas d'application est déployé sur l'aire urbaine de Lyon.

4 Références

- Amrhein, C.G., 1995. Searching for the elusive aggregation effect: evidence from statistical simulations. *Environ. Plan. A* 27, 105–119. <https://doi.org/10.1068/a270105>
- Amrhein, C.G., Flowerdew, R., 1992. The effect of data aggregation on a Poisson regression model of Canadian migration. *Environ. Plan. A* 24, 1381–1391. <https://doi.org/10.1068/a241381>
- Batty, M., Sikdar, P.K., 1982a. Spatial aggregation in gravity models. 1. An information-theoretic framework. *Environ. Plan. A* 14, 377–405. <https://doi.org/10.1068/a140377>
- Batty, M., Sikdar, P.K., 1982b. Spatial aggregation in gravity models: 2. One-dimensional population density models. *Environ. Plan. A* 14, 525–553. <https://doi.org/10.1068/a140525>
- Batty, M., Sikdar, P.K., 1982c. Spatial aggregation in gravity models: 3. Two-dimensional trip distribution and location models. *Environ. Plan. A* 14, 629–658. <https://doi.org/10.1068/a140629>
- Batty, M., Sikdar, P.K., 1982d. Spatial aggregation in gravity models: 4. Generalisations and large-scale applications. *Environ. Plan. A* 14, 795–822. <https://doi.org/10.1068/a140795>
- Bonnell, P., 2004. *Prévoir la demande de transport*. Presses de l'École nationale des ponts et chaussées, [Paris].
- Bovy, P.H.L., Jansen, G.R.M., 1983. Network Aggregation Effects upon Equilibrium Assignment Outcomes: An Empirical Investigation. *Transp. Sci.* 17, 240.
- Chan, Y., 1976. A method to simplify network representation in transportation planning. *Transp. Res.* 10, 179–191. [https://doi.org/10.1016/0041-1647\(76\)90073-3](https://doi.org/10.1016/0041-1647(76)90073-3)
- Chang, K., Khatib, Z., Ou, Y., 2002. Effects of zoning structure and network detail on traffic demand modeling. *Environ. Plan. B Plan. Des.* 29, 37–52. <https://doi.org/10.1068/b2742>
- Chapleau, R., De Cea, J., 1983. La perception de l'offre par les usagers du transport en commun sous la perspective d'un modèle d'affectation.
- Clark, W.A.V., Avery, K.L., 1975. The Effects of Data Aggregation in Statistical Analysis. *Geogr. Anal.* 8, 428–438. <https://doi.org/10.1111/j.1538-4632.1976.tb00549.x>
- Connors, R.D., Watling, D.P., 2008. Aggregation of Transport Networks Using Sensitivity Analysis. Association for European Transport.
- Dafermos, S.C., 1980. Continuum modelling of transportation networks. *Transp. Res. Part B Methodol.* 14, 295–301. [https://doi.org/10.1016/0191-2615\(80\)90009-0](https://doi.org/10.1016/0191-2615(80)90009-0)
- Daganzo, C.F., 1980. Network representation, continuum approximations and a solution to the spatial aggregation problem of traffic assignment. *Transp. Res. Part B Methodol.* 14, 229–239. [https://doi.org/10.1016/0191-2615\(80\)90002-8](https://doi.org/10.1016/0191-2615(80)90002-8)
- Daniele, P., Idone, G., Maugeri, A., 2003. The Continuum Model of Transportation Problem, in: *Equilibrium Problems and Variational Models*. Springer US, Boston, MA, pp. 53–60.
- Fotheringham, A.S., Wong, D.W.S., 1991. The modifiable areal unit problem in multivariate statistical analysis. *Environ. Plan. A* 23, 1025–1044. <https://doi.org/10.1068/a231025>

Friesz, T.L., 1985. TRANSPORTATION NETWORK EQUILIBRIUM, DESIGN AND AGGREGATION: KEY DEVELOPMENTS AND RESEARCH OPPORTUNITIES. *The automobile* 7, 296.

Gehlke, C.E., Biehl, K., 1934. Certain Effects of Grouping Upon the Size of the Correlation Coefficient in Census Tract Material. *J. Am. Stat. Assoc.* 29, 169. <https://doi.org/10.2307/2277827>

Guo, J.Y., Bhat, C.R., 2004. Modifiable Areal Units: A Problem or a Matter of Perception in the Context of Residential Location Choice Modeling?

Hensher, D.A., Button, K. (Eds.), 2000. *Handbook of transport modelling*, 1st ed. ed, *Handbooks in transport*. Pergamon, New York.

Hensher, D.A., Button, K.J., Haynes, K.E., Stopher, P.R., 2004. *Handbook of transport geography and spatial systems*. Elsevier Amsterdam.

Jeon, J.-H., Kho, S.-Y., Park, J.J., Kim, D.-K., 2012. Effects of spatial aggregation level on an urban transportation planning model. *KSCE J. Civ. Eng.* 16, 835–844. <https://doi.org/10.1007/s12205-012-1400-4>

Khatib, Z., Chang, K., Ou, Y., 2000. Measuring the Effect of Zoning Structure on Statewide Traffic Demand Estimates. *ITE J.* 75.

Leurent, F., Benezech, V., Samadzad, M., 2011. A stochastic model of trip end disaggregation in traffic assignment to a transportation network. *Procedia - Soc. Behav. Sci.* 20, 485–494. <https://doi.org/10.1016/j.sbspro.2011.08.055>

Manout, O., 2014. *Codification des connecteurs de zones pour les transports en commun*.

Miller, H.J., 1999. Potential Contributions of Spatial Analysis to Geographic Information Systems for Transportation (GIS-T). *Geogr. Anal.* 31, 373–399. <https://doi.org/10.1111/j.1538-4632.1999.tb00991.x>

Nicolas, J.P., Bonnel, P., Cabrera, J., Godinot, C., Homocianu, M., Routhier, J.L., Toilier, F., Zuccarello, P., 2009. *Simuler les MoBilités pour une Agglomération Durable*. Rapport final du projet Simbad, Rapport de recherche du Laboratoire d'Économie des Transports Pour le compte de la DRI (MEEDDAT) et de l'ADEME dans le cadre du groupe.

Openshaw, S., 1983. *Modifiable Areal Unit Problem*. Geo Books, Norwich [Norfolk].

Openshaw, S., 1978. An empirical study of some zone-design criteria. *Environ. Plan. A* 10, 781–794.

Openshaw, S., 1977. Optimal zoning systems for spatial interaction models. *Environ. Plan. A* 9, 169–184. <https://doi.org/10.1068/a090169>

Ortúzar, J. de D., Willumsen, L., 2011. *Modelling transport*, 4th ed. Wiley-Blackwell, Oxford.

Páez, A., Scott, D.M., 2005. Spatial statistics for urban analysis: A review of techniques with examples. *GeoJournal* 61, 53–67. <https://doi.org/10.1007/s10708-005-0877-5>

Sean Qian, Z., Zhang, H.M., 2012. On centroid connectors in static traffic assignment: Their effects on flow patterns and how to optimize their selections. *Transp. Res. Part B Methodol.* 46, 1489–1503. <https://doi.org/10.1016/j.trb.2012.07.006>

Sheffi, Y., 1984. Urban transportation networks: equilibrium analysis with mathematical programming methods. Prentice-Hall, Englewood Cliffs, N.J.

Stewart Fotheringham, A., Rogerson, P.A., 1993. GIS and spatial analytical problems. *Int. J. Geogr. Inf. Syst.* 7, 3–19. <https://doi.org/10.1080/02693799308901936>

Viegas, J.M., Martínez, L.M., Silva, E.A., 2009. Effects of the modifiable areal unit problem on the delineation of traffic analysis zones. *Environ. Plan. B Plan. Des.* 36, 625–643. <https://doi.org/10.1068/b34033>

Wong, S.C., 1998. Multi-commodity traffic assignment by continuum approximation of network flow with variable demand. *Transp. Res. Part B Methodol.* 32, 567–581. [https://doi.org/10.1016/S0191-2615\(98\)00018-6](https://doi.org/10.1016/S0191-2615(98)00018-6)