

Chaire Mobilité

*Chaire de recherche sur l'évaluation et la mise
en œuvre de la durabilité en transport*

Rapport d'activités 2021



Chaire Mobilité

Chaire de recherche sur l'évaluation et la mise en œuvre de la durabilité en transport

Rapport d'activités 2021

Titulaire

Pr Catherine Morency, ing., Ph.D.,
Département des génies civil, géologique et des mines
Polytechnique Montréal

Équipe de recherche

Hubert Verreault, Jean-Simon Bourdeau, Pierre-Léo Bourbonnais, Brigitte Milord, Geneviève Bastien – associés de recherche
Plusieurs étudiant-e-s de doctorat, maîtrise et baccalauréat

Partenaires

Ville de Montréal
Autorité régionale de transport métropolitain
Ministère des transports
Société de transport de Montréal
Exo (Réseau de transport métropolitain)

Août 2022

Le présent rapport témoigne des activités de recherche et développement de la Chaire Mobilité. Les résultats, analyses et constats présentés sont la seule responsabilité de la Chaire et n'engagent pas les partenaires. On ne peut présumer, non plus, que ceux-ci partagent les conclusions qui sont tirées.

Citation préférée: Chaire Mobilité (2022). Rapport d'activités 2021, Polytechnique Montréal, 141 pages.

Table des matières

Table des matières

1	Introduction.....	1
1.1	Mission générale de la Chaire Mobilité.....	1
1.2	Structure organisationnelle.....	2
1.2.1	Partenaires	2
1.2.2	Collaborateurs	2
1.2.3	Comité scientifique.....	3
1.3	Programme de recherche.....	4
1.4	Contenu du rapport.....	5
2	Activités de recherche	6
2.1	Thèse de doctorat complétée en 2021	6
2.2	Maîtrises complétées en 2021	6
2.3	Comprendre les facteurs de dépendance à l'automobile.....	7
2.3.1	Avancement général du projet de recherche doctoral	7
2.3.2	Introduction.....	7
2.3.3	Revue de littérature.....	8
2.3.4	Méthodologie.....	9
2.3.5	Résultats	17
2.3.6	Discussion et conclusion.....	29
2.4	Combinaison de l'enquête Origine-Destination de Montréal et de flux de données passives pour permettre un suivi longitudinal de la mobilité.....	31
2.4.1	Introduction.....	31
2.4.2	Revue de la littérature.....	31
2.4.3	Données et méthodologie.....	33
2.4.4	Résultats	38
2.4.5	Conclusion	43
2.5	Méthode de synthétisation spatialisée de la population et de sa mobilité.....	45
2.5.1	Introduction.....	45
2.5.2	Structure du projet de recherche.....	46
2.5.3	Question de recherche	47
2.5.4	Contributions.....	49
2.5.5	Revue de littérature.....	49
2.5.6	Méthodologie	52

2.5.7	Résultats	57
2.5.8	Conclusions et perspectives	60
2.6	Multimodalité et scénarios post-pandémie	62
2.6.1	Introduction.....	62
2.6.2	Principales questions de recherche.....	62
2.6.3	Approches de modélisation potentielles.....	62
2.6.4	Données.....	63
2.6.5	Résultats prévus	63
2.6.6	Schéma	63
2.7	Indicateurs de vulnérabilité des réseaux de transport en commun	65
2.7.1	Introduction.....	65
2.7.2	État des connaissances.....	66
2.7.3	Contributions attendues.....	68
2.8	Méthodologie d'évaluation de la qualité des accès actifs aux stations de métro	69
2.8.1	Introduction.....	69
2.8.2	Objectifs.....	69
2.8.3	Revue de littérature.....	69
2.8.4	Méthodologie	70
2.8.5	Choix des indicateurs.....	74
2.9	Interactions entre chaînes de déplacements et choix modal	76
2.9.1	Introduction.....	76
2.9.2	État des connaissances	77
2.9.3	Méthodologie	78
2.10	Fardeau du répondant.....	82
2.10.1	Introduction.....	82
2.10.2	Principales questions de recherche.....	82
2.10.3	État des connaissances.....	82
2.10.4	Objectifs de recherche	84
2.11	Impacts de scénarios de télétravail.....	85
2.11.1	Introduction.....	85
2.11.2	État des connaissances.....	85
2.11.3	Méthodologie	86
2.11.4	Résultats	91
2.11.5	Conclusion	95
2.12	Fusion inédite de données sur les véhicules et constats	96

2.12.1	Introduction.....	96
2.12.2	Contexte	96
2.12.3	Méthodologie de fusion de données	96
2.12.4	Constats.....	102
2.12.5	Conclusion	104
2.13	Qualité des informations géographiques et impacts sur les calculs	105
2.13.1	Introduction.....	105
2.13.2	Problématiques et contexte	105
2.13.3	OpenStreetMap.....	106
2.13.4	Alternatives à OpenStreetMap.....	107
2.13.5	Impacts de la qualité des données sur les temps de parcours piéton et l'accès au transport collectif	108
2.13.6	Ressources estimées requises pour valider et compléter les réseaux OpenStreetMap.....	111
2.13.7	Conclusion et perspectives.....	112
2.14	Perspective économique de la transformation du parc de véhicules.....	113
2.14.1	Contexte	113
2.14.2	Publications et résultats associés.....	113
2.14.3	Perspectives de recherche	115
3	Rayonnement	117
3.1	Publications (2021).....	117
3.2	Principales participations à des congrès	117
3.2.1	Congrès annuel du Transportation research board 2020	117
4	Références.....	119

Liste des figures

Figure 1-1. Représentation schématique du programme de recherche de la Chaire Mobilité	5
Figure 2-1 : Carte du territoire d'étude considéré pour l'analyse avec le taux pondéré de possession automobile en véhicule par adulte au sein des ménages à l'échelle des secteurs municipaux de l'enquête OD 2018.....	10
Figure 2-2: Exemple de tampons de distance de marche dans un quartier urbain (haut) et suburbain (bas).....	12
Figure 2-3: Représentation schématique de la procédure de modélisation.....	17
Figure 2-4 : Influence relative des variables d'environnement bâti pour les Mén1adulte selon différentes distances tampons testées.	20
Figure 2-5 : Influence relative des variables d'environnement bâti pour les Mén2+adultes selon différentes distances tampons testées	21
Figure 2-6 : Diagramme de dépendance partielle pour la distance à vol d'oiseau au centre-ville de Montréal.....	22
Figure 2-7 : Diagramme de dépendance partielle pour la distance de marche à la plus proche station de métro.....	22
Figure 2-8: Diagramme de dépendance partielle pour la densité de population.....	23
Figure 2-9 : Diagramme de dépendance partielle pour le nombre d'arrêts de bus.....	24
Figure 2-10: Diagramme de dépendance partielle pour le nombre de véhicules d'autopartage basé station.....	25
Figure 2-11 : Diagramme de dépendance partielle pour le nombre d'épicerie/fruiteries.....	26
Figure 2-12 Procédure de fusion de données entre l'enquête Origine-Destination de Montréal et des données passives	35
Figure 2-13 Décompositions STL (a) additive et (b) multiplicative des embarquements dans le métro	39
Figure 2-14 Décomposition STL multiplicative des comptages (a) de vélos et (b) d'automobiles privées	40
Figure 2-15 Courbes projetée (a) et annualisée (b) du nombre total de déplacements par jour dans le métro	41
Figure 2-16 Projection des parts modales quotidiennes de trois modes de transport entre 2013 et 2018.....	42
Figure 2-17 (a) Parts modales quotidiennes annualisées de trois modes en 2018 et (b) parts modales moyennes sur une semaine d'automne 2018	42
Figure 2-18. Des formulaires du recensement à la population synthétique.....	46
Figure 2-19. (a) Population synthétique de δ en utilisant δ comme strate de référence ; (b) Population recensée de δ	48
Figure 2-20. Limites des RMR de Montréal, Toronto et Vancouver (de gauche à droite)	53
Figure 2-21. Limites des SdR de Montréal, Toronto et Vancouver (de gauche à droite).....	53
Figure 2-22. Limites des ADA de Montréal, Toronto et Vancouver (de gauche à droite).....	53
Figure 2-23. Limites des SR de Montréal, Toronto et Vancouver (de gauche à droite).....	54
Figure 2-24. Limites des AD de Montréal, Toronto et Vancouver (de gauche à droite).....	54
Figure 2-25. Variation de α en fonction de la strate de référence.....	58
Figure 2-26. Variation de β en fonction de la strate de référence.....	58
Figure 2-27. Variation de γ en fonction de la strate de référence	59
Figure 2-28. Variation de $\beta+\gamma$ en fonction de la strate de référence	59
Figure 2-29. Variation de $\beta+\gamma$ au niveau de l'AD en fonction de la configuration utilisée.....	60

Figure 2-30. Vue schématique du projet de recherche de “multimodalité et scénarios post-pandémie”	64
Figure 2-31 Méthodologie générale.....	71
Figure 2-32 Création des bassins d'attraction.....	72
Figure 2-33 Décomposition du métro montréalais.....	73
Figure 2-34 Stations actuellement étudiées dans le cadre du projet	73
Figure 2-35 Étapes de la méthodologie.....	87
Figure 2-36 Exemple de la modification des déplacements et des chaînes de déplacements pour un travailleur assigné au télétravail	90
Figure 2-37 Taux d'utilisation par rapport à la capacité des ponts selon l'heure de la journée...	94
Figure 2-38 Dimensions des véhicules disponibles dans la base de données de la Canadian Vehicle Specification (CVS) (Transports Canada, 2012)	98
Figure 2-39 Mécanisme de fusion de données	99
Figure 2-40 Évolution du taux d'association entre les données de la SAAQ et les données CVS entre 1994 et 2019.....	100
Figure 2-41 Évolution de la similarité des candidats de la méthode des trigrammes entre 1994 et 2019.....	101
Figure 2-42 Distribution de la similarité des candidats de la méthode des trigrammes entre 1994 et 2019.....	102
Figure 2-43 Comparaison des temps de marche pour un échantillon aléatoire de déplacements de transport collectif	108
Figure 2-44 Pourcentage des arrêts de bus dans un rayon de 10 minutes à pied encore accessibles selon le niveau de précision des données.....	109
Figure 2-45 Comparaisons réseau routier seulement vs ajout du réseau piéton et pénalités aux feux. Orange/rouge: moins d'arrêts de bus accessibles à pied en 10 minutes qu'avec la marche sur le réseau routier seulement. Vert: plus d'arrêts de bus accessibles à pied qu'avec la marche sur le réseau routier seulement	110
Figure 2-46 Différence de temps d'accès entre le réseau OpenStreetMap complété et validé et le réseau original non validé (Drummondville).....	110
Figure 2-47 : Illustration d'un effet de revenu	113
Figure 2-48 : Illustration d'un effet de substitution	113
Figure 2-49 : Illustration théorique des options de financement sur le choix d'un véhicule, Source : figure inspirée des travaux de (ACFC 2016)	114
Figure 2-50 : Ventes nettes d'essence et véhicules immatriculés par ménage, Canada, 1987-2019	115

Liste des tableaux

Tableau 2-1 : Statistiques descriptives des variables	13
Tableau 2-2 : Description des hyperparamètres et valeurs testées.....	15
Tableau 2-3 : Résultats des modèles GBM-xgboost de références.....	19
Tableau 2-4 : Sommaire des résultats pour la modélisation des ménages à 1 adulte.....	27
Tableau 2-5 : Sommaire des résultats pour la modélisation des ménages à 2 adultes et plus. ...	28
Tableau 2-6 Valeurs sélectionnées des paramètres de lissage des composantes de saisonnalité pour chaque mode et chaque forme de décomposition	38
Tableau 2-7 Force des composantes de tendance et de saisonnalité (indicateur F) dans la décomposition multiplicative des données de chaque mode.....	40
Tableau 2-8 Taux d'erreur entre le nombre de déplacements projeté et observé pour les jours de semaine d'automne 2018.....	41
Tableau 2-9 Variations des parts modales quotidiennes au cours de l'automne 2018 (jours de semaine uniquement)	43
Tableau 2-10 Impacts potentiels de l'agrégation de la strate de référence sur la synthèse de population	48
Tableau 2-11 Variables de contrôle.....	52
Tableau 2-12 Scénarios.....	55
Tableau 2-13 - Évolution des types de chaînes de 1998 à 2018	79
Tableau 2-14 Évolution de la répartition modale des chaînes de déplacements de 1998 à 2018	80
Tableau 2-15 Évolution des systèmes d'activités de 1998 à 2018	81
Tableau 2-16 Réduction en pourcentage des nombres de déplacements et des distances selon les scénarios.....	92
Tableau 2-17 Coefficient d'élasticité des nombres de déplacements et des distances parcourues en fonction des scénarios et de la région du domicile.....	93
Tableau 2-18 Sources de données utilisées pour la fusion de données	97
Tableau 2-19 Description des variables disponibles dans les données de véhicules immatriculés de la SAAQ (SAAQ, 2022).....	97
Tableau 2-20: Exemple de similarité entre deux séries de trigrammes (Morency, 2021)	99
Tableau 2-21 Évolution de la longueur moyenne, de l'empreinte au sol et de l'empreinte volumétrique des véhicules immatriculés au Québec entre 2000 et 2019	102
Tableau 2-22 Capacité estimée de stationnement sur rue dans l'arrondissement Outremont en fonction de la distance inter-véhiculaire.....	103
Tableau 2-23 Évolution de la consommation moyenne combinée d'essence de tous les véhicules et des 10 véhicules les plus éco énergétiques entre 2005 et 2019.....	104
Tableau 2-24 Statistiques des ressources requises pour différentes tâches de codification dans OpenStreetMap (zone urbaine de densité moyenne).....	111

1 Introduction

1.1 Mission générale de la Chaire Mobilité

La Chaire Mobilité a amorcé son second mandat à la fin de l'année 2016. Cette Chaire a été créée en 2010, dans le contexte où plusieurs autorités de transport se questionnaient sur l'évaluation du niveau de durabilité de leur plan, politique et projet de transport. Le premier mandat de cette Chaire s'est déroulé sur une période de 5 ans, de 2010-2015, et a permis de mettre sur pied un programme de recherche orienté autour de l'évaluation et de la mise en œuvre de la durabilité en transport, principalement par le biais de recherches de maîtrise et doctorat. En effet, la Chaire assure une mission claire de formation de personnel hautement qualifié.

Les objectifs de la Chaire ont d'ailleurs été articulés dès le début de ses activités selon six grands axes :

- **Veille.** Tous les projets de recherche impliquent le suivi de l'état des connaissances et ce, afin de bien identifier les limites des approches disponibles et les opportunités de contributions.
- **Recherche.** La Chaire assure principalement des activités de recherche qui visent à développer de nouvelles connaissances et contribuer à une compréhension plus nette et objectives des enjeux associés à la planification des transports et réseaux, à la modélisation de la mobilité et à l'évaluation du niveau de durabilité. Tout ceci, dans une perspective de contribuer à la prise de décision.
- **Développement.** Plusieurs des contributions sont de NATURE méthodologique ce qui suggère des potentialités de développement importantes. Les recherches visent à démontrer certaines fonctionnalités et éventuellement à proposer des développements d'outils opérationnels, dans le cadre d'autres ententes. Les travaux de la Chaire assurent les développements conceptuels et méthodologiques qui permettent de soutenir la modélisation des comportements et l'évaluation de différents scénarios.
- **Évaluation du niveau de durabilité.** La notion de niveau de durabilité est sous-jacente à l'ensemble des recherches entreprises à la Chaire. La formalisation d'un cadre conceptuel du niveau de durabilité s'inscrit aujourd'hui en filigrane des différentes recherches.
- **Formation.** La Chaire donne une grande place aux étudiants et leur assure un rôle de premier plan dans l'articulation des recherches. En outre, ces recherches se retrouvent rapidement au sein de la formation des futurs ingénieurs civils et alimentent l'articulation des prochaines recherches.
- **Transfert.** Assurer le transfert des connaissances et méthodes est un défi de premier plan et la Chaire assure une présence importante dans plusieurs activités de diffusion, en plus d'inviter la communauté à entendre ses résultats lors du colloque annuel. La mise en ligne des maîtrise, doctorats et d'un rapport annuel témoigne de la préoccupation de diffuser les travaux.
- **Pôle multidisciplinaire.** Le génie des transports de Polytechnique se porte très bien : les professeurs attirent un nombre important et constant d'étudiants provenant de différents horizons. D'ailleurs, c'est bientôt cinq professeurs qui seront actifs en génie civil – transports, ce qui multipliera les opportunités de recherche et développement. Tous seront évidemment invités à collaborer aux activités de la Chaire Mobilité.

1.2 Structure organisationnelle

1.2.1 Partenaires

La Chaire Mobilité a été initiée en 2010, par quatre partenaires institutionnels : la Ville de Montréal, la Société de transport de Montréal, le ministère des Transports du Québec ainsi que l'Agence métropolitaine de transport.

Avec les changements dans la gouvernance métropolitaine des transports, le 2^e mandat de la Chaire Mobilité est soutenu par cinq partenaires : la Ville de Montréal, la Société de transport de Montréal, le ministère des Transports du Québec, l'Agence régionale de transport métropolitain et Exo (réseau de transport métropolitain).

1.2.2 Collaborateurs

La professeure Catherine Morency est titulaire de la Chaire Mobilité et assure la direction scientifique et administrative de ses activités. Elle profite de la collaboration de différents collègues notamment pour la codirection d'étudiant-e-s, à savoir :

- Pr Martin Trépanier, professeure titulaire au département de mathématiques et génie industriel
- Pr Nicolas Saunier, professeur titulaire au département des génies civil, géologique et des mines
- Pr Bruno Agard, professeure titulaire au département de mathématiques et génie industriel
- Pr Owen Waygood, professeur agrégé au département des génies civil, géologique et des mines
- Pre Geneviève Boisjoly, professeure adjointe au département des génies civil, géologique et des mines
- Pr Francesco Ciari, professeur adjoint au département des génies civil, géologique et des mines

La Chaire compte aussi sur la contribution de cinq professionnels de recherche :

- **Hubert Verreault**, M.Sc.A., associé de recherche. Titulaire d'un baccalauréat et d'une maîtrise en génie civil à Polytechnique Montréal, Hubert Verreault est associé de recherche pour la Chaire Mobilité à Polytechnique Montréal depuis 2010. En plus de son travail d'associé de recherche, il agit également en tant que chargé de cours à Polytechnique dans différents cours portant sur le transport au baccalauréat. Il est également membre de l'Ordre des Ingénieurs du Québec. En plus de son expérience en recherche à Polytechnique Montréal, il a aussi travaillé à la défunte Agence Métropolitaine de Transport (AMT) à la réalisation de l'enquête OD 2008 de la région de Montréal. Son mémoire de maîtrise portait sur l'utilisation cohérente des enquêtes Origines-Destination. Ses expertises diversifiées portent principalement sur l'analyse de données en transport, les enquêtes en transport, les statistiques, le traitement de données, les bases de données, la programmation et la modélisation des transports.
- **Pierre-Léo Bourbonnais**, Ph.D., associé de recherche. Après avoir terminé un baccalauréat en génie mécanique à Polytechnique Montréal, Pierre-Léo Bourbonnais a complété ses études en transport (Baccalauréat-maîtrise intégré avec maîtrise en génie civil/transport) avec passage accéléré au doctorat. Ses qualités de chercheur, entre autres, sa capacité de synthèse et sa facilité à réaliser des tâches dans différents domaines

avec rigueur et ponctualité l'ont conduit à s'investir dans cette spécialisation. Le transport est un de ces domaines dont les phases de recherches ont fort à gagner de l'intégration accrue de la collecte, de l'analyse et de la diffusion des données grâce à des outils technologiques dont il maîtrise plusieurs aspects. Au cours des années, Pierre-Léo Bourbonnais a acquis une forte réputation dans la gestion de données de mobilité et la production d'objets de visualisation innovateurs favorisant autant la diffusion auprès des partenaires et du public que la compréhension des enjeux par les analystes et intervenants du milieu.

- **Jean-Simon Bourdeau**, M.Sc.A., associé de recherche. Titulaire d'un baccalauréat et une maîtrise en génie civil de Polytechnique Montréal, Jean-Simon travaille à titre d'associé de recherche pour la Chaire Mobilité de Polytechnique Montréal depuis plus de cinq ans. Ses travaux de maîtrise ont porté sur l'analyse de l'offre et de l'utilisation des stationnements à Montréal. Il a aussi travaillé à titre d'associé de recherche sur plusieurs mandats de recherche et développement pour différents mandataires, notamment la ville de Montréal (politique de stationnement), le MTQ, les villes de Longueuil et Québec, Vélo Québec et la Communauté Métropolitaine de Montréal. Jean-Simon se spécialise dans les calculateurs de trajets, le traitement de flux de données en transport, les enquêtes Origine-Destination et l'analyse spatiale. En plus de ses activités de recherche, Jean-Simon a aussi été chargé de cours et de laboratoires à Polytechnique Montréal dans plusieurs cours de transport : Systèmes de Transport (CIV2710), Technologies informationnelles en génie civil (CIV1120), Gestion de données en transport (CIV8760), Circulation (CIV4740).
- **Brigitte Milord**, M.Sc.Econ., associée de recherche. Titulaire d'une maîtrise en sciences économiques de l'Université de Montréal, Brigitte Milord possède 10 années d'expérience dans les milieux de la consultation et de la recherche académique. Avant de rejoindre la Chaire Mobilité, à titre d'associée de recherche, elle a travaillé pour le Centre de recherche en analyse des organisations (CIRANO), pour les firmes de consultants AECOM et Deloitte ainsi que pour l'organisme à but non lucratif Vivre en Ville. Elle se spécialise dans l'évaluation de politiques publiques dans les secteurs des transports, de l'environnement, de la santé et de la fiscalité municipale.
- **Geneviève Bastien**, M.Sc.A., associée de recherche. Titulaire d'un baccalauréat et d'une maîtrise en génie informatique à Polytechnique Montréal, Geneviève a plus de 15 ans d'expérience en développement logiciel en grande entreprise, comme travailleuse autonome, puis dans le milieu académique. Avant de rejoindre la Chaire Mobilité en 2020, elle était, et est toujours, associée de recherche au laboratoire Dorsal en génie informatique, qui se spécialise dans l'analyse de performance de systèmes informatiques complexes, parallèles et/ou distribués. Elle a eu l'occasion de contribuer au développement de plusieurs projets logiciels libres et open source. Son intérêt pour les questions de mobilité l'ont conduit à la Chaire où elle participe au développement de plateformes logicielles permettant aux étudiants et partenaires de développer et mettre en application divers approches/algorithmes/modèles innovants.
- Deux stagiaires post-doctoraux ont aussi contribué aux activités de la Chaire en 2021 : Nazmul Arefin Khan et Aliasghar (Meghdad) Mehdizadeh Dastjerdi.

1.2.3 Comité scientifique

Dès son lancement, la Chaire a créé un comité scientifique constitué d'experts internationaux sur différents thèmes d'intérêt de la Chaire, d'une part pour assurer son rayonnement international

et l'émergence de collaborations de recherche et d'autre part pour assurer qu'un regard critique et indépendant soit porté sur ses activités. Le comité scientifique est actuellement composé de :

- Prof Kelly Clifton (depuis 2020): Civil and Environmental Engineering, Portland State University (facteurs qui influencent les décisions de transport);
- Dr Patrick Bonnel (depuis 2010): Laboratoire d'Économie des transports de Lyon (Analyse des comportements de mobilité urbaine, modélisation de la demande de transport);
- Pr Kostas Goulias: Laboratoire Geotrans de l'Université de Californie à Santa Barbara (Modélisation de la demande de transport, modèles d'activités, simulations par agents);
- Pr Antonio Paez: École de géographie et des sciences de la terre de l'Université McMaster (analyse spatiale et statistique, comportements de mobilité);
- Pr Matthew Roorda: Département de génie civil de l'Université de Toronto (modélisation du transport des marchandises et passagers).

1.3 Programme de recherche

Le programme de la Chaire s'articule autour de cinq grands axes de recherche, tel qu'illustré sur la Figure 1-1 : 1) la modélisation des comportements individuels de mobilité, 2) la modélisation des facteurs ayant une incidence sur les comportements de mobilité, 3) la modélisation des impacts des comportements de mobilité, 4) l'expérimentation et le développement de méthodes de collecte, traitement et analyse de données et 5) la formulation et l'estimation de scénarios.

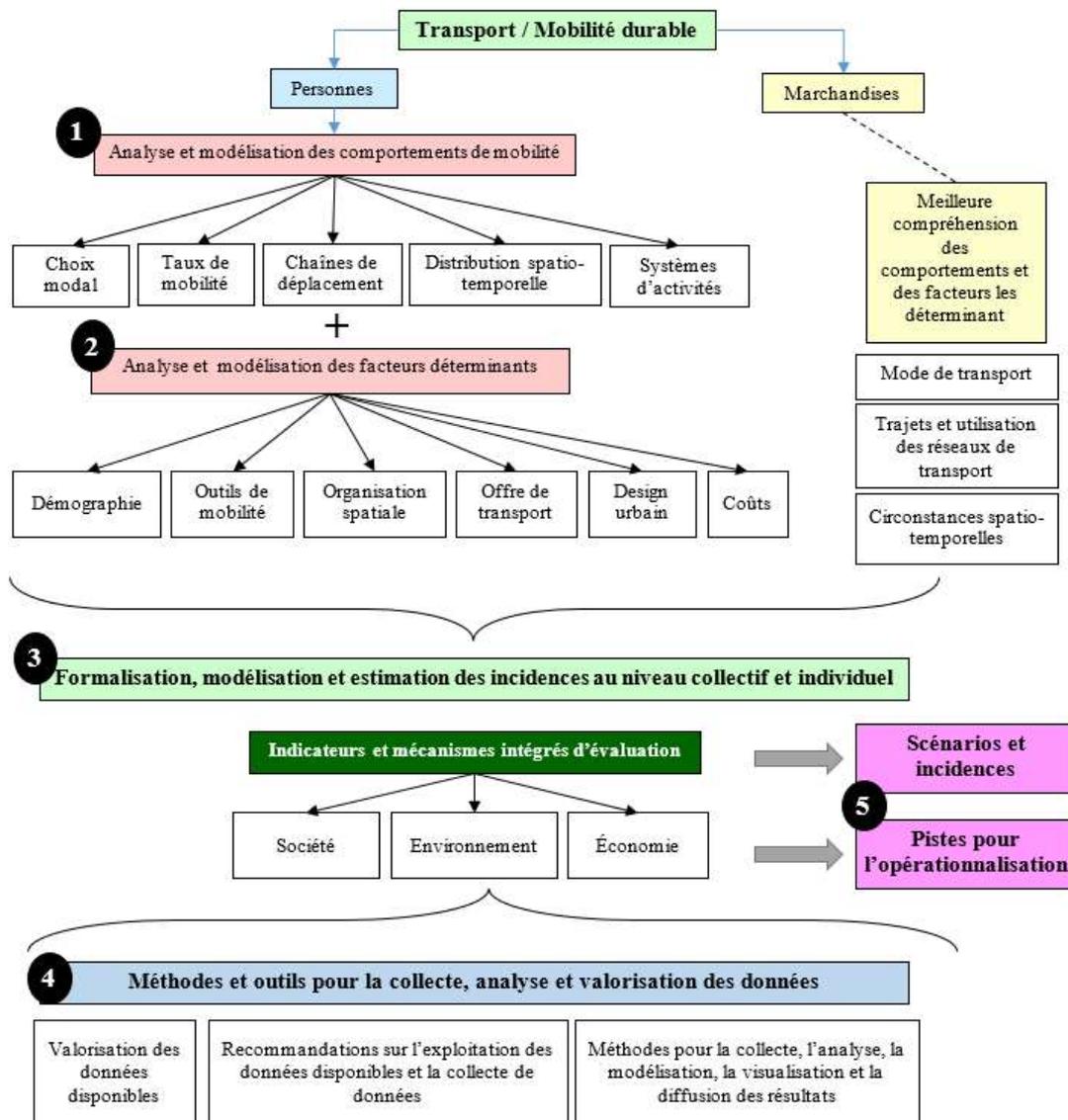


Figure 1-1. Représentation schématique du programme de recherche de la Chaire Mobilité

1.4 Contenu du rapport

Ce document fait état des activités de recherche conduites du 1^{er} janvier au 31 décembre 2021. Les différentes activités de recherche conduites par les étudiants et professionnels de la Chaire sont décrites dans la prochaine section. Différentes activités de rayonnement sont aussi présentées.

Les textes relatifs aux travaux de recherche des étudiantes sont souvent des extraits préliminaires de leurs futurs mémoires et thèses. Les mémoires et thèses complétés ne sont pas incluses dans le rapport mais sont référencés et peuvent être consultés par les partenaires.

2 Activités de recherche

2.1 Thèse de doctorat complétée en 2021

Lefebvre-Ropars, Gabriel (2021). Méthodologie d'évaluation du partage spatiotemporel de la rue. (Thèse de doctorat, Polytechnique Montréal). Directrice : Catherine Morency, Codirectrice : Paula Negron-Poblete.

2.2 Maîtrises complétées en 2021

Morissette, J. (2021). Caractérisation des impacts environnementaux de l'autopartage à Montréal. (Mémoire de maîtrise, Polytechnique Montréal). Directeur : Martin Trépanier, Codirectrice : Catherine Morency.

Narrainen, A. (2021). Développement d'une plateforme de consultation et d'analyse de l'industrie du taxi. (Mémoire de maîtrise, Polytechnique Montréal). Directrice : Catherine Morency.

Poirier, Charlotte (2021). Évaluation de la qualité et de la diversité des alternatives de transport en commun, (Mémoire de maîtrise, Polytechnique Montréal). Directrice : Catherine Morency.

2.3 Comprendre les facteurs de dépendance à l'automobile

- Par : Jérôme Laviolette
- Supervision : Pre Catherine Morency / Pr Owen Waygood / Pr Kostas Goulias (UCSB)
- Contexte : Chaire Mobilité / Chaire en transformation du transport
- Fin prévue : décembre 2022

Cette section est constituée d'extraits de la thèse de doctorat en cours de rédaction de Jérôme Laviolette.

2.3.1 Avancement général du projet de recherche doctoral

Ce projet de recherche doctoral vise à enrichir les connaissances sur les facteurs de dépendance à l'automobile. L'accent est mis sur la possession automobile comme élément central de la dépendance à l'auto. À travers quatre étapes d'analyse et trois échelles de modélisation, le projet cherche à raffiner notre compréhension des facteurs qui influencent les décisions de motorisation des ménages. La présente section rapporte la contribution trois de ce projet doctoral. Une approche d'apprentissage machine appelé Gradient Boosting Decision trees est appliquée aux données de l'enquête OD 2018 de Montréal pour comprendre 1) quelles variables de l'environnement bâti et d'accessibilité ont la plus grande influence sur la motorisation, 2) quelles distances d'accès et tailles de tampons doivent être considérées pour adéquatement capturer l'influence de ces variables et 3) est-ce que des effets seuils existent ? Il s'agit d'un article présenté au Congrès annuel 2022 du *Transportation Research Board* à Washington DC (Laviolette, Morency et Waygood, 2022) et soumis pour publication au *Journal of Transport Geography* en 2021 (Laviolette, Morency et Waygood, 2022, en révision).

2.3.2 Introduction

La possession automobile est liée à divers problèmes environnementaux, sociaux et économiques (Gärling & Steg, 2007) et de nombreuses villes cherchent à réduire le nombre de voitures sur leurs routes en réduisant les besoins en matière de possession de voitures. Il est attendu qu'une réduction de la possession automobile permettrait aussi de réduire l'utilisation de la voiture puisque les ménages ne possédant pas de véhicules font statistiquement moins de trajets en voiture et parcourent moins de véhicules-kilomètres (Sioui, Morency, & Trépanier, 2012; von Behren, Bönisch, Niklas, & Chlond, 2020). L'un des moyens qui a attiré beaucoup d'attention au fil des ans est la transformation des caractéristiques de l'environnement bâti connues comme déterminants clés des comportements de déplacement (Ewing et Cervero, 2010). Malgré la grande variété de modèles proposés pour étudier la relation entre la motorisation et les caractéristiques de l'environnement bâti, l'accessibilité et les comportements de mobilité (Anowar, Eluru et Miranda-Moreno, 2014), des questions restent sans réponse. L'une concerne les seuils de distance nécessaires pour qu'une caractéristique ait un effet et la possibilité que différentes caractéristiques de l'environnement bâti aient des « rayon d'influence » différents. Par exemple, il est courant dans de nombreuses études d'utiliser la même distance ou le même temps d'accès (par exemple, 1 km, 1 mile ou 10 minutes) pour toutes les variables d'environnement bâti et d'accessibilité malgré la logique suggérant que le rayon d'influence d'une station de métro est probablement plus grand que celui d'un arrêt de bus par exemple. En outre, de nombreuses études publiées ont tendance à estimer ces caractéristiques à un niveau agrégé (par exemple, dans le secteur de recensement ou ZAT) au lieu d'utiliser des tampons spécifiques aux ménages.

La présente étude tente de combler cette lacune en examinant d'abord comment la relation entre la possession automobile et les attributs de l'environnement bâti (densité, conception, diversité et mesures d'accessibilité aux ressources de mobilité et aux commodités locales) peut varier lorsque les variables sont calculées à diverses tailles de tampon. Deuxièmement, l'analyse vise à évaluer si les variables considérées présentent des effets seuils sur les décisions de motorisation des ménages.

L'analyse est réalisée en construisant une série de modèles à l'aide d'un algorithme d'apprentissage automatique connu sous le nom de Gradient Boosting Machines (GBM). Plus précisément, les arbres de décision GBM sont un bon moyen d'étudier les relations non linéaires en produisant des diagrammes de dépendance partielle. Ces diagrammes montrent comment la variable prédite varie en fonction d'une variable explicative, en tenant compte de l'effet moyen de toutes les autres variables du modèle.

2.3.3 Revue de littérature

Il a été démontré que la possession d'une voiture est une variable médiatrice entre l'environnement bâti et les comportements de mobilité et a donc été un important sujet de recherche en transport. Bien qu'il existe des revues du cadre de modélisation de la possession d'une voiture (par exemple, Anowar, Eluru, et al., 2014), il n'y a pas de revue publiée des facteurs influençant la possession automobile similaires à celles publiées pour l'impact de l'environnement bâti sur les comportements de mobilité (par exemple, Ewing & Cervero, 2010). Trois grandes catégories de facteurs sont souvent incluses dans la modélisation de la possession automobile : 1) les données sociodémographiques, y compris le revenu, la composition du ménage et les cycles de vie, le statut professionnel, le permis de conduire, 2) les attitudes, y compris les attitudes liées à l'environnement et aux déplacements dans les tentatives de contrôle de l'auto-sélection résidentielle (J. Cao & Cao, 2014; X. Cao, Mokhtarian, & Handy, 2007) et 3) les caractéristiques de l'environnement bâti. Cette troisième catégorie est d'un grand intérêt pour les planificateurs et les décideurs et comprend généralement un ou plusieurs des éléments suivants : densité (résidentielle, emploi), diversité (entropie ou autre indice de diversité d'usage des sols), design (taille moyenne des blocs, nombre d'intersections, kilomètres de rues, d'autoroutes ou de pistes cyclables), disponibilité du transport en commun (proximité/présence d'une gare ferroviaire et d'arrêts d'autobus), destinations diverses par divers modes (voiture, marche, transport en commun) et distance au centre-ville ou aux centres locaux. En de rares occasions, les études incluaient des variables liées au trajet domicile-travail telles que la distance, le temps ou le coût généralisé pour les membres du ménage (par exemple, Bhat et Guo, 2007) ou la disponibilité du stationnement (Guo, 2013 ; Yin, Shao et Wang, 2018).

Malgré le large éventail de variables incluses et de techniques de modélisation utilisées, un aspect important a reçu beaucoup moins d'attention : quelle est le rayon d'influence approprié pour les différentes caractéristiques de l'environnement bâti ? De plus, dans le cas des variables d'accessibilité, est-ce qu'il y a une influence du *nombre* de ces caractéristiques (par exemple, une épicerie/une station de métro à une certaine distance a-t-elle la même influence que 2 ou 3) ? Pour tirer des conclusions fiables sur la direction et l'ampleur de l'influence des caractéristiques de l'environnement bâti, il faut déterminer la bonne échelle et les bonnes méthodes de mesure pour chacune.

La majorité des modèles désagrégés utilisent des zones d'analyse administratives, statistiques ou de transport pour calculer des mesures agrégées. Bien qu'il y ait eu un effort récent pour traiter la dépendance spatiale résultant d'une telle agrégation à l'aide de modèles à plusieurs niveaux

(par exemple, Wang, Shao, Yin et Zhuge, 2018), il est loin d'être clair si ces unités prédéfinies représentent adéquatement les « quartiers ». Plusieurs auteurs ont tenté de résoudre ce problème en calculant le « voisinage » spécifique au ménage à l'aide d'un tampon euclidien ou d'un tampon de distance réseau autour de chaque emplacement de ménage. Cependant, les seuils de distance utilisés pour ces calculs sont des distances couramment utilisées dans les recherches sur les comportements de mobilité (500 m, 1 km, un demi-mile, etc.) et sont très rarement justifiés par référence à des résultats empiriques. Certaines études n'ont fourni aucun détail sur l'échelle d'estimation des variables de l'environnement bâti incluses dans leurs modèles (par exemple, Anowar, Yasmin, Eluru et Miranda-Moreno, 2014 ; Moeckel et Yang, 2016). Seules quelques études ont mentionné avoir testé différentes tailles de tampon ou de zone pour certaines de leurs estimations de variables incluses sans trop expliquer pourquoi une certaine distance a été sélectionnée pour les modèles finaux (par exemple, Anowar, Eluru, & Miranda-Moreno, 2016 ; Bhat, Sen, & Eluru, 2009 ; J. Cao et Cao, 2014 ; Jiang, Gu, Chen, He et Mao, 2017). En résumé, ces exemples montrent l'ampleur du problème auquel est confronté tout chercheur cherchant à modéliser la possession automobile et tout décideur politique qui souhaite soutenir des modes de vie sans voiture ou avec moins de véhicule par ménage. Bien que le problème ne soit pas critique pour la modélisation des caractéristiques de l'environnement bâti considérées comme des substituts pour d'autres choses (par exemple, la densité, la conception du réseau routier), il est particulièrement important lors du calcul de l'accessibilité locale aux commodités ou aux ressources de mobilité. En effet, si différentes tailles de zone tampon ou échelles de voisinage sont prises en compte, conduisent-elles à des résultats différents susceptibles d'affecter les recommandations de politiques publiques ?

Zhang, Zhao, Cao, Lu et Chai (2020) ont spécifiquement étudié la relation non linéaire entre les variables décrivant l'accessibilité régionale, l'accessibilité locale et l'accès aux transports en commun en appliquant des arbres de décision par Gradient Boosting Machine. Leur approche comporte toutefois plusieurs limitations de précision (localisation des ménages approximés au centroïde de quartier) et n'évalue pas l'effet de plusieurs tailles de tampon.

2.3.4 Méthodologie

2.3.4.1 Sources des données

La principale source de données pour cette analyse est l'Enquête Origine-Destination 2018 de la région du Grand Montréal. L'édition 2018 a recueilli des données auprès de 73 421 ménages abritant une population de 168 883 personnes. L'enquête est menée à l'aide d'une méthode d'échantillonnage stratifié et de pondération basée sur le Recensement canadien de 2016 afin d'être représentative de la population de la région enquêtée. L'ensemble de données contient des variables au niveau du ménage, y compris des coordonnées géocodées précises et le nombre de véhicules possédés, des variables au niveau individuel pour tous les membres de chaque ménage enquêté et des informations sur chaque voyage effectué le jour précédant l'entretien d'enquête. À l'aide des coordonnées de chaque ménage, les données de l'enquête OD sont enrichies des variables d'environnement bâti considérées pour l'analyse. La densité de population est calculée à l'aide des données démographiques du Recensement canadien de 2016 en utilisant la plus petite unité statistique disponible (bloc de diffusion). L'ensemble de données d'utilisation du sol nécessaire pour calculer les mesures d'entropie de la diversité des sols a été obtenu auprès de la Communauté métropolitaine de Montréal pour l'année 2016. Les stations de métro et l'emplacement des arrêts d'autobus sont extraits du General Transit Feed Specification (GTFS) des sociétés de transport dans la région. Le jeu de données des stations d'autopartage a été obtenu de Communauto et celui comprenant la localisation des stations de vélos en libre-service BIXI de

la Ville de Montréal. Les opportunités commerciales (épiceries) ont été extraites de la base de données DMTI Spatial pour l'année 2016. Le réseau routier et piétonnier d'OpenStreetMap a été utilisé pour le calcul des tampons d'accessibilité à pied et des caractéristiques de la route (par exemple, la densité des intersections). Tous les jeux de données ont été intégrés dans une base de données géospatiale dans PostgreSQL, en utilisant l'extension spatiale PostGIS.

2.3.4.2 Territoire d'étude

La région pour l'analyse actuelle est l'île de Montréal et les banlieues de Laval et Longueuil. Selon le recensement de 2016, la zone d'étude compte une population totale de 2,78 millions d'habitants répartis dans 1,21 million de ménages, avec une densité de population moyenne de 2700 personnes par km². En 2018, le système de transport en commun comptait trois composantes, un système de métro de quatre lignes et 68 stations, un système de bus de 348 lignes et un système de trains de banlieue de six lignes convergeant vers le centre-ville. La zone d'étude compte 46 517 observations de ménages dans l'enquête OD de 2018 et est illustrée à la Figure 2-1.

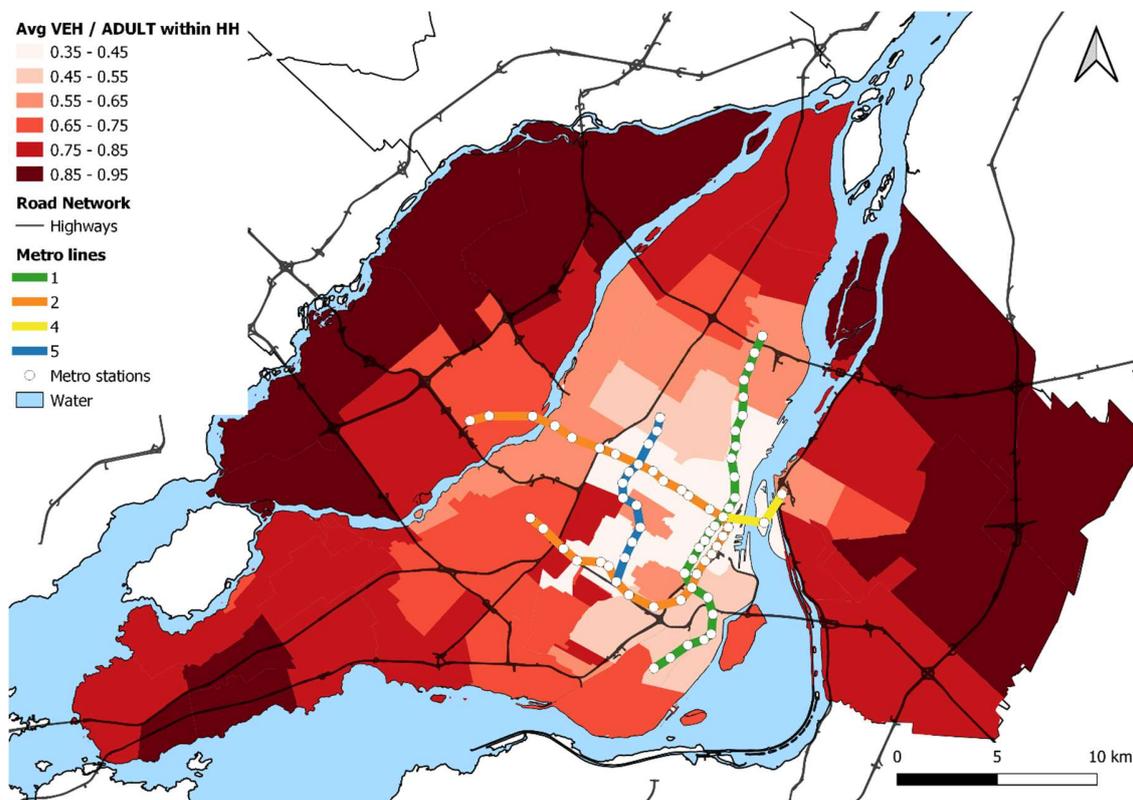


Figure 2-1 : Carte du territoire d'étude considéré pour l'analyse avec le taux pondéré de possession automobile en véhicule par adulte au sein des ménages à l'échelle des secteurs municipaux de l'enquête OD 2018.

2.3.4.3 Préparation des variables

Les variables d'environnement bâti et d'accessibilité considérées ont été sélectionnées sur la base de recherches antérieures. Pour calculer les variables, des tampons de distance de marche sont calculés autour de chaque ménage enquêté en utilisant les coordonnées exactes dérivées de l'adresse géocodée du ménage. Les zones tampons de distance de marche ont été sélectionnées car elles sont censées mieux refléter la véritable accessibilité des ménages dans leur quartier que des tampons à vol d'oiseau. Des tampons de sept distances différentes - 400, 600, 800, 1000,

1200, 1400 et 1600 m - ont été construits à l'aide de l'outil d'analyse de réseau de QGIS. Des exemples des sept tailles de zones tampons sont illustrés à la Figure 2-2 pour un quartier central et suburbain. La configuration du réseau routier dans les quartiers de banlieue entraîne des zones plus petites couvertes par des zones tampons accessibles à pied et une accessibilité plus faible par rapport aux quartiers urbains.

La densité de population a été calculée en additionnant la fraction de la population de l'îlot de diffusion du Recensement canadien de 2016 intersectant le polygone de chaque zone tampon et en divisant par la superficie totale de la zone tampon. Comme mesure de diversité, l'entropie a été calculée au sein de chaque zone tampon selon la méthode décrite par Potoglou et Kanaroglou (2008) en considérant cinq catégories d'utilisation des sols : résidentielle, commerciale, bureaux, loisirs et/ou parcs et institutionnelle. Pour évaluer le design du quartier, la densité d'intersections a été combinée avec le pourcentage d'intersections à quatre directions pour obtenir la densité d'intersections quatre directions dans chaque zone tampon. Cet indicateur semble être un bon discriminant entre les quartiers en grille traditionnels et les quartiers en cul-de-sac de type suburbain.

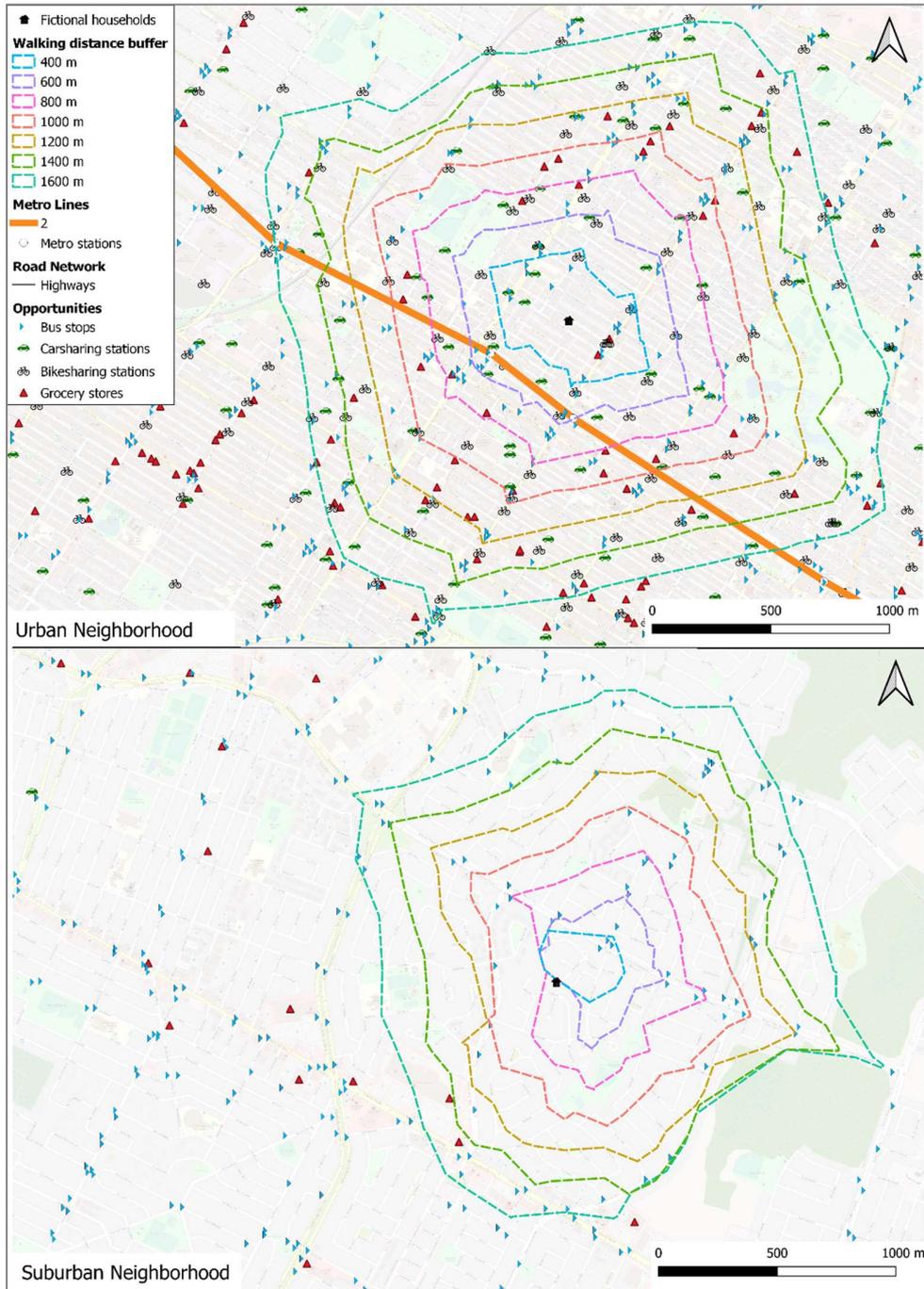


Figure 2-2: Exemple de tampons de distance de marche dans un quartier urbain (haut) et suburbain (bas)

L'accessibilité aux ressources de mobilité a été calculée en additionnant indépendamment chacune des opportunités suivantes au sein de chaque tampon pour les sept distances sélectionnées : nombre d'arrêts de bus, nombre de stations de métro, nombre de véhicules d'autopartage en station, nombre de stations de vélos en libre-service. De même, l'accessibilité aux opportunités locales a été calculée en additionnant le nombre d'épiceries locales et de chaînes. L'accessibilité à d'autres commodités (pharmacies, garderies, écoles primaires) a été

testée mais a été abandonnée par la suite parce qu'elle ne contribuait pas de manière significative aux modèles.

Les variables de contrôle sociodémographiques (SD) ont été sélectionnées sur la base de recherches antérieures sur la possession d'une voiture et extraites directement des données de l'enquête OD. Le revenu du ménage, la présence d'au moins une personne à mobilité réduite dans le ménage, l'âge moyen des adultes dans le ménage, l'amplitude d'âge des adultes du ménage (âge maximum moins minimum), le nombre d'enfants (17 ans et moins), la somme de travailleurs (la somme des travailleurs à temps partiel, des travailleurs à temps plein et des étudiants âgés de 18 ans ou plus) sont inclus. L'amplitude d'âge des adultes du ménage permet de contrôler pour la présence de deux générations dans le ménage (par exemple, deux parents dans la soixantaine et deux enfants au début de la vingtaine ou deux adultes dans la cinquantaine avec un parent âgé de 75 ans et plus).

Les statistiques descriptives pour toutes les variables sont présentées dans le Tableau 2-1 : Statistiques descriptives des variables. Seules les variables de l'environnement bâti calculées dans une zone tampon de marche de 1 km sont incluses pour faciliter la lisibilité.

Tableau 2-1 : Statistiques descriptives des variables

Variable	Description / Catégories	Ménage à 1 adulte		Ménage à 2+ adultes	
		n	%	n	%
Variable dépendance	n observations	16 384		30 123	
Statut de possession automobile	0 véh.	7 171	43.8%	3 199	10.6%
	< 1 véh. / adulte	na	na	16 003	53.1%
	>= 1 véh. / adulte	9 213	56.2%	10 921	36.3%
Caractéristiques sociodémographiques		n	%	n	%
Incapacité dans le ménage	Yes	1 371	8%	1 942	6%
	0	9 097	56%	6 665	22%
	1	7 287	44%	5 516	18%
NB d'actifs (travailleurs et étudiants adultes)	2+	-	-	17 942	60%
	0	15 282	93%	20 950	70%
	1	607	4%	3 697	12%
NB d'enfants (<= 17 ans)	2+	495	3%	5 476	18%
	< 30 000 \$	5 092	31%	2 458	8%
	30 000 à 89999 \$ (REF)	6 483	40%	10 296	34%
Revenu	> 90 000 \$	1 116	7%	10 251	34%
	Ne sait pas / Refus	3 693	23%	7 118	24%
		Moy.	Coeff. Var.	Moy.	Coeff. Var.
	Âges moyen des adultes du ménage	59.7	29%	49.9	31%

Amplitude des âges des adultes	(Max - Min)	-	-	11.8	120%
Caractéristiques d'environnement bâti					
Tampon de 1 km de distance de marche	Distance à vol d'oiseau au CV [km]	8.88	62%	11.00	57%
	Distance au métro [m]	3 977	122%	5 858	103%
	Densité de pop. [pers. / km ²]	7 239	46%	6 148	54%
	Densité d'intersections 4 directions	27.35	52%	22.94	62%
	Entropie 5-catégories [échelle 0-1]	0.58	26%	0.52	30%
	Nb d'arrêts de bus	50.96	44%	43.23	52%
	Présence de station de métro (0/1)	0.33	142%	0.23	182%
	Nb de véhicules d'autopartage	10.03	154%	6.66	198%
	Nb de station de vélopartage	6.06	157%	3.76	203%
NB d'épicerie locales et chaînes	7.27	105%	5.34	130%	

2.3.4.4 Cadre de modélisation

Pour répondre aux questions de recherche et obtenir des informations sur la distance qui compte pour chaque caractéristique de l'environnement bâti, un algorithme d'apprentissage automatique, le Gradient Boosting Machines (GBM), est appliqué à une série de sous-modèles. Le GBM fonctionne en développant un ensemble de petits arbres de décision où, à chaque itération, un nouvel arbre est ajusté aux résidus de l'ensemble actuel et ajouté à l'ensemble tout en laissant les arbres précédents inchangés (Hastie, Tibshirani, & Friedman, 2009, ch. dix).

Il utilise les principes de l'algorithme de descente de gradient où les paramètres sont modifiés de manière itérative dans le sens du gradient de la fonction de perte. Pour les résultats binaires, par exemple si un ménage possède au moins une voiture (1) ou non (0), l'algorithme *xgboost* du package R du même nom utilise la fonction de log de vraisemblance négative. Pour les problèmes de classification multiple, il utilise la fonction de log de vraisemblance généralisée pour les résultats multinomiaux (Chen & He, 2021). Pour réduire les chances de ne trouver qu'un minimum local, la descente de gradient stochastique peut être utilisée en échantillonnant une fraction de l'échantillon d'apprentissage (Friedman, 2002). L'un des inconvénients des modèles GBM est qu'ils sont sujets au surajustement, ce qui peut gonfler l'impact des valeurs aberrantes et donc nécessiter une méthode de validation croisée pour arrêter le processus de recherche avant qu'il ne se produise. Une validation croisée à 5 sous-ensembles est utilisée pour la présente analyse.

Un avantage des arbres de décision GBM est qu'ils nécessitent très peu de prétraitement des variables car ils peuvent gérer facilement différents types de variables explicatives (continues, catégorielles, de comptage) et sont capables de gérer les valeurs manquantes (Hastie et al., 2009). De plus, si l'objectif est la prédiction, une corrélation élevée entre les variables n'est pas un problème car l'algorithme ne choisira pas une variable faible partageant beaucoup de variances avec une variable plus forte lors de la division des variables pendant le processus de croissance de l'arbre. Ceci est utile lors de l'exploration des caractéristiques de l'environnement bâti qui influencent le comportement de déplacement car ces variables ont tendance à être fortement corrélées. Cependant, si l'objectif est de comprendre la relation entre les variables explicatives et le résultat, la colinéarité entre les variables incluses peut être un problème car l'algorithme peut

sélectionner les variables les plus pertinentes d'un point de vue purement statistique, mais celle-ci ne sont pas nécessairement pertinente d'un point de vue théorique.

La modélisation est réalisée à l'aide du package *xgboost* dans le logiciel R développé par Chen et He (2021). Le package offre une implémentation efficace du cadre de renforcement de gradient développé par Friedman (2001).

Les arbres de décision GBM sont très flexibles en raison des nombreux hyperparamètres qui peuvent être ajustés aux besoins spécifiques du problème. Le Tableau 2-2 résume les hyperparamètres utilisés dans les modèles roulés avec l'algorithme *xgboost*. Les valeurs testées pour chaque hyperparamètres sont celles recommandées dans la littérature sur les arbres de décision GBM (Friedman, 2001 ; Hastie et al., 2009).

Tableau 2-2 : Description des hyperparamètres et valeurs testées

Hyperparamètres	Description	Valeurs testées
Nombre d'arbre (ou itérations)	Nombre total d'arbre utilisé dans l'ensemble par l'algorithme de descente de gradient.	Nb maximal utilisé : 3000
Taux d'apprentissage	Permet d'ajuster la vitesse de descente de gradient de la fonction de perte par l'algorithme en ajustement la contribution de chaque arbre par la valeur du taux d'apprentissage.	0,01; 0,05; 0,1
Profondeur des arbres	Nombre maximal de nœud de séparation par arbre. Une plus grande profondeur d'arbre permet d'avantage d'interactions entre les variables.	3, 5 et 7
Nb min. d'observations par nœud	Nb minimal d'observation à avoir à un nœud de séparation pour que la séparation puisse avoir lieu.	5, 10 et 15
Fraction d'échantillonnage (% de ligne)	Proportion des observations utilisée à chaque itération pour construire un nouvel arbre. Une valeur < 1 rend le processus stochastique.	0,5; 0,6; 0,75; 1
Fraction des variables (% de colonne)	Proportion des variables explicatives utilisée à chaque itération pour construire un nouvel arbre	0,8; 0,9; 1

En tout, 324 combinaisons d'hyperparamètres sont testés pour chacun des deux sous-ensembles de données (ménage à 1 adulte et ménage à 2+ adultes). La combinaison d'hyperparamètres permettant d'obtenir la plus petite valeur de la fonction de perte est sélectionnée pour le reste de la procédure.

Enfin, pour interpréter les résultats et éclairer nos questions de recherche, deux méthodes principales d'analyse des résultats du GBM sont produites et comparées à différentes étapes de la démarche générale décrite dans la section suivante.

Lors de l'utilisation d'arbres de décision GBM, il n'est pas possible de tester si les variables explicatives sont statistiquement significatives. Cependant, Friedman (2001) a étendu une méthode existante d'identification de l'importance relative de chaque variable dans une seule décision à un ensemble d'arbres de décision tels que produits par GBM. Suivant la nomenclature utilisée par Friedman, l'influence d'une variable explicative j moyennée sur l'ensemble de M arbres T (T_m) peut être approchée par :

$$I_j^2 = \frac{1}{M} \sum_{m=1}^M \sum_{t=1}^{J-1} i_t^2 1(v_t = j) \quad (1)$$

Où l'influence de la variable j dans chaque arbre est calculée en additionnant les nœuds non terminaux t (c'est-à-dire les divisions) du nœud terminal J de l'arbre T lorsque la variable de division actuelle v_t est la même que la variable j que nous interrogeons. Ceci est multiplié par l'amélioration empirique au carré i_t^2 résultant de cette division. Il est alors courant, comme c'est le cas dans le package R *xgboost*, que l'influence relative de chaque variable dans le modèle soit standardisée pour totaliser 100 %. Notez que l'influence relative ne donne aucune indication sur la *direction* de l'influence de chaque variable sur le résultat. Pour cela, nous nous tournons vers les diagrammes de dépendance partielle.

2.3.4.5 *Diagramme de dépendance partielle*

Les diagrammes de dépendance partielle (DDP) sont un puissant outil de visualisation pour étudier les relations non linéaires entre les variables explicatives et la variable de réponse dans les modèles GBM. Ils montrent le changement moyen de la variable de résultat (ou dans la probabilité de prendre la valeur 1 dans le cas d'une variable binaire) lorsque nous modifions la variable explicative inspectée, en maintenant toutes les autres variables constantes à leur moyenne. Si leur pertinence peut être limitée lorsqu'il existe de fortes interactions entre les variables explicatives, elles peuvent néanmoins fournir des informations utiles sur l'effet des variables les plus importantes du modèle. Ces diagrammes seront utilisés pour analyser la relation entre les attributs de l'environnement bâti et le statut de possession automobile des ménages en tenant compte de différentes tailles de tampon.

2.3.4.6 *Cadre général d'analyse*

L'objectif principal de la recherche est double : 1) étudier quelle taille de tampon (distance) doit être prise en compte lors du calcul des variables d'environnement bâti et d'accessibilité pour la modélisation de la possession d'une voiture et 2) identifier les effets seuil des variables (c'est-à-dire combien de cette variable) sur les décisions de motorisation des ménages et ce, pour différentes distances de tampon considérées. Une procédure générale est proposée pour examiner ces questions de recherche en utilisant une série d'arbres de décision GBM.

Pour comprendre l'effet structurel des caractéristiques d'environnement bâti étudiées sur les décisions de possession automobile, le nombre de voitures est normalisé à la taille du ménage en divisant le nombre de voitures du ménage par le nombre d'adultes (18 ans ou plus).

La variable de résultat est ensuite obtenue en classant chaque ménage dans trois statuts de possession de véhicule distincts : pas de véhicule, un ou plusieurs véhicules partagés par deux adultes ou plus (c'est-à-dire moins d'une voiture / adulte) et une motorisation complète (c'est-à-dire un ou plus véhicule / adulte). Cette approche exclut de l'équation tout véhicule « excédentaire » peu susceptible d'être expliqué par les caractéristiques de voisinage (une vieille voiture, une décapotable, etc.). Toutefois, en formulant la variable dépendance de cette façon, les ménages avec un seul adulte ne peuvent pas être classés dans la catégorie de motorisation partielle (moins d'une voiture / adulte) et doivent donc être modélisés séparément. Pour cette raison, l'échantillon de l'enquête OD est divisé entre les ménages d'un adulte (Mén1adulte) et les ménages de plusieurs adultes (Mén2+adultes). Pour les ménages d'un adulte, la variable de résultat prend une forme binaire (0 : pas de véhicule, 1 : 1 véhicule ou plus). Cette séparation des échantillons nous permet cependant d'étudier et de comparer les effets de l'environnement bâti sur les ménages d'un seul adulte et les ménages de plusieurs adultes. Les modèles GBM sont construits à partir d'un ensemble d'apprentissage contenant 70 % des observations pour chaque

sous-échantillon. Les 30 % restants de chaque sous-échantillon sont utilisés pour calculer la précision de la prédiction. Un processus en 4 étapes est établi et résumé dans la Figure 2-3.

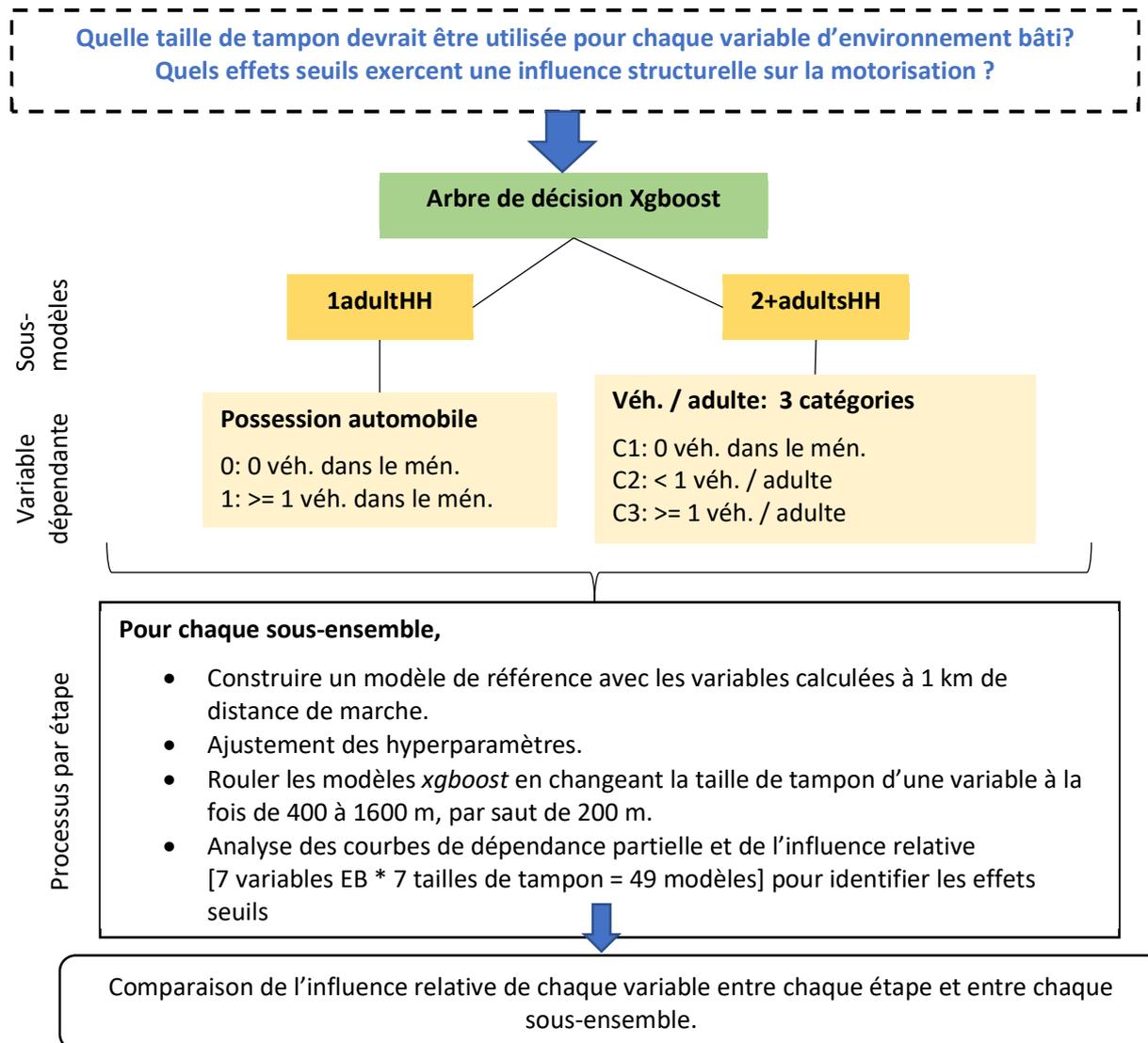


Figure 2-3: Représentation schématique de la procédure de modélisation

2.3.5 Résultats

Dans la première étape du cadre général, un modèle GBM de référence est exécuté avec les variables sociodémographiques et d'environnement bâti présentées au Tableau 2-3 pour les deux sous-échantillons. Dans ces modèles de référence, la taille du tampon pour la mesure des caractéristiques d'environnement bâti est fixée à une distance de marche de 1 km. Les statistiques de performance des modèles et l'influence relative (IR) de chaque variable explicative sont présentées dans le Tableau 2-3 avec les hyperparamètres finaux utilisés.

La première observation est que le modèle de *Mén2+adultes* atteint une précision de prédiction plus faible (64,6 %) que le modèle de *Mén1adulte* (70,7 %) qui a une classe de résultats en moins. La variable sociodémographique dominante pour les *Mén1adulte* est la variable pour les ménages

à faible revenu (< 30 000 \$/an). Pour les *Mén2+adultes*, les variables d'âge (âge moyen et amplitude d'âge des adultes du ménage) se classent deuxième et quatrième en IF, pour une influence combinée de 20 %. Comme les indicateurs catégoriels ne sont pas autorisés dans l'algorithme *xgboost*, le revenu a été encodé sous la forme de quatre variables binaires avec la catégorie la plus représentée (30 000 à 90 000 \$/an) comme référence. L'influence relative totale du revenu est de 26,3 % pour les *Mén1adulte* et deux fois moindre pour les *Mén2+adultes* (11,7 %). Pour les *Mén1adulte*, l'IR du revenu du ménage est comparable aux 31 % obtenus par Zhang et al. (2020) pour leur modèle comprenant toutes les tailles de ménage. Une personne dans le ménage ayant une mobilité réduite est, sans surprise, plus importante pour expliquer le statut de motorisation pour les *Mén1adulte*, que pour les *Mén2+adultes*. Le nombre d'adultes actifs dans le ménage - la somme des travailleurs à temps partiel et à temps plein et des étudiants (âgés de 18 ans ou plus) - a une très faible influence pour les *Mén1adulte* (1,5%), se classant 15^e sur 16 variables. La variable joue un rôle plus important dans l'explication du statut de motorisation des *Mén2+adultes* (6,6 %, 5^e). Fait intéressant, dans les deux modèles, le nombre d'enfants ne joue pas un rôle important, se classant dernier et avant-dernier avec un IR d'environ 2 %.

Malgré les disparités dans le classement des variables sociodémographiques entre les sous-échantillons, l'influence relative combinée des variables sociodémographiques est très similaire (44,8 % pour *Mén1adulte* et 43,2 % pour *Mén2+adultes*) et offre moins de pouvoir prédictif que l'influence des variables d'environnement bâti combinées (55,25 % et 56,8 %). Le ratio d'influence des variables sociodémographiques et d'environnement bâti est l'inverse de ce qui a été trouvé par Zhang et al. (2020) en Chine où l'IR combiné était de 62,4 % pour les variables sociodémographiques et de 37,6 % pour les variables d'environnement bâti. Cela pourrait potentiellement s'expliquer par les variables sélectionnées, la méthode de calcul du tampon (à partir des centroïdes de voisinage dans leur cas) et par la nature de la variable de résultat modélisée.

La densité de population, qui est généralement considérée comme une variable proxy pour diverses caractéristiques de l'environnement bâti, se classe première parmi cette catégorie de variables et respectivement deuxième et première pour les *Mén1adulte* et les *Mén2+adultes* globalement avec des influences relatives très similaires (16,9 % et 16,2 %). Les deux autres caractéristiques de l'environnement bâti « traditionnelles » mesurées à l'intérieur du tampon de marche d'1 km, à savoir l'indice d'entropie et la densité des intersections à quatre voies, ont une influence modérée sur les deux modèles, confirmant que la diversité et la conception sont associées aux décisions de possession d'une voiture par les ménages.

Quatre variables incluses servent à mesurer l'influence de l'accessibilité aux ressources de mobilité. Dans une première itération des modèles (résultats disponibles sur demande), l'accessibilité au réseau de métro a été mesurée avec une variable binaire indiquant si au moins une station de métro était située à moins d'un km de marche de chaque ménage. Sous cette forme, la variable a donné de mauvais résultats, avec une influence relative de moins de 1 % pour les deux sous-échantillons de ménages. C'était contre-intuitif puisque le réseau du métro est l'épine dorsale du système de transport en commun de Montréal. Dans l'itération finale des modèles (résultats présentés dans le Tableau 2-3), l'accessibilité au réseau de métro a été mesurée comme la distance de marche la plus courte jusqu'à la station de métro la plus proche. Sous cette forme, la variable est la troisième plus influente dans les deux modèles, avec un IR de 10,8 % pour les *Mén1adulte* et de 10,2 % pour les *Mén2+adultes*.

Tableau 2-3 : Résultats des modèles GBM-xgboost de références

Sous-ensemble de ménages		Mén1adulte		Mén2+adultes		
	Variable prédite	Véh. dans le ménage (0 vs 1+)		Véh. / adulte (3 catégories)		
	Nb de catégorie	2		3		
	Nb d'observations (n)	16384		30123		
Hyperparamètres	Nb max. d'arbres (itérations)	3000		3000		
	Fonction de perte	Logloss		Multinomial logloss		
	Profondeur d'interactions	3		5		
	Taux d'apprentissage	0.01		0.01		
	Nb minimum d'observations aux noeuds	10		10		
	# NB de sous-ensemble de validation	5		5		
	Fraction d'échantillonnage	0.75		0.75		
	Fraction des colonnes sélectionnés par arbre	0.8		0.8		
Modèles de références (toutes les variables calculées à 1 km de marche)						
Résultats	Meilleure itération	773		985		
	Valeur <i>logloss</i> sur ensemble de test (30 %)	0.557		0.752		
	Prédiction sur l'ensemble de validation	70.8 %		64.7 %		
		Rang	IR (%)	Rang	IR (%)	
Variables sociodémographiques						
	Revenu: < 30 k \$	1	18.5%	10	4.9%	
	Revenu: >= 90 k \$	8	4.0%	7	6.3%	
	Revenu: inconnu / refus	9	3.8%	17	0.5%	
	Mobilité réduite dans le mén. (bin.)	4	8.3%	15	2.6%	
	Âge moyen des adultes du ménage	5	6.8%	4	9.8%	
	Amplitude des âges des adultes du ménage	-	-	2	10.3%	
	NB d'actifs (travailleurs, étudiants > 18 ans)	16	1.5%	5	6.6%	
	NB d'enfants (< 17 ans)	15	1.8%	16	2.2%	
Influence totale - Sociodémographie			44.75%	43.19%		
Environnement bâti (EB)						
	Densité de population [pers./km ²]	1-km marche	2	16.9%	1	16.2%
	Entropie 5-catégories	1-km marche	12	3.0%	8	5.2%
	Densité d'intersections 4-dir.	1-km marche	13	2.9%	12	3.7%
	Distance vol d'oiseau au CV [km]	-	10	3.2%	9	5.1%
	Distance min. de marche au métro	-	3	10.8%	3	10.2%
	NB d'arrêts de bus	1-km marche	14	2.8%	13	3.1%
	NB de véhicules d'autopartage	1-km marche	6	6.6%	6	6.4%
	NB de stations de vélopartage	1-km marche	11	3.1%	14	3.0%
	Nb d'épiceries	1-km marche	7	6.0%	11	3.9%
Influence totale – Environnement bâti			55.25%	56.81%		

Notes: IR= Influence relative

L'accessibilité aux véhicules d'autopartage est la 6^e variable la plus importante dans les deux modèles avec un IR d'environ 6,5 %, ce qui indique que la disponibilité de ce service de mobilité

pourrait affecter les décisions de possession de voiture des ménages à proximité. L'accessibilité aux arrêts de bus et aux stations de vélos en libre-service exerce une influence moindre dans les deux modèles, avec des IR d'environ 3 %. Enfin, l'accessibilité aux épiceries locales semble jouer un rôle plus important dans les probabilités d'être motorisés pour les *Mén1adulte* (IR = 6 %) que pour les ménages de plusieurs adultes (IR = 3,9 %). Ce résultat est intuitif puisqu'il est probablement plus facile pour un ménage d'un seul adulte de faire ses courses à pied que pour les ménages plus grands puisque la quantité de nourriture à transporter est moindre.

2.3.5.1 Résultats des analyses de tailles de tampon

Dans l'étape suivante du cadre de modélisation, 49 modèles GBM pour chaque sous-échantillon de ménages ont été calculés, chacun faisant varier la taille du tampon de l'une des sept variables d'environnement bâti, en maintenant la distance de toutes les autres variables à une zone tampon de 1 km. Les Figure et Figure illustrent comment l'influence relative de chaque variable dans chaque modèle varie en fonction de la taille des tampons de distance de marche utilisés pour le calcul. Plus précisément, pour les deux sous-échantillons, l'influence relative du nombre de véhicules d'autopartage et du nombre de stations de vélos en libre-service varie considérablement entre la plus petite zone tampon (400 m) et la plus grande zone tampon (1 600 m). Le changement est le plus extrême pour l'autopartage où il passe de l'IR la plus basse (0,5 %) à 400 m au 5e rang (8,9 %) à 1600 m pour les *Mén1adulte*, soit une multiplication par 16,8. Pour les *Mén2+adultes*, la variable passe du 16^e rang (1,1 %) à 400 m au 5^e rang (7,6 %) à 1600 m, soit une multiplication par 7. On peut affirmer que puisque ces deux variables sont exprimées en termes absolus (nombre d'opportunités) et non en termes relatifs (densité), les tampons plus grands présenteraient une plus grande variabilité que les plus petits. Cependant, d'autres variables mesurées en termes absolus telles que le nombre d'arrêts d'autobus et le nombre d'épiceries ne montrent pas de changements aussi drastiques. Les arrêts de bus sont même la deuxième variable la plus cohérente sur la taille du tampon après l'entropie avec une IR moyenne de 2,8 % (*Mén1adulte*) et 3,11 % (*Mén2+adultes*) et un écart type (entre les tailles de tampon) de 0,71 % et 0,44 % respectivement pour les *Mén1adulte* et les *Mén2+adultes*.

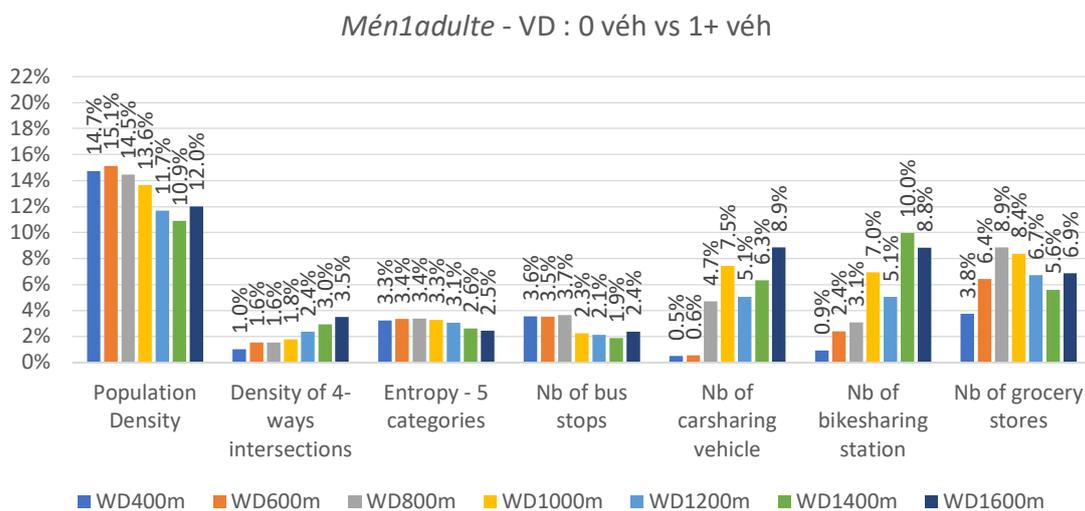


Figure 2-4 : Influence relative des variables d'environnement bâti pour les *Mén1adulte* selon différentes distances tampons testées.

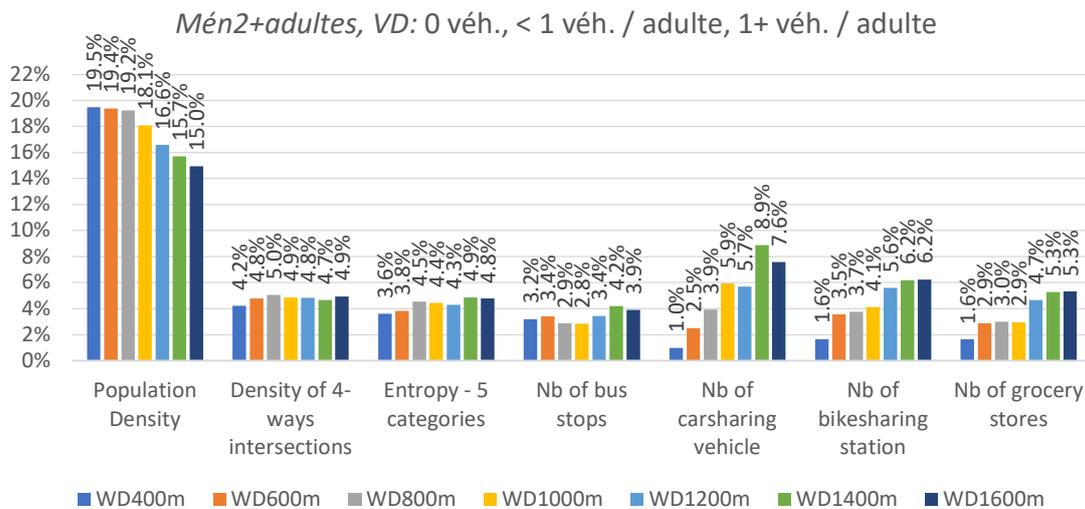


Figure 2-5 : Influence relative des variables d'environnement bâti pour les Mén2+adultes selon différentes distances tampons testées

Alors que l'influence relative des variables incluses dans les arbres de décision par GBM sont des indicateurs utiles, ils ne fournissent pas d'informations sur la direction de l'influence de chaque variable. Pour cela, des diagrammes de dépendance partielle, tels que décrits dans la section méthodologie, sont produits pour chaque variable explicative continue. Pour le modèle de résultat binaire des *Mén1adulte*, chaque courbe montre le changement de probabilité de posséder un véhicule en fonction des changements dans la valeur de la variable d'environnement bâti investiguée, en maintenant toutes les autres variables constantes à leur moyenne. Pour les ménages *Mén2+adultes*, trois courbes de dépendance partielle sont produites pour chaque variable. Chaque courbe représente le changement de probabilité d'appartenir à l'une des trois classes de résultats possibles (pas de voiture, moins d'une voiture par adulte et une voiture ou plus par adulte) lorsque la valeur de la variable d'environnement étudié change, en maintenant toutes les autres variables constantes à leur moyenne. Les probabilités pour les trois classes totalisent toujours 100 %. La valeur absolue de la probabilité dépend des interactions et de la distribution spécifique du résultat de toutes les variables incluses dans le modèle. Ce qui importe, c'est la **différence** entre les probabilités maximale et minimale pour une variable donnée et la forme de la distribution qui pourrait aider à identifier les effets de seuil. Pour éviter une interprétation risquée des modèles basés uniquement sur les valeurs aberrantes, chaque courbe de dépendance partielle est représentée pour une distribution allant du 1^{er} au 99^e centiles de chaque variable pour chaque sous-échantillon.

Pour les variables de l'environnement bâti basées sur la distance (distance au centre-ville et distance de marche à la station de métro la plus proche), des diagrammes de dépendance partielle (DDP) sont produits une fois pour chaque modèle de sous-échantillon et présentés sur les Figure et Figure .

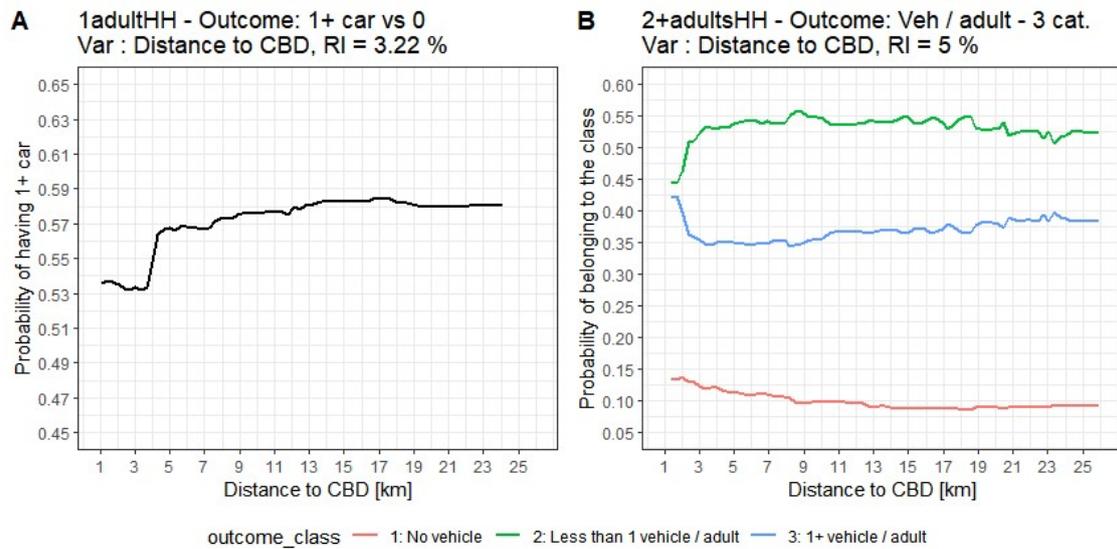


Figure 2-6 : Diagramme de dépendance partielle pour la distance à vol d'oiseau au centre-ville de Montréal

En considérant la distance au centre-ville pour le modèle des *Mén1adulte* (Figure , graphique A), il semble y avoir un effet de seuil entre 3,5 et 4,5 km où la probabilité d'être motorisé augmente fortement suivi d'une augmentation progressive jusqu'à 14 km. Pour les ménages *Mén2+adultes* (Figure , graphique B), la forte hausse est plus proche (entre 2 et 3,5 km du centre) et ne concerne que la probabilité d'être entièrement motorisé (classe 3) versus le partage de véhicules entre adultes du ménage (classe 2). Ceci indique que toutes choses étant égales par ailleurs, les ménages très proches du centre sont moins susceptibles d'avoir besoin d'un véhicule par adulte que ceux plus éloignés. La probabilité pour les *Mén2+adultes* de ne pas avoir de véhicule (courbe rouge sur la Figure , graphique B) diminue légèrement et lentement jusqu'à 9 km du centre-ville avant de rester stable au fur et à mesure que la distance augmente.

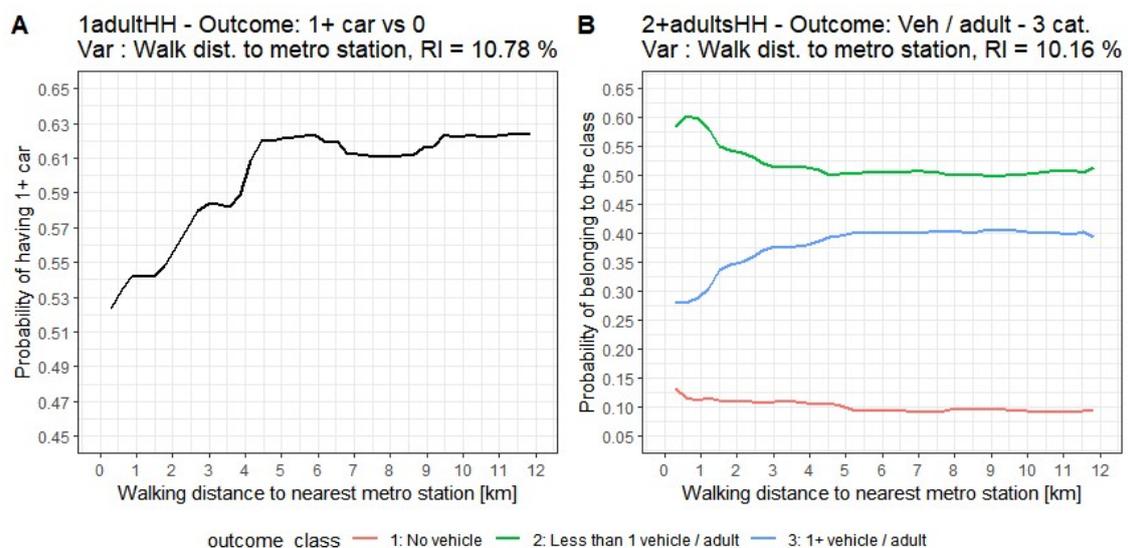


Figure 2-7 : Diagramme de dépendance partielle pour la distance de marche à la plus proche station de métro

L'accessibilité au réseau de métro semble avoir un effet clair sur les décisions de possession automobile chez les *Mén1adulte*. Le DDP (Figure , graphique A) montre trois augmentations régulières de la probabilité de posséder une voiture, suivies de trois plateaux. Le premier plateau va de 800 m à 1500 m (soit environ 10 à 20 minutes de marche), un second se dessine à environ 3 km représentant potentiellement une zone d'influence proche où les stations sont facilement accessibles en bus ou à vélo et un dernier apparaît à 4,5 km délimitant potentiellement l'ensemble du territoire considéré comme les « quartiers centraux » de Montréal. Pour les ménages de plusieurs adultes (Figure , graphique B), la probabilité d'avoir une ou plusieurs voitures par adulte (classe 3, courbe bleue) augmente régulièrement jusqu'à 5 km avec une diminution miroir de la probabilité d'avoir moins d'une voiture par adulte (classe 2, courbe verte). Au-delà de 6 km, aucun changement dans la probabilité d'appartenir à l'une ou l'autre classe n'est observé. La probabilité de ne pas avoir de véhicule (classe 1, courbe rouge) apparaît peu affectée par la proximité au réseau de métro.

Pour les variables basées sur le tampon, les courbes de dépendance partielle pour quatre des sept tailles de tampon testées pour les deux sous-échantillons sont présentées aux figures suivantes.

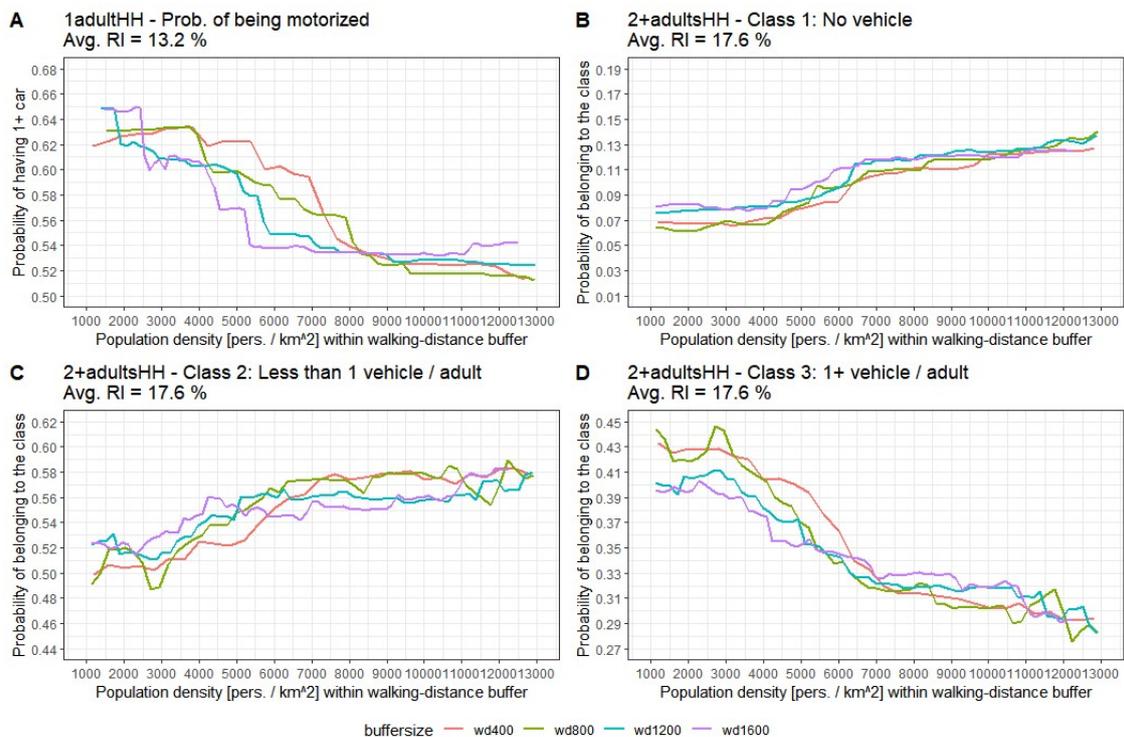


Figure 2-8: Diagramme de dépendance partielle pour la densité de population

La Figure indique que pour les *Mén1adulte* (graphique A), la probabilité d'être motorisé diminue assez régulièrement à mesure que la densité de population augmente jusqu'à 8500 personnes par km carré pour toutes les tailles de tampon sauf 1600 m où le plateau est atteint à 5500 personnes par km carré. Pour les *Mén2+adultes*, le point d'inflexion (formation d'un plateau) entre les trois classes est plutôt observé à une densité d'environ 7000 personnes par km carré (graphiques B, C et D). Des densités plus élevées ne semblent pas réduire davantage la probabilité de posséder une voiture pour les deux sous-échantillons. Les tampons plus petits présentent plus de plateaux

suivis de changements plus nets que les tampons plus grands où la densité est lissée parce qu'elle est moyennée sur des zones plus grandes.

Par soucis d'espace, les DDP pour les variables d'entropie, de densité d'intersections et du nombre de stations de vélopartage ne sont pas inclus.

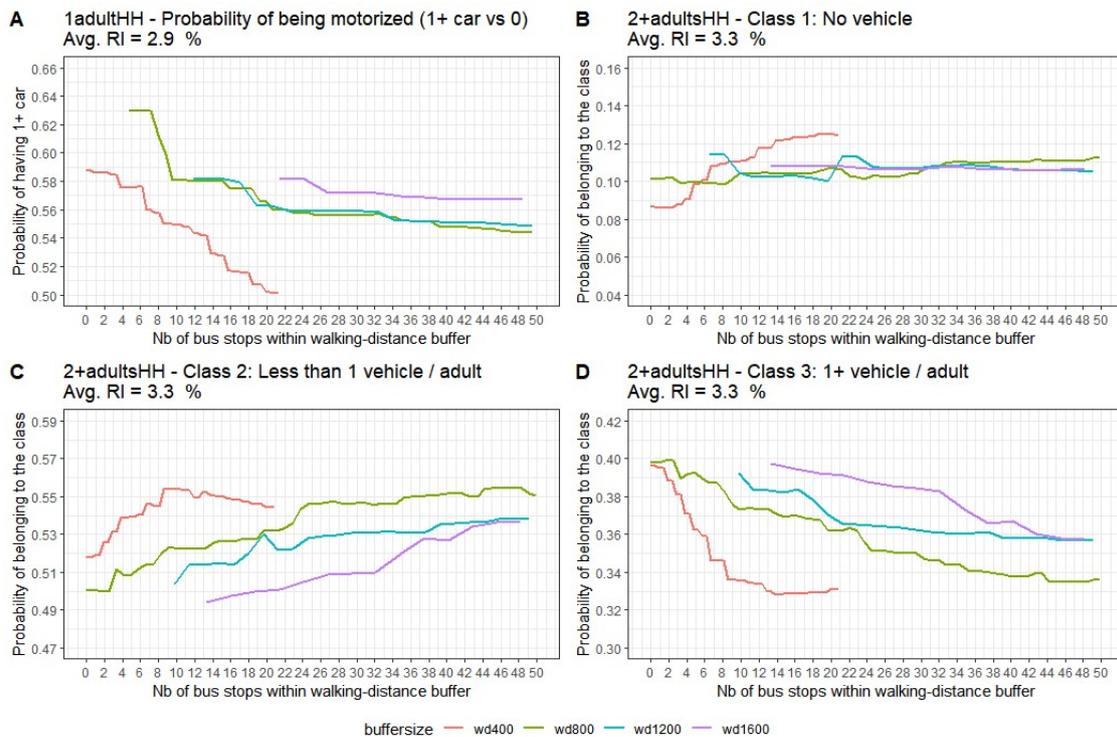


Figure 2-9 : Diagramme de dépendance partielle pour le nombre d'arrêts de bus

L'accès aux arrêts de bus est l'une des variables qui présentent la plus grande variabilité dans les distributions lorsque la taille des tampons passe de 400 m à 1600 m. Une diminution rapide des probabilités de possession d'une voiture pour les deux sous-échantillons de ménages est observée sur la Figure pour la plus petite zone tampon (400 m), tandis que pour les zones tampons plus grandes, les courbes sont plus lisses et ne se creusent pas aussi profondément. Cela pourrait être une indication que des tailles de tampon plus petites capturent mieux les effets de seuil de l'accessibilité des bus sur la possession automobile.

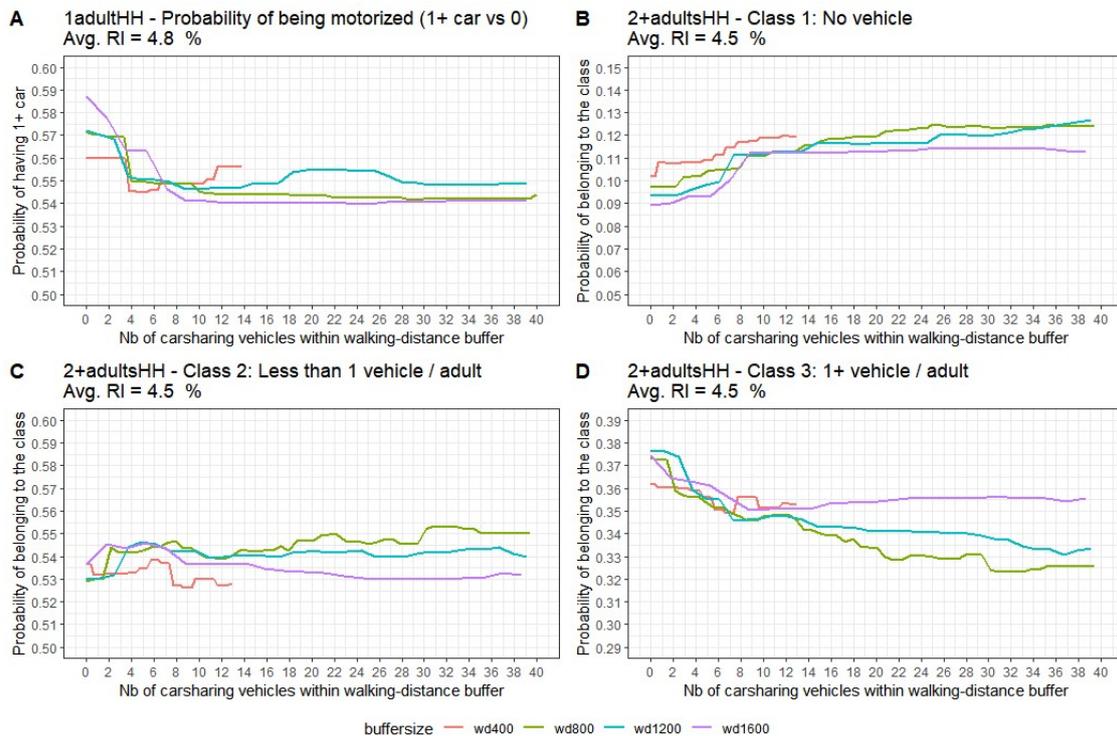


Figure 2-10: Diagramme de dépendance partielle pour le nombre de véhicules d'autopartage basé station

En ce qui concerne l'accessibilité aux véhicules d'autopartage, on observe que pour les *Mén1adulte* (Figure , graphique A), la probabilité de posséder une voiture chute fortement pour les zones tampons de 400, 800 et 1 200 m lorsque quatre véhicules d'autopartage en station sont disponibles à distance de marche. Pour les tampons de 1600 m, la probabilité la plus faible est atteinte à 8-9 véhicules. Pour les ménages *Mén2+adultes* (Figure 11, graphique B, C et D), la baisse globale la plus importante de la probabilité de posséder une voiture est observée lorsqu'une zone tampon de taille moyenne (800 m) est considérée. Pour ce tampon, la variation la plus nette est observée à deux véhicules, le point auquel la probabilité d'être entièrement motorisé (classe 3, graphique D) diminue tandis que la probabilité d'avoir moins d'un véhicule par adulte dans le ménage (classe 2, graphique C) et de ne pas avoir de véhicule (classe 1, graphique B) augmente en retour. Après cette première baisse, une baisse de la probabilité d'être entièrement motorisé est observée jusqu'à 20 véhicules. Cela semble indiquer qu'un nombre de plus en plus important de véhicules d'autopartage accessibles dans un rayon de 10 minutes à pied permet à davantage de ménages de plusieurs adultes de se débrouiller sans que chaque adulte n'ait besoin de sa propre voiture.

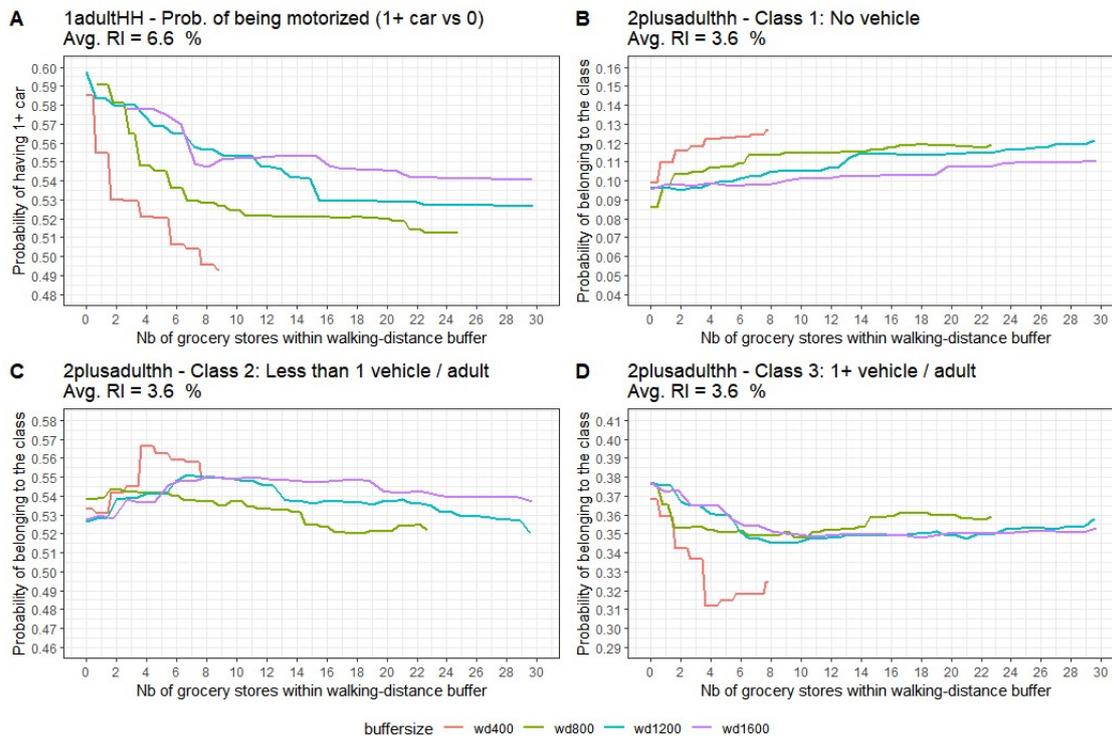


Figure 2-11 : Diagramme de dépendance partielle pour le nombre d'épicerie/fruiteries

Enfin, conformément aux autres variables d'accessibilité analysées (arrêts de bus, d'autopartage), des tailles de tampons plus petites (400 à 800 m) montrent une baisse rapide de la probabilité de posséder une voiture lorsque le nombre d'épicerie accessibles augmente (Figure). L'influence relative de la variable, qui affecte le changement de probabilité d'être motorisé, est plus élevée pour les *Mén1adulte* que pour les *Mén2+adultes* et donc le changement de probabilités est plus élevé dans le modèle de *Mén1adulte*. La variable épicerie comprend les grandes épicerie mais aussi les petites fruiteries locales très présentes dans les quartiers centraux de Montréal. Avoir accès à moins de 5 ou 10 minutes de marche (zones tampons de 400 m et 800 m) à plusieurs de ces commerces semble important pour aider les *Mén1adulte* à vivre sans voiture et, dans une moindre mesure, pour éviter la motorisation complète pour les *Mén2+adultes*. Pour les *Mén1adulte* (Figure , graphique A), lorsque l'on considère les courtes distances (400 m), seuls deux magasins semblent nécessaires pour exercer une forte influence sur la probabilité de posséder une voiture. Pour les ménages *Mén2+adultes* (Figure , graphique B, C et D), les tendances ne sont pas aussi claires, mais le fait d'avoir au moins quatre magasins à une très courte distance (400 m) est associé à une probabilité plus faible de motorisation complète (Classe 3, graphique D). En outre, la variable peut également être considérée comme un indicateur de l'accessibilité à un plus large éventail d'opportunités de commerces de proximité généralement situées dans les mêmes rues commerciales mais non capturées par les variables incluses.

Les principaux résultats pour chaque variable d'environnement bâti sont résumés dans le Tableau pour les *Mén1adulte* et dans le Tableau pour les *Mén2+adultes*.

Tableau 2-4 : Sommaire des résultats pour la modélisation des ménages à 1 adulte

Variables	Influence relative pour les 7 tampons testés			Éléments clés à retenir
	Moy.	Min IR	Max IR	
<i>Variables par tampon</i>				
Densité de population [pers. / km ²]	13.2%	1400 m 10.9%	600 m 15.1%	Les résultats sont cohérents pour toutes les tailles de tampon. Une convergence dans la probabilité d'être motorisé est observée à 8500 pers. / km ² pour tous les buffers (probabilité minimale observée).
Densité d'intersection 4-dir. [4-ways int. / km ²]	2.1%	400 m 1.00%	1600 m 3.50%	Les tampons plus large (1200-1600 m) montrent le plus de changement dans les probabilités d'être motorisé. Des plateaux de probabilité minimales sont atteint entre 10 to 15 intersection 4-directions / km ² .
Entropie 5 catégories [Entropie index 0-1]	3.1%	1600 m 2.50%	800 m 3.40%	Les tampons de >= 800 m performe bien avec des probabilités linéaires décroissantes dans la probabilité des motorisés lorsque l'entropie augmente de 0.25 to 0.75
NB d'arrêts de bus	2.8%	1400 m 1.90%	800 m 3.70%	Le tampon de 400 m montre la plus grande et la plus rapide diminution de probabilité de motorisation. Le plateau est atteint à 20 arrêts de bus à < 400 m de marche.
NB de véhicules d'autopartage	4.8%	400 m 0.50%	1600 m 8.90%	Le tampon de 1600 m montre la plus forte diminution de probabilité d'être motorisé. Le plateau le plus bas est atteint à 8 véhicules pour ce tampon. Pour les tampons plus petit (400, 800 et 1200 m), une diminution marquée est observée à 4 véhicules d'autopartage au sein du tampon.*
NB de stations de vélopartage	5.3%	400 m 0.90%	1400 m 10.00%	Le tampon de 1600 m montre la plus forte diminution de probabilité d'être motorisé. Le plateau le plus bas est atteint à 5 stations pour ce tampon. Pour les tampons plus petit (400, 800 et 1200 m), une diminution marquée est observée jusqu'à 6 stations de vélopartage au sein des tampons.*
Nb d'épiceries	6.7%	400 m 3.80%	800 m 8.90%	Le tampon de 400 m montre la plus grande et la plus rapide diminution de probabilité de motorisation. Le plateau est atteint à 8 épiceries à moins de 400 m de marche.
<i>Variables par distance</i>	RI dans le modèle de référence			
Distance à vol d'oiseau au centre-ville		3.2%		La plus rapide augmentation dans la probabilité d'être motorisée est observée entre 3 and 5 km du centre-ville, suivi par une faible augmentation jusqu'à 14 km.
Distance de marche au métro		10.8%		Trois augmentations linéaires sont observées : 0-800 m, 1,5 à 3 km et 3,8 to 4,5 km séparé par des plateaux. Aucune changement significatif n'est observé passé 4.5 km.

Tableau 2-5 : Sommaire des résultats pour la modélisation des ménages à 2 adultes et plus.

Variables	Influence relative pour les 7 tampons testés			Éléments clés à retenir
	Moy.	Min IR	Max IR	
<i>Variables par tampon</i>				
Densité de population [pers. / km ²]	17.6%	1600 m 15.0%	400 m 19.5%	Les résultats sont cohérents pour toutes les tailles de tampon. Changement dans les probabilités de motorisation complète (↓), motorisation partielle (↑) et 0 véhicule (↑) sont observés pour des densité allant jusqu'à 6500 to 8500 pers. / km ² .
Densité d'inter. 4-dir. [4-ways int. / km ²]	4.8%	400 m 4.2%	800 m 5.0%	Les résultats sont cohérents pour des tampons >= 800 m. Des changements dans les probabilités de motorisation sont observés pour des densités d'intersections à 4 directions entre 0 et 8 et entre 36 et 50 [int. / km ²] avec aucun changement entre ces deux groupes de valeurs.
Entropie 5 catégories [Entropie index 0-1]	4.3%	400 m 3.6%	1200 m 4.9%	Les résultats sont cohérents pour toutes les tailles de tampon. Lente déclin des probabilités de motorisation complètes sont observés pour des valeurs d'entropie entre 0,25 to 0,80 avec des (↑) dans les probabilités de 0 véhicule et de motorisation partielle.
NB d'arrêts de bus	3.4%	1000 m 2.8%	1400 m 4.2%	Le tampon de 400 m montre un changement le plus important et le plus rapide dans les statuts de motorisation. Pour ce tampon, une (↑) dans la probabilité d'avoir aucun véhicule et une (↓) dans la probabilité de motorisation complète jusqu'à 14 arrêts. Les probabilités de motorisation partielle (↑) lentement jusqu'à 9-10 arrêts de bus accessible dans le tampon.
NB de véhicules d'autopartage	5.1%	400 m 1.0%	1200 m 8.9%	Le tampon de 800 m montre le plus important changement dans les probabilités de motorisation. Pour ce tampon, les probabilités d'avoir aucun véhicule (↑) et celle d'être complètement motorisé (↓) jusqu'à un plateau de 20 véhicules dans le tampon. Les probabilités de motorisation partielle (↑) lentement jusqu'à 30 véhicules.
NB de stations de vélopartage	4.4%	400 m 1.6%	1600 m 6.2%	Le tampon de 800 m montre le plus important changement dans les probabilités de motorisation alors que celui de 400 m montre le plus rapide changement jusqu'à 6 stations de vélopartage. Pour le tampon de 800 m, les changements dans les probabilités de n'avoir aucun véhicule et d'être partiellement motorisé (↑) et les probabilités d'être complètement motorisé (↓) jusqu'à un plateau de 12 à 16 stations.
Nb d'épiceries	3.7%	400 m 1.6%	1600 m 5.3%	Le tampon de 400 m montre un changement le plus important et le plus rapide dans les statuts de motorisation. Les plus haute probabilité de n'avoir aucun véhicule et d'être partiellement motorisé et les plus faibles probabilités d'être complètement motorisé sont atteintes à 4 épiceries à moins de 400 m.
<i>Variables par distance</i>		RI, modèle de réf.		
Distance à vol d'oiseau au centre-ville		5.1%		Une (↑) rapide des prob.de motorisation et une (↓) équivalente dans celles d'être partiellement motorisé sont observées entre 1.5 et 3.5 km . La prob. de n'avoir aucun véhicule (↓) lentement jusqu'à 9 km.
Distance de marche au métro		10.2%		Une (↑) rapide des prob. de motorisation et une (↓) équivalente dans celles d'être partiellement motorisé sont observées entre 0.5 and 5 km . La prob. de n'avoir aucun véhicule (↓) lentement jusqu'à 5 km.

Notes: Zéro véhicule = classe 1: 0 véhicule, Motorisation partielle = classe 2: < 1 véhicule / adulte, Motorisation complète = Class 3: 1+ véhicule / adulte.

2.3.6 Discussion et conclusion

Cette analyse utilise des arbres de décision par Gradient Boosting Machines (GBM) pour étudier comment l'échelle de mesure des variables de l'environnement bâti affecte leur relation estimée avec le statut de possession automobile des ménages et pour étudier les effets de seuil potentiels. Pour ce faire, des variables d'intérêt ont été calculées à l'intérieur de sept zones tampons de distance de marche de plus en plus grandes (de 400 m à 1600 m par saut de 200 m) autour de chaque ménage observé de l'Enquête Origine-Destination 2018 de Montréal. Une série de modèles GBM a été exécutée pour deux sous-échantillons de ménages : *Mén1adulte* et *Mén2+adultes*, en faisant varier la taille de tampon pour une variable à la fois tout en maintenant le tampon des autres variables à une valeur de référence de 1 km. Pour les deux sous-échantillons, les variables d'environnement bâti incluses ont une influence relative combinée plus élevée que les variables sociodémographiques incluses. Les résultats de l'influence relative comparée et des diagrammes de dépendance partielle des trois mesures plus « traditionnelles » de l'environnement bâti (densité, diversité et conception) sont cohérents entre les tailles de tampon testées pour les deux sous-modèles, ce qui indique que ces mesures ne sont pas trop sensibles à la distance choisie pour leur mesure. Cela pourrait s'expliquer par leur nature relative, c'est-à-dire qu'elles sont moyennées sur des surfaces, et le fait que celles-ci sont généralement considérées comme des mesures « proxy » de l'environnement bâti. Ces variables présentent également les relations les plus linéaires de toutes les variables avec la probabilité d'être entièrement motorisé et peuvent donc être utilisés « en toute sécurité » dans des modèles de régression logistique couramment utilisés pour modéliser la possession automobile.

En revanche, les mesures d'accessibilité aux ressources de mobilité partagée (autopartage, vélopartage) et aux opportunités commerciales locales (épicerie) présentent toutes des changements importants dans leur influence relative au sein des modèles lorsque la taille de la zone tampon utilisée pour le calcul passe progressivement de 400 m à 1 600 m. De plus, les diagrammes de dépendance partielle de ces variables montrent des effets de seuil clairs sur la probabilité pour les deux types de ménages d'être entièrement motorisés lorsque de petits tampons sont utilisés (≤ 800 m). En termes d'accessibilité au transport en commun, la distance de marche à la station de métro la plus proche est la variable d'accessibilité avec la plus grande influence sur la probabilité des motorisés, atteignant une influence relative d'environ 10 % pour les deux sous-échantillons de ménages, juste derrière la densité de population parmi les variables d'environnement bâti incluses.

L'influence relative de l'accessibilité aux arrêts de bus (le nombre dans la zone tampon) était parmi les plus faibles pour les variables d'environnement bâti incluses (~ 3 %), mais était très cohérente entre les tailles de zone tampon testées. Cependant, la forme des courbes de dépendance partielle change radicalement entre les tailles de tampon testés. La plus petite zone tampon (400 m) a montré la plus forte baisse de la probabilité de posséder une voiture à mesure que le nombre d'arrêts de bus augmente et confirme ainsi qu'elle est mieux adaptée pour capturer l'effet sur la possession d'une voiture que des zones tampon plus grandes telles que 1200 m ou 1600 m. Bien que l'accès aux arrêts de bus soit une mesure décente, l'effet de l'accessibilité des bus sur la possession d'une voiture pourrait être mieux capturé dans les recherches futures en incluant un indicateur de niveau de service à ces arrêts de bus. Dans l'ensemble, les résultats indiquent que le choix de la bonne échelle (taille du tampon ou des unités d'agrégation) lors du calcul des variables d'accessibilité pour la modélisation de la possession d'une voiture est essentiel car cela peut conduire à des résultats très différents. De plus, la plupart des variables d'accessibilité présentent une non-linéarité avec les probabilités de possession d'une voiture, c'est-à-dire que

les changements de probabilités montrent des effets de seuil où des changements rapides sont observés avant une certaine valeur, suivis de changements nuls ou très lents après cette valeur. Les effets de seuil sont les plus forts lorsque les variables sont calculées à l'intérieur de petits tampons. Cela indique que la prudence est recommandée lors de l'inclusion de ces variables continues dans les régressions logistiques qui supposent généralement l'homoscédasticité.

En termes d'implications dans les politiques de transport et de planification, les valeurs seuils identifiées et résumées dans les Tableau 2-4 et Tableau 2-5 fournissent des indices empiriques sur les valeurs à viser (par exemple, le nombre d'opportunités, la valeur de densité) et les distances à prendre en compte pour atteindre ces valeurs et ainsi soutenir la vie avec une moindre motorisation pour une diversité de ménages.

Quelques limites doivent être mentionnées. Premièrement, bien que le revenu ait été contrôlé dans les modèles, son hétérogénéité connue par rapport à l'influence de l'environnement bâti (Bhat & Guo, 2007) pourrait ne pas être correctement capturée par la méthode des diagrammes de dépendance partielle (DDP) utilisée. Deuxièmement, les mesures de l'accessibilité aux possibilités d'emplois n'ont pas été incluses et devraient l'être dans les recherches futures. D'un point de vue méthodologique, les DDP produits à partir des modèles GBM ne peuvent qu'approximer la relation entre les variables explicatives et la variable dépendante et leur précision peut être faible en présence d'interactions complexes entre variables (Hastie et al., 2009). Une autre limite importante est l'incapacité du modèle à distinguer l'effet « réel » de l'environnement bâti sur les décisions de propriété d'une voiture de l'effet potentiel causé par l'auto-sélection résidentielle. Les quartiers à forte densité et à usage mixte avec une grande accessibilité aux ressources de mobilité et aux opportunités locales pourraient attirer des ménages moins enclins à posséder une voiture, ce qui pourrait expliquer une partie de la relation observée entre les variables analysées et les probabilités d'être motorisé des ménages.

Enfin, les recherches futures devraient utiliser des méthodes de modélisation alternatives avec des ensembles de données d'autres villes pour valider les valeurs seuils obtenues ici - c'est-à-dire tenter d'évaluer « combien » de chaque attribut - sont nécessaires pour influencer les décisions de motorisation des ménages et, surtout, valider quelles distances réseaux sont nécessaires pour atteindre ces valeurs.

2.4 Combinaison de l'enquête Origine-Destination de Montréal et de flux de données passives pour permettre un suivi longitudinal de la mobilité

- Par : Elodie Deschaintres, doctorat
- Supervision : Pre Catherine Morency / Pr Martin Trépanier
- Contexte : Chaire de recherche du Canada sur la mobilité des personnes
- Fin prévue : automne 2022

Cette section est constituée d'extraits de la thèse de doctorat en cours de rédaction d'Élodie Deschaintres.

2.4.1 Introduction

Des enquêtes ménages déplacements sont réalisées depuis longtemps pour guider la planification des systèmes de transport et de la mobilité. De plus, au cours des dernières années, le rapide développement des technologies de l'information et de la communication a rendu disponible une grande variété d'ensembles de données passives : entre autres, des traces GPS, des données de capteurs, des comptages, des données cellulaires, des données de cartes à puce ou d'autres données transactionnelles (pour les services de mobilité partagée, par exemple). Toutefois, malgré cette multitude de sources de données, la recherche en transport souffre encore d'un manque d'informations qui soient à la fois multimodales et longitudinales. En effet, il n'existe pas de base de données unique décrivant de manière continue l'utilisation de tous les modes de transport en même temps.

D'une part, les déplacements effectués avec tous les modes sont recueillis dans les enquêtes traditionnelles, mais les modes moins fréquents et plus irréguliers sont généralement sous-observés. En outre, les informations sont transversales (i.e., collectées à des moments précis, à plusieurs années d'intervalle) et les carnets de déplacements sont regroupés pour représenter un jour de semaine moyen « typique ». Par conséquent, ces données ne tiennent pas compte de la variabilité des comportements de mobilité, et les informations compilées deviennent rapidement datées et moins représentatives des conditions actuelles. D'autre part, les données passives sont longitudinales, i.e. disponibles en continu sur de longues périodes, mais elles sont collectées de manière indépendante et hétérogène pour chaque mode ; elles rendent donc difficile l'analyse des interactions entre les modes.

La fusion de données est une solution proposée dans la littérature pour tirer profit de toutes les sources de données et compenser leurs limites respectives. Dans cette perspective, cette recherche vise à enrichir une enquête ménages déplacements (l'enquête Origine-Destination de Montréal, menée tous les 5 ans) avec des flux de données passives pour permettre un suivi longitudinal de l'utilisation de plusieurs modes de transport. Le principe consiste à appliquer les variations détectées dans les données passives aux indicateurs de mobilité moyens estimés avec l'enquête. À cette fin, une procédure automatisée basée sur la décomposition de série temporelle est développée. Cette méthode est appliquée à l'enquête OD de Montréal 2013 et validée avec l'enquête 2018. Elle permet d'annualiser les parts modales quotidiennes de trois modes (métro, vélo, automobile) et de suivre leur évolution sur cinq ans, entre les deux enquêtes, en s'appuyant sur la saisonnalité et la tendance long-terme mises en évidence dans les données passives.

2.4.2 Revue de la littérature

La fusion (ou intégration) de données n'est pas un concept nouveau en transport. Les enquêtes sont combinées depuis longtemps, par exemple pour regrouper des données de préférences révélées et déclarées (Hensher et al., 1998), ou des données de recensement et d'enquête sur les

déplacements (Venigalla, 2004), ce dernier cas étant également impliqué dans la pondération de données échantillonnées. Cependant, l'utilisation de ce terme s'est intensifiée ces dernières années. Trois principales causes ont été avancées par Bayart et al. (2009). Premièrement, les enquêtes de déplacements sont devenues moins représentatives en raison de la baisse des taux de réponse et de la réception de formulaires moins complets/précis. Cette tendance peut s'expliquer par différents facteurs, notamment l'exposition à un nombre croissant de demandes d'enquête, provoquant des effets de fatigue, mais aussi une plus grande anxiété à l'idée de révéler des informations personnelles. En outre, certains sous-groupes de population sont plus difficiles à atteindre. Ainsi, parmi les exemples récents de combinaison d'enquêtes, on trouve la fusion de plusieurs modes d'enquête, ou l'intégration d'une enquête satellite à vocation spéciale (par exemple, ciblant les étudiants) à une enquête ménages déplacements centrale (Verreault et Morency, 2018 ; Wang, K. et al., 2021). Une deuxième cause de l'intérêt grandissant pour la fusion de données selon Bayart et al. (2009) est la prise de conscience de la complexité des phénomènes urbains. Cela conduit les chercheurs à croiser les bases de données du transport avec des informations issues de différents domaines, comme l'urbanisme et la santé, ou sous différentes dimensions (économiques, environnementales et sociales par exemple). Enfin, une troisième cause (celle qui est à l'origine de cette recherche) est la disponibilité croissante de systèmes d'observation pour collecter des données passives.

Ces données émergentes sont communément considérées comme une opportunité pour compléter les données traditionnelles, notamment les enquêtes (Bonnell & Munizaga, 2018 ; Miller et al., 2018). La comparaison des deux types de données (données passives versus enquêtes) met en évidence les avantages potentiels de leur fusion. Les enquêtes sont transversales et souvent basées sur des carnets de déplacements d'une journée, alors que les données émergentes sont longitudinales et fournissent des informations spatio-temporelles très précises. Il existe des alternatives aux enquêtes transversales d'une journée, par exemple les enquêtes multi-jours, les enquêtes par panel et les enquêtes continues - voir Ampt (2013) et Ortúzar et al. (2011) pour des définitions et une revue sur le sujet -, mais elles sont beaucoup plus coûteuses et laborieuses que les données émergentes, collectées passivement et donc sans fardeau pour le répondant. Cependant, les données émergentes sont unidimensionnelles, axées sur la mobilité sans informations contextuelles (pour des raisons de confidentialité), tandis que les enquêtes traditionnelles contiennent des attributs sociodémographiques, nécessaires pour expliquer les comportements de mobilité (Bayart et al., 2009 ; Cherchi & Bhat, 2018). Par conséquent, les données émergentes et les enquêtes traditionnelles répondent respectivement aux questions « comment ? » et « pourquoi ? » (Callegaro & Yang, 2018). D'autres arguments ont été avancés pour justifier que les données émergentes ne peuvent pas remplacer les enquêtes traditionnelles, notamment les préoccupations concernant la qualité, la complétude et la représentativité des informations disponibles, ainsi que l'accès aux données (Bonnell & Munizaga, 2018 ; Miller et al., 2018). En particulier, peu de travaux bénéficient de suffisamment de données pour étudier simultanément différents services de transport (Morency et al., 2018). De plus, les données émergentes sont généralement massives (big data) et donc difficiles à traiter, et les méthodes de collecte associées sont encore en évolution (Miller et al., 2018).

Par conséquent, les deux types de données ont des avantages et des inconvénients qui peuvent être exploités dans une perspective de fusion de données pour maximiser leur valeur et dépasser leurs limites. L'objectif du problème de la fusion de données est de transmettre des informations supplémentaires d'une base de données donneuse à une base de données réceptrice grâce à des variables qui sont communes aux deux bases de données. Les procédures de fusion de données sont souvent classées en deux catégories : l'appariement exact (*exact matching*), lorsque deux

enregistrements (un de chaque base de données) sont appariés à l'aide d'un identifiant unique, ou l'appariement statistique (*stastistical matching*), basé sur des modèles explicites ou implicites (Bayart et al., 2009 ; Zhu et al., 2018). Cependant, il y a peu d'exemples d'application de ces méthodes dans un contexte de transport. Ce manque est dû à plusieurs défis techniques, tels que l'hétérogénéité des formats et des unités de mesure (ou indicateurs d'usage). En particulier, les données passives ayant été initialement produites pour d'autres fonctions que l'analyse et la modélisation de la mobilité, elles ne contiennent pas de déplacements tels que définis dans les enquêtes mais des transactions, des comptages, des points GPS ou d'autres informations spécifiques (Bonnell & Munizaga, 2018). Au-delà de cette dissemblance sémantique, il existe également des incompatibilités temporelles et spatiales entre les sources de données, qui peuvent avoir été collectées avec des niveaux de résolution différents. En particulier, le territoire enquêté et la population de référence correspondante sont différents entre les enquêtes de déplacements et les données passives (résidents de l'ensemble de la région versus utilisateurs, incluant résidents et visiteurs, d'un système donné) (Miller et al., 2018). Toutes ces disparités conduisent à un manque de variables communes et à la nécessité d'agréger les données. Par conséquent, de nombreux auteurs s'accordent à dire que de nouveaux outils sont nécessaires pour permettre l'intégration de diverses sources de données, ainsi que pour mieux définir le rôle des données traditionnelles et émergentes (Cherchi & Bhat, 2018).

Les cas d'application les plus courants de techniques de fusion de données sur des données passives proviennent du domaine de l'ingénierie du trafic routier et des systèmes de transport intelligents (El Faouzi & Klein, 2016). Plusieurs ensembles de données passives ont également été recoupés pour valider ou compléter certaines informations, comme les origines et/ou les destinations de déplacements (Giraud et al., 2016 ; Lovelace et al., 2016). Cependant, pour les raisons précédemment évoquées, les exemples d'intégration entre des données passives et des enquêtes traditionnelles sont beaucoup plus rares. Néanmoins, certains auteurs ont exploité des enquêtes ménages déplacements et des données de cartes à puce pour comparer ou valider des distributions spatiales et temporelles (Munizaga et al., 2014 ; Spurr et al., 2015), et des procédures d'enrichissement (basées sur des modèles ou par appariement) ont été développées pour ajouter aux données passives des attributs de déplacement tels que le motif (Kusakabe & Asakura, 2014) ou des informations sociodémographiques (Grapperon et al., 2016) provenant d'enquêtes. Par conséquent, dans les travaux précédemment cités, une enquête de déplacements (donneur) a été exploitée pour compléter les informations dans une base de données passives (récepteur) ou les deux sources de données (les deux en tant que donneurs) ont été valorisées à des fins de validation/comparaison, mais aucune étude n'explore la transmission d'informations depuis des données passives (donneur) vers une enquête traditionnelle (récepteur). Pourtant, le modèle de collecte de données noyau-satellite proposé dans plusieurs articles (Miller et al., 2018 ; Wang, K. et al., 2021) considère l'enquête ménages déplacements comme centrale et les ensembles de données passives comme auxiliaires. C'est également l'approche adoptée dans ce travail, qui vise à appliquer les variations capturées dans des données longitudinales à une enquête transversale.

2.4.3 Données et méthodologie

2.4.3.1 Description et traitement des données

Cette recherche bénéficie de deux types de données. Premièrement, l'enquête Origine-Destination (OD) de Montréal, Canada, est une enquête téléphonique (et web depuis 2018) menée tous les cinq ans dans la région métropolitaine de Montréal pour enregistrer les déplacements d'environ 4 à 5% de tous les ménages résidents sur un jour de semaine d'automne. Les deux dernières enquêtes, à savoir celles de 2013 et 2018 (78 731 et 75 266 ménages

respectivement), sont utilisées dans ce travail. Toutes deux ont été pondérées en appliquant une méthode de calage sur marges (les totaux des catégories étant tirés du recensement canadien) afin de dépeindre des comportements de déplacement représentatifs d'un jour typique de semaine. Elles fournissent donc des informations transversales ; elles permettent de calculer des indicateurs de mobilité quotidienne moyens, tels que des parts modales, pour l'automne 2013 et l'automne 2018.

Deuxièmement, plusieurs flux de données passives sont exploités pour décrire les variations longitudinales dans l'utilisation de trois modes de transport du 1er janvier 2013 au 31 décembre 2018 : des données de validation pour le métro, et des données de comptage pour les vélos et les automobiles privées. Toutes les validations (quel que soit le support : carte à puce ou ticket) collectées par le système OPUS à l'entrée des stations de métro (*tap-in*) sur la période d'étude ont été fournies par la STM. Les données sur le vélo proviennent du système Eco-compteur de la Ville de Montréal, tandis que les compteurs (situés sur plusieurs autoroutes et ponts) du système CIR du MTQ sont utilisés pour le trafic routier. Les comptages vélos comprennent les vélos personnels et les vélos en libre-service (BIXI), alors que seules les voitures particulières (sans les camions) sont extraites des données de circulation classifiées. Cependant, dans les deux jeux de données, les informations de quelques compteurs seulement sont (presque) continues entre 2013 et 2018. Deux compteurs par mode sont donc sélectionnés, en fonction du nombre d'observations manquantes et de leur localisation (les compteurs sur les ponts sont privilégiés pour la voiture car ils sont généralement moins affectés par la congestion). Les compteurs vélos (Berri et Parc) sont situés près du centre-ville, tandis que les compteurs automobiles sont positionnés sur des ponts (Viau et Lachapelle) reliant l'île de Montréal à la banlieue. Ainsi, par leur emplacement, ces compteurs captent principalement des déplacements domicile-travail, qui sont généralement bien représentés dans l'enquête OD de Montréal (Spurr et al., 2015).

Plusieurs étapes de prétraitement des données passives sont nécessaires pour agréger l'utilisation de chaque mode (en nombre de validations ou de comptages) en séries d'observations journalières sur six ans. Les validations et les comptages sont additionnés par 24 heures, de 4h00 à 28h00 (4h00 le lendemain) : c'est la définition d'une journée dans l'enquête OD, où les déplacements collectés de 24h00 à 27h59 sont des déplacements qui ont commencé le jour suivant celui enquêté et qui permettent typiquement de conclure une chaîne de déplacements. Lorsque les validations manquent sur plusieurs heures et pour plusieurs stations, la somme journalière est également considérée comme manquante dans les données du métro. Les comptages vélos et automobiles sélectionnés sont d'abord sommés par compteur, c'est-à-dire sans distinguer la direction, puis par mode. Lorsqu'une observation quotidienne est manquante pour un compteur, la somme avec l'autre compteur est également considérée comme manquante. Finalement, on obtient des séries temporelles de $365 \times 5 + 366 = 2191$ jours par mode, avec 3, 8 et 122 valeurs manquantes pour le métro, le vélo et la voiture particulière respectivement. Par conséquent, toutes les séries de données utilisées dans ce travail sont incomplètes, mais aussi hétérogènes (validations et comptages dans les données passives, versus déplacements dans l'enquête OD). De plus, les volumes de vélos et d'automobiles ne sont pas représentatifs de l'ensemble de la région. Cependant, la méthode développée tient compte des valeurs manquantes et des valeurs aberrantes, et seules les variations dans les données passives (ni les valeurs, ni la métrique) sont considérées.

2.4.3.2 Aperçu de la méthode

La procédure de fusion développée pour combiner l'enquête OD et les flux de données passives est présentée dans la Figure 2-12. Elle consiste, pour chaque mode de transport, à appliquer les

variations longitudinales détectées dans les données passives (disponibles de 2013 à 2018 et converties en une série temporelle agrégée par jour) à l'utilisation quotidienne moyenne observée dans l'enquête OD, par exemple l'enquête de 2013. Les variations dans l'utilisation du mode capturées dans les données passives sont mises en évidence à l'aide d'une méthode de décomposition de série temporelle. Cette méthode décompose les données en trois types de composantes (la tendance à long terme, la saisonnalité et le reste) liées par une équation additive ou multiplicative, selon si les composantes sont additionnées ou multipliées entre elles pour revenir aux données initiales. Plus précisément, l'algorithme STL (*Seasonal and Trend decomposition using Loess*) est utilisé. Cette approche est détaillée dans la section suivante. Les composantes sont ensuite normalisées par la moyenne des jours de semaine d'automne. L'opération conduit à des coefficients normalisés qui sont appliqués à un indicateur d'utilisation modale, tel que le nombre de déplacements effectués avec un mode donné, calculé à partir de l'enquête OD 2013. Il en résulte une projection et une annualisation de l'indicateur d'utilisation typique observé à l'automne 2013 sur six années complètes, incluant les fluctuations journalières. La dernière enquête OD (2018) sert enfin à valider les résultats en comparant le nombre moyen de déplacements prévu par la méthode pour l'automne 2018 à la valeur observée dans l'enquête. Ainsi, les patrons longitudinaux extraits des données passives permettent de combler l'écart de 5 ans entre les deux enquêtes, et de rendre compte de la variabilité des comportements de mobilité dans le temps (plutôt que de seulement considérer un jour typique de semaine d'automne).

Pour chaque mode de transport :

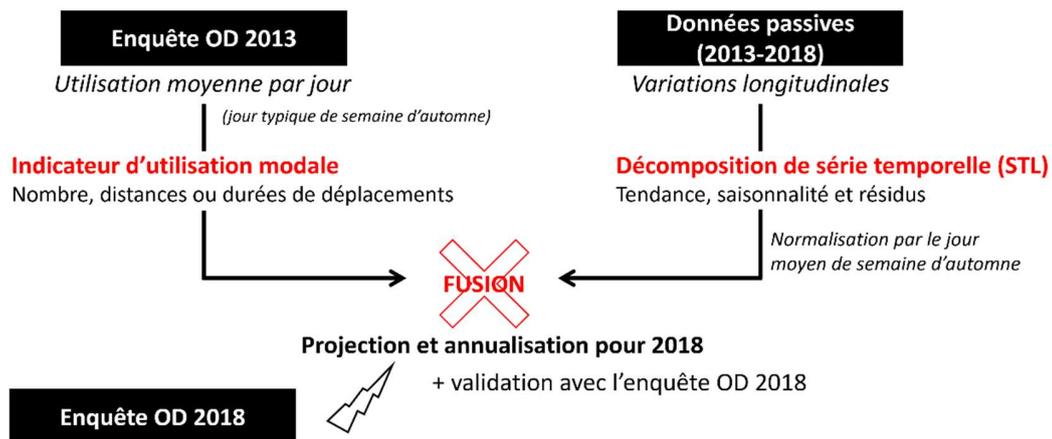


Figure 2-12 Procédure de fusion de données entre l'enquête Origine-Destination de Montréal et des données passives

2.4.3.3 Décomposition STL (*Seasonal and Trend decomposition using Loess*)

Il existe plusieurs méthodes de décomposition de série temporelle, mais la méthode STL, développée par Cleveland et al. (1990), est choisie dans ce travail car elle dispose de nombreux avantages par rapport aux autres. En particulier, la décomposition STL est plus flexible et permet la présence de valeurs manquantes dans les données (ce qui est le cas dans les données utilisées). L'objectif de la méthode est de décomposer une série temporelle Y_t en une composante tendancielle T_t (i.e., les variations de basse fréquence dans les données et les changements de niveau à long terme), une ou plusieurs composantes de saisonnalité S_t (i.e., les variations dans les données étant donnée une certaine fréquence) et une composante résiduelle (i.e., les résidus) R_t . La forme additive s'écrit comme suit, pour toutes les observations t :

$$Y_t = T_t + S_t + R_t$$

La méthode STL ne fournit pas directement de décomposition multiplicative, mais on peut y parvenir en prenant le logarithme des données, tel que :

$$\log(Y_t) = \log(T_t) + \log(S_t) + \log(R_t) \quad \text{ou} \quad Y_t = T_t * S_t * R_t$$

Pour obtenir cette décomposition, la méthode STL applique une séquence d'opérations de lissage basées sur une régression pondérée localement de type Loess (*locally estimated scatterplot smoothing*), c'est-à-dire l'estimation locale d'un polynôme de degré d pour chaque valeur à partir de ses q voisins les plus proches pondérés par des poids décroissant avec leur distance depuis la valeur. L'algorithme complet est détaillé ci-dessous. Il comprend une boucle interne, composée de plusieurs étapes pour estimer la tendance et les composantes de saisonnalité. Celle-ci est imbriquée dans une boucle externe, qui calcule la composante résiduelle et qui peut réduire l'influence des points aberrants (*outliers*) ; elle permet ainsi une estimation robuste. Si nécessaire, la robustesse est obtenue par ajustement itératif des moindres carrés pondérés dans les opérations de lissage : les poids de voisinage sont multipliés par des poids de robustesse (les plus grandes valeurs en absolu des résidus reçoivent les poids de robustesse les plus faibles).

STL – ALGORITHME GÉNÉRAL

<p>Boucle externe (n_o itérations)</p> <p>Boucle interne (n_i itérations)</p> <ul style="list-style-type: none"> - <u>Étape 1</u> : Détendancialisation à $Y_t - T_t$ - <u>Étape 2</u> : Lissage des sous-séries du cycle avec $q = n_s$ et $d = 1$ à C_t - <u>Étape 3</u> : Application d'un filtre passe-bas¹ sur les sous-séries lissées avec $q = n_l$ et $d = 1$ à L_t - <u>Étape 4</u> : Détendancialisation des sous-séries lissées à $S_t = C_t - L_t$ - <u>Étape 5</u> : Désaisonnalisation à $Y_t - S_t$ - <u>Étape 6</u> : Lissage de la tendance avec $q = n_t$ et $d = 1$ à T_t <p>Calcul de $R_t = Y_t - T_t - S_t$ et de poids de robustesse pour chaque observation (détection des valeurs aberrantes)</p>
--

où Y_t est la série temporelle correspondant aux données initiales, T_t la composante tendancielle, S_t la composante de saisonnalité, R_t la composante résiduelle, C_t une composante intermédiaire regroupant l'ensemble des sous-séries du cycle lissées, L_t une composante intermédiaire regroupant l'ensemble des sous-séries du cycle lissées et filtrées.

STL – PARAMÈTRES DE LA DÉCOMPOSITION

<p>n_p : nombre d'observations dans chaque période ou cycle de la composante de saisonnalité basé sur les propriétés des données</p> <p>n_o : nombre d'itérations de la boucle externe $n_o = 5$ ou 10 si on veut contrôler la robustesse, $n_o = 0$ sinon</p> <p>n_i : nombre d'itérations de la boucle interne $n_i = 1$ si on veut contrôler la robustesse, $n_i = 2$ sinon</p> <p>n_s : paramètre de lissage de la composante de saisonnalité entier impair supérieur ou égal à 7 ; à déterminer graphiquement et/ou sur la base de connaissances préalables</p> <p>n_l : paramètre de lissage du filtre passe-bas $n_l = [n_p]_{odd}$</p> <p>n_t : paramètre de lissage de la composante tendancielle $n_t = [1.5n_p / (1 - 1.5/n_s)]_{odd}$</p> <p><u>Notation</u> : $[x]_{odd}$ est le premier entier impair supérieur ou égal à x</p>
--

La boucle interne est composée de six étapes interconnectées qui visent à empêcher les composantes tendancielle et saisonnière de se disputer les mêmes variations dans les données. Tout d'abord, les données sont détendancialisées, la tendance étant nulle à la première itération de la boucle (étape 1). La série détendancialisée est ensuite utilisée dans les étapes 2, 3 et 4 pour évaluer la composante de saisonnalité. n_p , le nombre d'observations dans chaque période (ou cycle) de la composante de saisonnalité, dépend des propriétés des données. Par exemple, si la série est mensuelle avec une périodicité annuelle, alors $n_p = 12$. Les sous-séries du cycle sont

¹ Filtre passe-bas : Transformation mathématique qui laisse passer les basses fréquences et qui atténue les hautes fréquences.

les sous-séries de valeurs à chaque position du cycle saisonnier, par exemple toutes les valeurs de janvier, de février, etc. Ces sous-séries sont d'abord lissées par la méthode Loess et toutes les valeurs lissées sont rassemblées dans la série temporaire C_t (étape 2). Ensuite, un filtre consistant en l'application de plusieurs moyennes mobiles suivies d'un lissage par la méthode Loess est exécuté (étape 3), ce qui conduit à L_t . Enfin, les sous-séries lissées sont détendancialisées en soustrayant L_t de C_t , opération dont découle la composante de saisonnalité S_t (étape 4). Les données initiales sont ensuite désaisonnalisées (étape 5), et la composante tendancielle est obtenue par lissage de la série désaisonnalisée (étape 6). Par conséquent, trois opérations de lissage de type Loess sont incluses dans la procédure (étapes 2, 3 et 6). Si la série temporelle a deux composantes de saisonnalité ou plus, la procédure est répétée : les composantes sont estimées successivement (de la composante avec la période la plus courte à la composante avec la période la plus longue), soustraites, et la composante suivante est déterminée à partir des résidus. La composante tendancielle finale est alors calculée à la fin de toutes les itérations.

Cleveland et al. (1990) ont prescrit des valeurs pour presque tous les paramètres (nombres d'itérations et paramètres de lissage) de la procédure, sur la base des propriétés mathématiques de la décomposition. Il est possible de remarquer que tous les degrés polynomiaux impliqués dans les opérations de lissage sont égaux à 1 ; cela signifie que les composantes sont localement linéaires. Seul le choix de n_s , le paramètre de lissage de la composante de saisonnalité (c'est-à-dire le nombre d'observations consécutives à utiliser pour estimer C_t à l'étape 2), qui est intrinsèquement lié à n_t , le paramètre de lissage de la composante tendancielle, est moins direct. Ce paramètre contrôle la quantité de variations dans les données qui constituent la composante de saisonnalité. En d'autres termes, il détermine la vitesse à laquelle la composante peut changer. Les sous-séries du cycle peuvent être tracées pour décider de la rapidité avec laquelle les valeurs doivent évoluer d'un cycle à l'autre (plus n_s est élevé, plus l'évolution est lente et plus la courbe est lisse). Par conséquent, le choix de n_s est généralement basé sur des graphiques de diagnostic et/ou des connaissances préalables sur les données.

Voir Cleveland et al. (1990) pour plus de détails méthodologiques et Hyndman & Athanasopoulos (2018) pour des applications pratiques avec le logiciel statistique R (package *fpp2*), qui est également utilisé dans ce travail. Par ailleurs, Wang, X. et al. (2006) ont proposé un indicateur dont la valeur est comprise entre 0 et 1 pour mesurer la force de chaque composante :

$$F_T = \max_t \left(0; 1 - \frac{\text{Var}(R_t)}{\text{Var}(T_t + R_t)} \right) \text{ et } F_S = \max_t \left(0; 1 - \frac{\text{Var}(R_t)}{\text{Var}(S_t + R_t)} \right) \text{ où } \text{Var} \text{ est la variance}$$

La force de la composante est plus élevée lorsque F est plus proche de 1, c'est-à-dire lorsque les données désaisonnalisées ($T_t + R_t$) ou les données détendancialisées ($S_t + R_t$) varient beaucoup plus que la composante restante (R_t).

2.4.3.4 Projection et annualisation

Dans cette recherche, les composantes extraites des données passives avec la méthode STL sont utilisées pour estimer des courbes projetées et annualisées (en nombre total de déplacements quotidiens) par mode. La courbe projetée, définie à l'ensemble des t observations (jours) de 2013 à 2018, équivaut à la composante tendancielle normalisée par sa valeur moyenne sur la période de l'enquête OD 2013 (\bar{x} étant la notation de la moyenne), puis appliquée au nombre typique de déplacements calculé à partir de l'enquête OD 2013 :

$$Projection_{t(2013 \rightarrow 2018)}^{mode} = \frac{T_t(2013 \rightarrow 2018)}{\bar{T}_t(\text{période EOD 2013})} * \#déplacements_{EOD 2013}$$

La prévision du nombre de déplacements dans l'enquête OD 2018 est faite en prenant la moyenne des valeurs obtenues sur la période de l'enquête OD 2018. Cette méthode permet donc d'estimer une projection qui n'est pas affectée ni par la saisonnalité (ce qui est également le cas dans l'enquête OD), ni par les données aberrantes.

De même, l'annualisation pour une année spécifique (par exemple 2018) sans l'influence de la tendance se fait en appliquant la composante de saisonnalité (qui peut être une somme ou une multiplication de plusieurs composantes), normalisée par sa valeur moyenne durant la période de l'enquête OD 2018, au nombre typique de déplacements calculé à partir de l'enquête OD 2018. L'équation diffère entre la forme additive et la forme multiplicative, la première étant plus compliquée et nécessitant une normalisation supplémentaire pour pouvoir additionner des valeurs avec la même unité. Pour les deux formes, l'équation est formulée de manière que la moyenne de la courbe annualisée sur la période de l'enquête OD 2018 soit égale au nombre typique de déplacements observés dans l'enquête OD 2018.

$$\begin{aligned} \text{Annualisation [add]}_{t(2018)}^{mode} &= \#déplacements_{SEOD\ 2018} + (S_{t(2018)} - \overline{S_t(période\ EOD\ 2018)}) * \frac{\#déplacements_{SEOD2018}}{\overline{Y_t(période\ EOD\ 2018)}} \\ \text{Annualisation [mult]}_{t(2018)}^{mode} &= \frac{S_{t(2018)}}{\overline{S_t(période\ EOD\ 2018)}} * \#déplacements_{SEOD\ 2018} \end{aligned}$$

2.4.4 Résultats

2.4.4.1 Décomposition des données passives

La méthode STL est d'abord appliquée à chaque base de données passives. Les données ayant été agrégées au niveau quotidien, deux composantes de saisonnalité sont estimées : une pour la saisonnalité hebdomadaire et une pour la saisonnalité annuelle, avec respectivement $n_p = 7$ et $n_p = (365 * 5 + 366)/6 = 365,17$. Le calcul de la périodicité annuelle est lié à la période étudiée, composée de cinq années de 365 jours et d'une année (2016) de 366 jours. En effet, il n'est pas possible de supprimer le 29 février 2016 sans impacter la périodicité hebdomadaire. Le choix crucial de n_s est basé sur des graphiques de diagnostic (disponibles sur demande). Les changements d'une année à l'autre et d'une semaine à l'autre sont supposés être lents ; ainsi, n_s doit être élevé. De plus, une boucle itérative (testant successivement différentes valeurs impaires de n_s comprises entre 7 et 99) est exécutée afin de trouver la valeur optimale pour la validation de la procédure de la Figure 2-12, c'est-à-dire la valeur qui conduit à la meilleure projection de l'indicateur moyen pour l'automne 2018 (ou à l'erreur la plus faible). Les paramètres sélectionnés sont fournis dans le Tableau 2-6 pour chaque mode, chaque forme de décomposition et chaque composante de saisonnalité. Tous les autres paramètres sont déterminés à l'aide des valeurs prescrites. Une estimation robuste ($n_o = 15$) est réalisée car il peut y avoir des valeurs aberrantes dans les données passives, en particulier dans les comptages vélos et automobiles.

Tableau 2-6 Valeurs sélectionnées des paramètres de lissage des composantes de saisonnalité pour chaque mode et chaque forme de décomposition

Forme de décomposition		ADDITIVE		MULTIPLICATIVE	
Mode	Saisonnalité	Semaine	Année	Semaine	Année
MÉTRO		33	33	45	79
VÉLO		7	7	47	79
AUTOMOBILE		13	7	19	53

Les résultats de l'application de la méthode aux embarquements dans le métro sont présentés à la Figure 2-13, pour les deux formes de décomposition (additive et multiplicative). Les données brutes (série temporelle du nombre de validations par jour) sont illustrées dans la première ligne, et les graphiques suivants représentent les différentes composantes de la décomposition. Les deux décompositions étant très similaires, des commentaires communs sont faits ci-après.

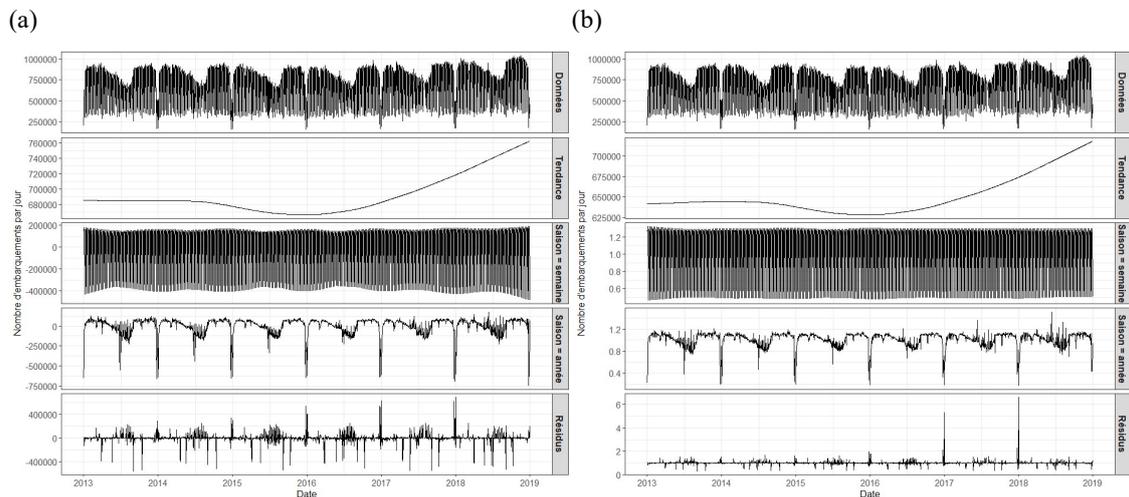


Figure 2-13 Décompositions STL (a) additive et (b) multiplicative des embarquements dans le métro

La composante tendancielle, tracée en deuxième position, indique une lente croissance de la demande quotidienne totale du métro, sauf en 2015. Cette courbe long-terme a été comparée à la courbe d'achalandage annuel publié par la STM dans ses rapports officiels et un profil similaire a été trouvé. Cependant, ici, la tendance est décrite avec une valeur par jour, et non une seule valeur par an. Par ailleurs, la tendance (et les composantes de saisonnalité) est disponible à tous les jours, même lorsque les données sont manquantes ; ceci est possible grâce aux propriétés de la méthode de lissage Loess.

Contrairement aux autres méthodes de décomposition, l'algorithme STL permet aux composantes de saisonnalité de changer dans le temps. Dans le cas d'application présenté, cela permet de capturer la variabilité d'un jour à l'autre, d'une semaine à l'autre et d'une année à l'autre. Cette flexibilité permet d'observer, pour le métro, que la différence entre les jours de semaine et les jours de fin de semaine est plus faible pendant l'été ; la diminution estivale de l'achalandage total est plutôt considérée dans la saisonnalité annuelle. Toutefois, la composante de saisonnalité annuelle est assez régulière d'une année à l'autre entre 2013 et 2018. Les différences les plus importantes sont dues aux jours fériés, qui ne sont pas positionnés au même moment (au même nombre de jours depuis le 1^{er} janvier) au fil des ans.

En ce qui concerne la dernière composante (reste), on observe une augmentation des résidus pendant l'été et les périodes de vacances. Cela pourrait s'expliquer par des facteurs externes temporaires, tels que des événements, activités spéciales ou les conditions météorologiques. Néanmoins, comme l'estimation est robuste, l'impact de ces cas spécifiques sur le calcul des autres composantes est réduit.

Les décompositions obtenues à partir des comptages vélos et automobiles sont exposées dans la Figure 2-14. Seule la décomposition multiplicative de chaque mode est présentée car cette forme donne la meilleure projection pour 2018 (voir section suivante). Pour les automobiles, la tendance

long-terme est à la baisse, avec une accélération en 2018. Les composantes de saisonnalité ont des points communs avec celles du métro, mais elles sont moins régulières. En effet, les plus petites valeurs de n_s dans le Tableau 2-6 indiquent que les changements sont plus rapides. L'utilité d'une estimation robuste est démontrée par les résidus plus élevés dus aux comptages inhabituels en 2015. Pour les vélos, la tendance est beaucoup moins claire. La saisonnalité hebdomadaire varie dans le temps, mais la saisonnalité annuelle est très prononcée.

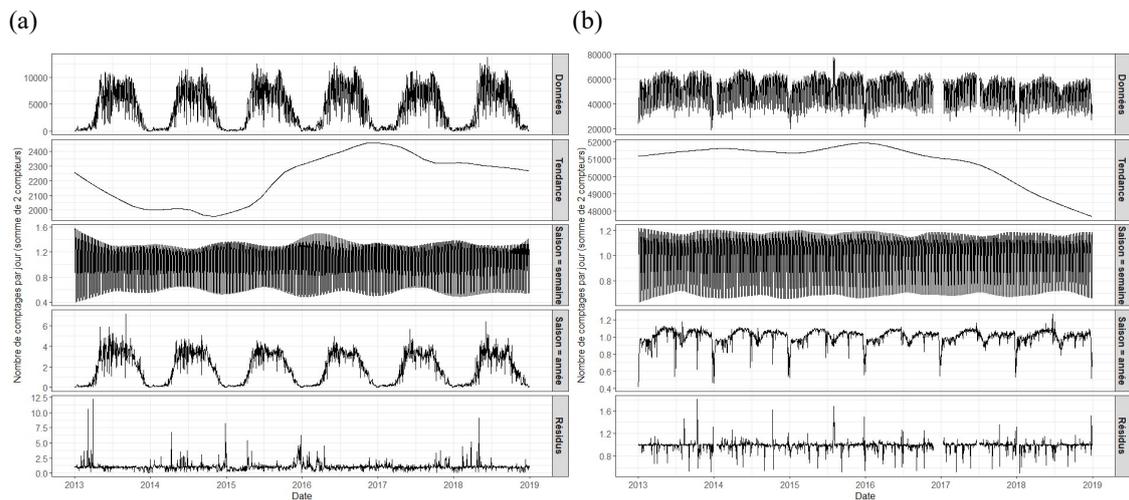


Figure 2-14 Décomposition STL multiplicative des comptages (a) de vélos et (b) d'automobiles privées

Cette dernière observation est confirmée par les indicateurs de force du Tableau 2-7 (pour la décomposition multiplicative uniquement, mais les résultats sont très similaires pour la forme additive). Ceux-ci révèlent que la plus forte saisonnalité annuelle est mesurée pour le vélo. La force des composantes de saisonnalité est analogue entre le métro et la voiture, et des rythmes hebdomadaires importants sont mis en évidence dans les deux cas. Cependant, la tendance est beaucoup plus faible pour tous les modes ; cela peut être dû à la longueur de la période d'étude (6 ans), qui n'est pas suffisante pour observer de grands changements à long terme.

Tableau 2-7 Force des composantes de tendance et de saisonnalité (indicateur F) dans la décomposition multiplicative des données de chaque mode

Composante	MÉTRO	VÉLO	AUTOMOBILE
Tendance	0.05	0.05	0.07
Saisonnalité hebdomadaire	0.84	0.36	0.81
Saisonnalité annuelle	0.47	0.89	0.50

2.4.4.2 Parts modales projetées et annualisées

En utilisant la tendance de la décomposition précédente pour chaque mode, le nombre moyen de déplacements projeté pour les jours de semaine d'automne 2018 est estimé et comparé au nombre observé dans l'enquête OD 2018. Les taux d'erreur obtenus entre la projection (avec choix optimal du paramètre n_s) et l'observation sont fournis dans le Tableau 2-8 par mode et par forme de décomposition. De plus, ces pourcentages sont comparés à ceux obtenus en appliquant la méthode traditionnelle (désagrégée) de projection de la demande du MTQ. Cette méthode

projette le facteur de pondération de chaque déplacement individuel en fonction de l'évolution anticipée des caractéristiques socio-démographiques. Un ajustement de type Furness est également effectué pour moduler la distribution spatiale des déplacements à motif travail. L'erreur rapportée dans le Tableau 2-8 correspond à la comparaison entre le nombre de déplacements calculés avec les facteurs de projection de 2016 (estimés à partir de l'enquête OD 2013) et le nombre observé dans l'enquête OD 2018 (pondéré avec le recensement de 2016).

Tableau 2-8 Taux d'erreur entre le nombre de déplacements projeté et observé pour les jours de semaine d'automne 2018

Méthode	MÉTRO	VÉLO	AUTOMOBILE
Méthode MTQ (facteurs de projection 2016)	1.8 %	4.5 %	8.6 %
Procédure de fusion – Forme additive	3.4e-04 %	11.1 %	6.5 %
Procédure de fusion – Forme additive multiplicative	1.8e-03 %	9.4e-04 %	6.4 %

Dans tous les cas (sauf un), les erreurs évaluées avec la procédure proposée sont plus petites qu'avec la méthode traditionnelle du MTQ. Par conséquent, la méthode développée fonctionne relativement bien. Cela n'était pas évident au départ en raison de toutes les disparités sémantiques, temporelles et spatiales décrites précédemment entre les données passives et les données d'enquête. Pour le métro, la différence est presque nulle, quelle que soit la forme de décomposition, alors que les erreurs sont plus importantes pour les vélos et les automobiles, mais la qualité des données de comptage est moindre et ces modes sont plus difficiles à prévoir. Néanmoins, la décomposition multiplicative semble plus performante, c'est pourquoi elle est choisie pour les analyses suivantes. Cette meilleure performance peut être due au fait que le logarithme permet de réduire l'influence des valeurs extrêmes et procure ainsi une distribution plus proche de la loi normale pour la régression locale.

Les courbes projetées et annualisées dérivées de la décomposition multiplicative des validations du métro (par application des composantes tendancielle ou saisonnières au nombre typique de déplacements estimé avec l'enquête OD) sont présentées dans la Figure 2-15. La ligne noire représente le nombre de déplacements en métro observé dans l'enquête OD de 2013 ou 2018.

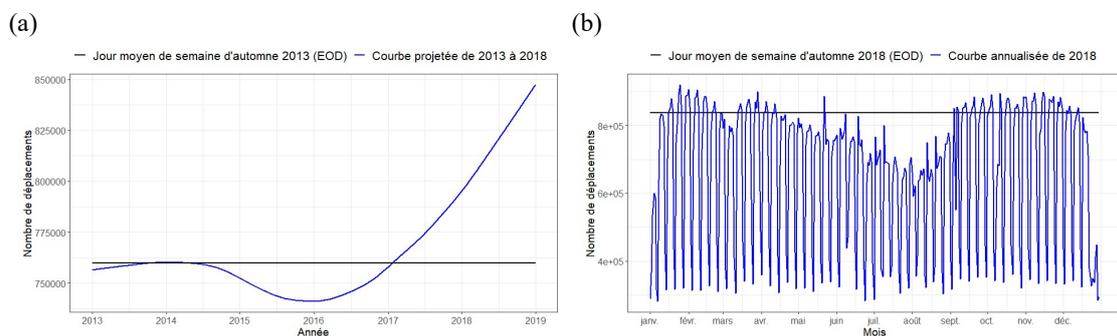


Figure 2-15 Courbes projetée (a) et annualisée (b) du nombre total de déplacements par jour dans le métro

Les mêmes courbes sont estimées pour chaque mode, puis les parts modales correspondantes (en proportion du total des déplacements pour les trois modes) sont calculées. Les parts modales projetées de 2013 à 2018 sont représentées à la Figure 2-16. Ainsi, plutôt que d’avoir seulement un point en automne 2013 et un point en automne 2018 (i.e., les parts modales quotidiennes typiques observées dans les deux enquêtes), la méthode permet un suivi longitudinal de ce qui s’est passé pendant l’intervalle de 5 ans. Les variations sont faibles, mais cela était attendu car les changements mesurés dans les comportements de mobilité agrégés sont généralement lents.

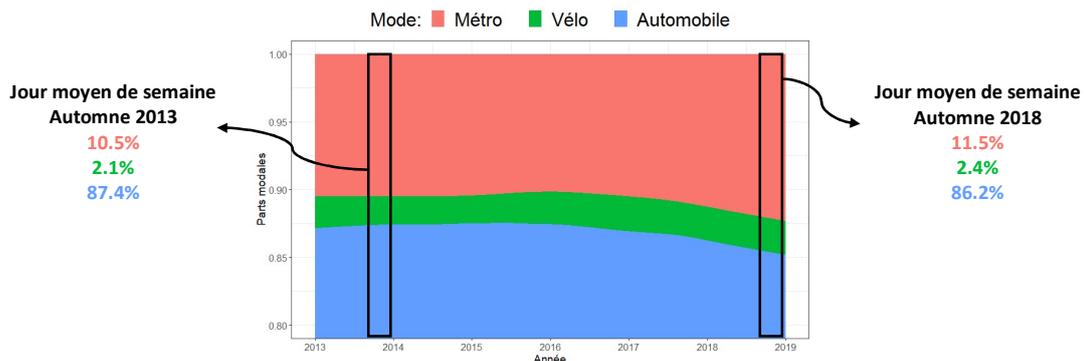


Figure 2-16 Projection des parts modales quotidiennes de trois modes de transport entre 2013 et 2018

Les parts modales annualisées (ou saisonnalisées) pour 2018, ainsi que la semaine moyenne d’automne 2018, sont illustrées à la Figure 2-17. Ainsi, au lieu d’une valeur moyenne pour toute l’année, la méthode permet d’obtenir un profil qui varie selon le jour de la semaine et le jour de l’année. En observant les variations par type de jour sur une semaine moyenne calculée sur la période de l’enquête OD 2018 plus les week-ends, on constate que les parts modales ne sont pas constantes, même du lundi au vendredi. Cela remet en question la représentativité du jour de semaine « typique » de l’enquête OD, déjà questionnée par Verreault et Morency (2011).

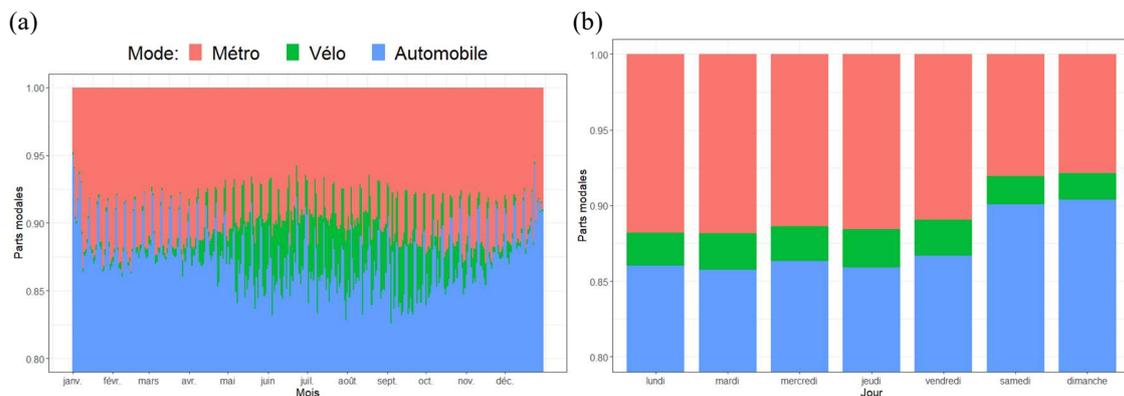


Figure 2-17 (a) Parts modales quotidiennes annualisées de trois modes en 2018 et (b) parts modales moyennes sur une semaine d’automne 2018

En outre, les variations des parts modales quotidiennes sélectionnées sur l’automne 2018 pour les jours de semaine uniquement (donc sans même considérer la variabilité entre les jours de semaine et de fin de semaine) sont décrites dans le Tableau 2-9. Les résultats prouvent qu’il y a beaucoup de variations autour de la moyenne, notamment pour le vélo.

Tableau 2-9 Variations des parts modales quotidiennes au cours de l'automne 2018 (jours de semaine uniquement)

MODE	Minimum	Moyenne	Maximum
MÉTRO	7.8 % (- 33 %)	11.5 %	13.0 % (+ 13 %)
VÉLO	0.2 % (- 91 %)	2.3 %	6.3 % (+ 167 %)
AUTOMOBILE	82.6 % (- 4 %)	86.2 %	90.1 % (+ 5 %)

2.4.5 Conclusion

Ce travail propose une méthode pour transmettre des informations temporelles depuis des données passives (donneur) vers une enquête ménage transversale constituée de carnets de déplacements d'une journée (récepteur). La combinaison de ces deux types de données, surtout dans ce sens (c'est-à-dire pour enrichir une enquête), a été peu explorée dans la littérature en raison du manque de variables communes entre les données. Cependant, dans cette recherche, l'agrégation des données au niveau quotidien a permis de créer des liens entre elles. La procédure développée, basée sur la décomposition STL de séries temporelles, consiste à appliquer les fluctuations longitudinales (tendance localement linéaire et saisonnalité) extraites des données passives à l'utilisation quotidienne typique observée dans l'enquête pour un mode de transport donné. Cette technique possède des propriétés intéressantes pour travailler avec des données passives : elle permet la présence de valeurs manquantes et est robuste aux valeurs aberrantes, pouvant ainsi compenser l'incomplétude et la faible qualité des données. En outre, les calculs sont simples et rapides, ce qui permet de traiter des ensembles de données massifs.

La méthode a été appliquée à l'enquête OD de Montréal, menée tous les cinq ans sur un jour typique de semaine d'automne, pour trois modes : le métro, le vélo et la voiture particulière. Alors qu'il n'était initialement possible que de calculer des nombres moyens de déplacements par mode pour l'automne 2013 et l'automne 2018 (à partir des enquêtes), la méthode a permis de suivre leur évolution tout au long de l'intervalle de cinq ans, ainsi que d'estimer un profil qui tient compte des variations annuelles et hebdomadaires. De cette manière, les parts modales, généralement disponibles de manière ponctuelle, peuvent devenir continues, et des informations à la fois multimodales et longitudinales peuvent donc être obtenues grâce à la complémentarité entre enquêtes traditionnelles et données émergentes. Dans cette perspective, le rôle des enquêtes traditionnelles serait de collecter des patrons moyens d'utilisation des modes et des informations contextuelles, tandis que les données passives les complèteraient en capturant la variabilité au cours de l'année et entre deux enquêtes. Un tel schéma pourrait rendre superflue la réalisation d'enquêtes longitudinales (au moins au niveau agrégé), qui sont sujettes à des enjeux économiques et techniques. Toutefois, les nouveaux services de mobilité étant encore sous-représentés dans les enquêtes de déplacements, davantage d'informations sur ces modes devront être collectées afin de pouvoir considérer à la fois les modes traditionnels et émergents.

Les travaux futurs viseront à tester la méthode proposée avec d'autres modes, tels que le bus ou le taxi, et à réaliser des analyses plus désagrégées : par station de métro (composante spatiale), par produit tarifaire ou pour des moments spécifiques (par exemple, les heures de pointe). Les variations évaluées autour de la moyenne grâce aux données passives pourraient également être exploitées pour développer des facteurs de variabilité. Ceux-ci permettraient de partiellement corriger les facteurs de pondération actuels de l'enquête OD et de tenir compte des variations

journalières ou hebdomadaires au cours de l'automne. L'idée serait de générer une distribution de valeurs plutôt qu'une valeur moyenne à partir des mêmes données. De plus, la modélisation des résidus de la décomposition de série temporelle, ou l'estimation d'une composante de régression supplémentaire dans l'algorithme STL (par ajustement pondéré itératif), serait intéressante pour inclure les effets de la météo, des vacances et des événements particuliers.

Pour aller encore plus loin, un modèle longitudinal pourrait être développé pour modéliser la série temporelle désaisonnalisée (plutôt que de simplement l'observer dans les données) et ensuite prévoir la tendance future. Dans l'exemple présenté, la tendance a été ajustée (interpolée) avec deux observations, l'une provenant de l'enquête OD 2013 et l'autre de l'enquête OD 2018, mais l'objectif ultime serait de projeter cette courbe au-delà de 2018 en utilisant les données passives pour prévoir le nombre de déplacements qui sera observé dans l'enquête OD de 2023. Cet exercice est encore plus pertinent que jamais dans le contexte de la pandémie de COVID-19. En effet, les changements qui impactent actuellement les comportements de mobilité rendent plus difficile la prédiction des parts modales qui seront observées dans 5 ans. Avec des données passives plus récentes, l'effet de la crise sanitaire sur la tendance et la saisonnalité pourrait être analysé. Le télétravail, largement adopté à Montréal depuis mars 2020, pourrait avoir un impact important sur les rythmes hebdomadaires. Différents scénarios de projection de la mobilité pourraient alors être proposés.

2.5 Méthode de synthétisation spatialisée de la population et de sa mobilité

- Par : Mohamed Khachman, doctorat
- Supervision : Pre Catherine Morency / Pr Francesco Ciari
- Contexte : Chaire Mobilité
- Fin prévue : automne 2022

Cette section est constituée d'extraits de la thèse de doctorat en cours de rédaction de Mohamed Khachman.

2.5.1 Introduction

La planification stratégique des transports requiert une modélisation réaliste des comportements de mobilité de la population dans le présent, mais aussi à différents horizons. Cela permet de tester des scénarios de réponse à la demande changeante afin d'identifier l'offre de mobilité la plus efficace dans laquelle investir. Les approches classiques basées sur le modèle à quatre étapes ayant montré leurs limites (Kitamura et al. 2000), les microsimulations à grande échelle à base d'agents couplées aux modèles basés activités, telles que UrbanSim (Waddell, 2002), ILUTE (Salvini & Miller, 2005), MATSim (Balmer et al., 2009) et SimMobility (Adnan et al. 2016) entre autres, deviennent de plus en plus populaires dans la modélisation et la prévision de la demande en transport.

Les caractéristiques sociodémographiques des individus et de leurs ménages ainsi que la localisation spatiale des ménages sont des facteurs déterminants des comportements de mobilité. Ainsi, pour des microsimulations fiables de ces comportements, une liste totalement énumérée, caractérisée et spatialisée de la population de ménages et d'individus est requise en intrant. Les instituts statistiques (e.g., Statistique Canada) disposent d'une telle liste composée du recueil des formulaires du recensement complétés, mais celle-ci demeure inaccessible aux chercheurs et planificateurs pour des raisons relevant de la protection de la vie privée. Toutefois, deux produits en sont dérivés et rendus publics : un échantillon désagrégé mais non spatialisé et des résumés du recensement constitués de totaux zonaux des attributs recensés. Pour ce qui est de l'échantillon, au Canada, il s'agit du Fichier de Micro-données à Grande Diffusion (FMGD) comprenant un échantillon généralement d'environ 1% de la population dont la seule information spatiale connue est la région métropolitaine d'appartenance. Les totaux zonaux sont quant à eux calculés pour des régions géographiques normalisées, e.g., aires de diffusion, secteurs de recensement, etc. Il s'agit de découpages géographiques listés dans la Classification Géographique Type (CGT) de Statistique Canada. N'ayant pas accès à la liste totalement énumérée, caractérisée et spatialisée des instituts statistiques, les planificateurs ont cherché à en synthétiser une en utilisant les données accessibles, i.e., l'échantillon désagrégé non spatialisé et les résumés du recensement, afin de pouvoir rouler leurs microsimulations. C'est ce processus inversé que l'on appelle « synthèse de population ». La Figure 2-18 illustre le processus allant de la complétion des formulaires du recensement à la population synthétique pour un territoire imaginaire δ .

La synthèse de population est typiquement un processus se servant de données agrégées, provenant des résumés de recensement, et de données partielles désagrégées relatives à un échantillon d'enquête, provenant souvent de FMGD, pour produire une population totalement énumérée d'individus et/ou de ménages avec des caractéristiques sociodémographiques assignées et une localisation spatiale plus ou moins précise. Le but est de produire une population synthétique, i.e., pour des individus non identifiables, ayant des caractéristiques statistiquement

cohérentes avec les données disponibles décrivant la population réelle donc présentant un comportement de mobilité semblable au sien.

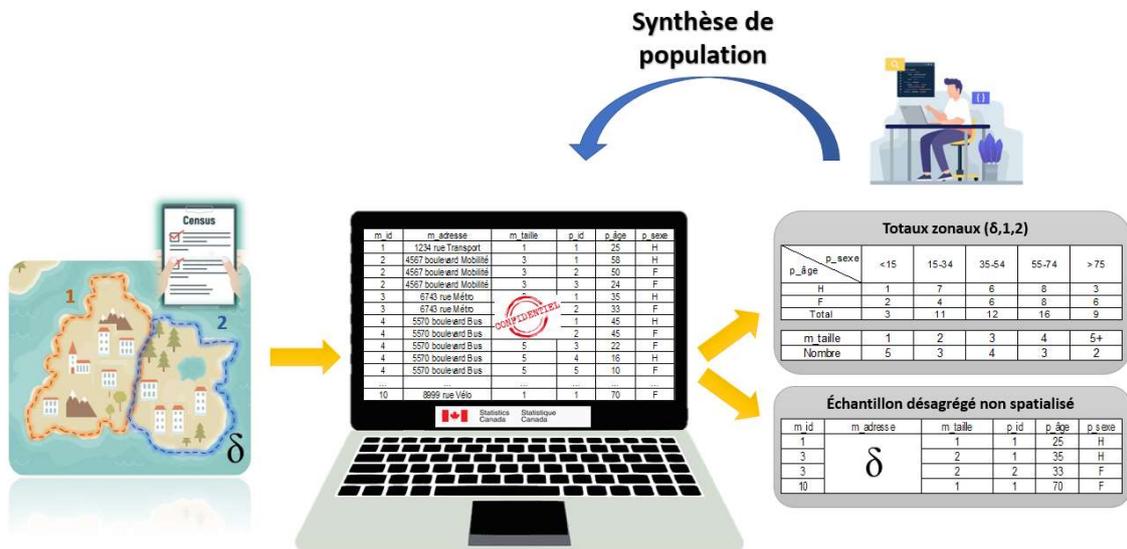


Figure 2-18. Des formulaires du recensement à la population synthétique

2.5.2 Structure du projet de recherche

Tel que mentionné précédemment, pour une modélisation fiable des comportements de mobilité de la population, une population synthétique de qualité, i.e., représentant fidèlement les caractéristiques de la population réelle, est requise. Ainsi, l'objectif principal du projet de recherche est de proposer une méthodologie de synthèse spatialisée de la population. Il s'agit d'une méthodologie permettant de produire une population synthétique optimale où les ménages synthétiques sont assignés à des bâtiments voire des unités de logement. Une application de cette méthodologie dans le contexte montréalais est aussi ciblée.

Pour atteindre l'objectif du projet, une structure en trois étapes, correspondant à trois contributions de recherche, est conçue.

- **Diagnostic** : il s'agit d'étudier les enjeux liés à la synthèse spatialisée de la population. Comme une population synthétique au niveau du bâtiment est convoitée, une étude de l'impact de la strate de référence sur la qualité de la population synthétique est particulièrement approfondie.
- **Synthèse spatialisée par ajustement** : les méthodes de synthèse par ajustement, semblables à des méthodes de pondération d'enquête, présentent les avantages d'être généralement simples à utiliser et à transférer. Ainsi, nous œuvrons dans cette étape à mettre en place une méthode de synthèse spatialisée par ajustement qui permet d'obtenir une population synthétique optimale au niveau du bâtiment.
- **Synthèse spatialisée probabiliste** : les méthodes probabilistes de synthèse, généralement moins faciles à opérer et transférer, présentent toutefois un potentiel important d'amélioration de la qualité de la synthèse spatialisée. Dans le contexte montréalais, celles-ci permettraient d'exploiter la richesse spatiale des données des enquêtes Origine-Destination (OD), où l'on dispose des coordonnées de chaque ménage

enquêté. Des modèles d'apprentissage automatique performants pourront ainsi être calibrés et appliqués à la synthèse spatialisée de la population.

Les sections suivantes sont basées sur le papier publié par Khachman et al. (2021) portant sur la première contribution du projet où l'impact de la strate de référence sur la qualité de synthèse est diagnostiqué.

2.5.3 Question de recherche

Comme indiqué dans l'introduction, les instituts statistiques publient les résumés du recensement pour plusieurs découpages géographiques qui peuvent chacun constituer la strate de référence pour l'exercice de synthèse. Pour clarifier l'idée de la strate de référence, considérons l'exemple du secteur δ (Figure 2-18). En effet, pour ce secteur nous disposons des totaux zonaux pour l'ensemble du territoire δ (e.g., nombre d'hommes sur δ) mais aussi pour les secteurs 1 et 2 séparément (e.g., nombre d'hommes dans le secteur 1 et nombre d'hommes dans le secteur 2). Ainsi, pour synthétiser la population du territoire δ , il existe deux choix. Le premier est de synthétiser directement l'ensemble de la population de δ en utilisant les résumés du recensement au niveau de δ . Cela revient à considérer δ comme strate de référence. Le deuxième est de synthétiser les populations des secteurs 1 et 2 qui composent δ séparément en utilisant les résumés du recensement relatifs à chacun. Cela revient à considérer les secteurs comme strate de référence. Compte tenu de la richesse de la CGT de Statistique Canada, ce choix est beaucoup plus varié en réalité. Par conséquent, tout comme le choix des variables de contrôle, i.e., des variables dont les distributions dans la population réelle doivent être reproduites dans la population synthétique, le choix de la strate de référence constitue un prérequis à l'opération de synthèse. Dans la suite nous appelons une zone de référence (ZdR) une unité géographique de la strate de référence. Par exemple, si l'on considère les secteurs comme strate de référence, celle-ci contient deux ZdR (secteur 1 et secteur 2).

En fait plusieurs enjeux sont liés à l'agrégation de la strate de référence. D'une part, plus la strate de référence est agrégée, plus les erreurs de spatialisation seraient importantes. Cela serait essentiellement dû à l'hypothèse d'uniformité spatiale sous-jacente (ou d'homogénéité des niveaux moins agrégés) que l'on émet en choisissant une strate de référence. Pour clarifier cette idée, considérons de nouveau l'exemple δ . Si l'on génère une population synthétique de 6 jeunes hommes et 6 femmes âgées en utilisant δ comme strate de référence, nous ne disposons d'aucune information sur l'appartenance de ces individus synthétiques aux secteurs 1 et 2. Ainsi, nous supposons implicitement qu'ils sont uniformément distribués sur les deux secteurs (Figure 2-19a). Cependant, la distribution réelle pourrait être différente avec une concentration de jeunes hommes dans le secteur 1 et de femmes âgées dans le secteur 2 (Figure 2-19b) ce qui générerait des comportements de mobilité drastiquement différents de ceux d'une distribution uniforme. Ainsi, plus la strate de référence est agrégée, plus l'hypothèse d'uniformité spatiale émise est forte, plus les erreurs de spatialisation seraient importantes, et plus les simulations alimentées par une telle population synthétique seraient altérées.

D'autre part, plus l'on désagrège la strate de référence, plus les erreurs d'ajustement seraient importantes, i.e., moins l'algorithme serait capable de reproduire correctement la population de δ . Trois facteurs expliqueraient cette supposition : 1) les combinaisons d'attributs non observées, 2) les totaux nuls arrondis et 3) la complexité du calcul. En effet, aux niveaux les plus désagrégés un échantillon a plus de chance de manquer d'observations d'individus et/ou ménages qui existent dans la population réelle, ce qui serait susceptible d'empêcher la convergence de l'algorithme de synthèse. De plus, comme les effectifs recensés deviennent plus faibles lorsque

l'on désagrège la strate de référence, les totaux arrondis à zéro deviennent plus fréquents pour empêcher l'identification des individus et/ou ménages recensés. Enfin, la désagrégation de la strate de référence augmente la complexité du processus de synthèse.



Figure 2-19. (a) Population synthétique de δ en utilisant δ comme strate de référence ; (b) Population recensée de δ

Par exemple, pour le territoire δ , si l'on synthétise au niveau des secteurs, il va falloir tourner l'algorithme deux fois (pour les secteurs 1 et 2 séparément) pour synthétiser deux populations en essayant de respecter deux ensembles de contraintes distincts. Or, pour une synthèse directe au niveau de δ , l'algorithme tourne une seule fois pour synthétiser une seule population selon un seul ensemble de contraintes. Le facteur de 2 entre les secteurs et δ est en réalité de milliers si l'on compare une synthèse au niveau des aires de diffusion à une synthèse au niveau d'une région métropolitaine. Ainsi, on risquerait d'amplifier les erreurs intrinsèquement liées à la méthodologie de synthèse des milliers de fois en désagrégant la strate de référence. Le Tableau 2-10 résume les avantages, inconvénients et impacts présumés de l'agrégation et de la désagrégation de la strate de référence.

Tableau 2-10 Impacts potentiels de l'agrégation de la strate de référence sur la synthèse de population

Agrégation de la strate de référence	Avantages	Inconvénients	Impacts sur la population synthétique
Plus agrégée	<ul style="list-style-type: none"> • Moins de combinaisons d'attributs manquantes • Moins de totaux arrondis à zéro • Moins de contraintes d'ajustement 	<ul style="list-style-type: none"> • Hypothèse d'homogénéité (distribution spatiale uniforme) plus forte 	<ul style="list-style-type: none"> • Moins d'erreurs potentielles d'ajustement • Plus d'erreurs potentielles de spatialisation
Moins agrégée	<ul style="list-style-type: none"> • Hypothèse d'homogénéité (distribution spatiale uniforme) plus faible 	<ul style="list-style-type: none"> • Plus de combinaisons d'attributs manquantes • Plus de totaux arrondis à zéro • Plus de contraintes d'ajustement 	<ul style="list-style-type: none"> • Plus d'erreurs potentielles d'ajustement • Moins d'erreurs potentielles de spatialisation

Comme nous supposons qu'agréger la strate de référence augmenterait les erreurs de spatialisation alors que la désagréger augmenterait les erreurs d'ajustement, des travaux sont menés pour évaluer l'impact de la strate de référence (ou de la résolution géographique) sur la qualité de la synthèse de population.

2.5.4 Contributions

Les contributions des travaux menés sont les suivantes : 1) introduire la quantification des erreurs de spatialisation dans l'évaluation de la qualité de la synthèse (jusque-là généralement assimilée à la qualité de l'ajustement), 2) quantifier les erreurs d'ajustement et de spatialisation et en évaluer les variations en fonction de la strate de référence, et 3) définir et évaluer des mesures potentielles d'atténuation des erreurs telles que le double contrôle et l'harmonisation. Ultiment, cela a permis de déterminer une bonne pratique pour ce qui a trait au choix de la strate de référence et des mesures possibles pour réduire l'erreur totale (i.e., la somme des erreurs d'ajustement et de spatialisation).

2.5.5 Revue de littérature

Plusieurs approches de synthèse de population basées sur différents principes sont documentées dans la littérature. Les méthodes de synthèse de population peuvent être séparées en trois grandes familles : les méthodes de synthèse par ajustement, l'optimisation combinatoire et les méthodes de synthèse par simulation ou méthodes probabilistes. Dans cette section, un aperçu des algorithmes de synthèse de population multi-niveau (i.e., pouvant contrôler des variables pour plusieurs types d'agents) et multi-résolution (i.e., pouvant contrôler des variables à différentes résolutions géographiques) est donné. L'évolution depuis l'ajustement proportionnel itératif (Iterative Proportional Fitting - IPF) à la mise à jour proportionnelle itérative renforcée (enhanced Iterative Proportional Updating - eIPU) est notamment exposée, puisque l'eIPU a été choisie pour mener les expérimentations permettant de répondre à la question de recherche.

2.5.5.1 Méthodes de synthèse de population par ajustement

Les méthodes de synthèse par ajustement ont comme objectif principal d'ajuster les totaux des variables de contrôle dans l'échantillon à celles recensées. Elles englobent les méthodes basées sur l'IPF et les méthodes dites d'optimisation combinatoire (Combinatorial Optimization – CO).

Méthodes de synthèse basées sur l'IPF

Principe de l'ajustement proportionnel itératif (IPF)

L'IPF – aussi appelée approche conventionnelle (Guo & Bhat, 2007) – a d'abord été utilisée dans la modélisation des systèmes de transport pour synthétiser des populations de ménages (Beckman et al., 1996). L'approche standard de synthèse de population basée sur l'IPF comporte deux étapes : l'estimation et la sélection (Templ et al., 2017). Dans l'étape d'estimation, la distribution conjointe des variables de contrôle est estimée par ajustement itératif des fréquences des différentes caractéristiques de l'échantillon sur leurs totaux respectifs extraits du recensement. Les fréquences des combinaisons de caractéristiques de l'échantillon sont inscrites dans les cases d'une matrice et leurs totaux respectifs constituent les marges d'ajustement. Une fois l'effectif connu dans la population synthétique pour chaque combinaison de caractéristiques, l'étape de sélection commence. Pour chacune des cases ajustées – correspondant à une combinaison de caractéristiques – chacun des ménages de l'échantillon se voit attribuer une probabilité de sélection proportionnelle à la similarité existante entre ses caractéristiques et celles de la case. Selon les probabilités calculées, des ménages sont sélectionnés de l'échantillon

jusqu'à atteindre les effectifs des cellules ajustées et une population synthétique de ménages est par conséquent obtenue. Les membres assignés à ces ménages dans l'échantillon forment une population synthétique d'individus mais celle-ci ne respecte pas les totaux du recensement pour les individus puisque l'ajustement a été effectué uniquement au niveau des ménages. L'étape de sélection peut être opérée suivant une fonction traduisant le degré de divergence ou de convergence entre les combinaisons cibles et celles de l'échantillon (Beckman et al., 1996) ou aléatoirement par une série de simulations de Monte Carlo. Des populations synthétiques variées peuvent ainsi être produites respectant toutes les contraintes imposées.

Méthodes multi-niveau

L'**IPU** (Ye et al., 2009) est un algorithme heuristique de synthèse de population implémenté dans PopGen permettant de surmonter quelques limitations de l'IPF. Alors que l'IPF se contente d'ajuster la distribution à un seul niveau, souvent celui des ménages, l'IPU ajuste les distributions des ménages et des individus simultanément en recalculant les facteurs de pondération attribués aux ménages selon leur composition en individus. Ainsi, une fois l'ajustement selon les contraintes des ménages terminé, les poids sont ajustés selon les types de personnes de manière que seuls les facteurs de pondération des ménages comportant des individus d'un certain type soient modifiés lorsque l'ajustement se fait pour ce type.

À chaque itération, un indicateur de la qualité d'ajustement ζ , consistant en la différence en valeur absolue entre la somme pondérée et la contrainte, est calculé pour chacune des catégories. La valeur de ζ ne décroît pas constamment au cours des itérations. Ainsi, à chaque itération, les poids correspondant à la valeur minimale de ζ sont conservés. La différence en valeur absolue entre les valeurs de l'indicateur ζ avant le début de l'itération (ζ_a) et après la fin de celle-ci (ζ_b) indique le gain en termes de qualité d'ajustement pour chacune des itérations. L'algorithme converge lorsque la valeur absolue de cette différence est en-dessous d'un seuil ε prédéfini donc lorsque le gain ou la variation des poids calculés par rapport au poids minimal retenu demeure minime.

Lors de la sélection, la probabilité d'un ménage d'être choisi vaut son poids divisé par la somme des poids des ménages du même type. Le nombre ainsi obtenu présente le nombre de ménages du même type avec une composition identique d'individus.

Il existe d'autres algorithmes de synthèse multi-niveau. L'**optimisation de l'entropie** introduite par Bar-Gera et al. (2009) assure le respect des contraintes aux niveaux des individus et des ménages en fournissant une solution où l'apport de nouvelles informations demeure minimal. L'idée est de garder une distribution de poids la plus proche possible de la distribution initiale tout au long du processus d'ajustement. L'**IPF hiérarchique** (Müller & Axhausen, 2011) assure un ajustement multi-niveau en alternant les ajustements des distributions de ménages et d'individus. **PopH**, une méthode heuristique proposée par Zhuge et al. (2017), fournit une population synthétique par ajustement multi-niveau en minimisant la moyenne des différences entre les totaux générés et les contraintes marginales imposées, mais aussi les écarts-types de ces différences. Fournier et al. (2021) ont récemment introduit un algorithme résolvant le problème de synthétisation de la population au niveau des ménages et des individus simultanément sous forme d'**optimisation matricielle**.

Méthodes multi-résolution

Konduri et al. (2016) ont développé une version avancée de l'IPU, l'**eIPU**, capable de considérer les contraintes aux niveaux des individus et des ménages à deux échelles géographiques simultanément démontrant ainsi une amélioration de la qualité de la population synthétique par

rapport à l'IPU. L'algorithme IPU renforcé est implémenté dans PopGen 2.0. Une autre extension de l'IPU permettant de contrôler simultanément des variables à trois résolutions géographiques a été élaborée par Moreno et Moeckel (2018).

Optimisation combinatoire (CO)

Williamson et al. (1998) proposent l'optimisation combinatoire comme alternative à l'ajustement proportionnel itératif. La méthode est opérée en sélectionnant d'abord le nombre d'agents synthétiques (e.g., ménages) requis. Ensuite, chaque ménage initialement sélectionné est itérativement remplacé par les ménages de l'échantillon et un indicateur de qualité est calculé. Si le remplaçant contribue à l'amélioration de l'ajustement, il est gardé. Sinon, l'itération n'est pas considérée et une autre itération sélectionnant un autre ménage de l'échantillon est lancée à partir de l'état précédent de la population synthétique. Lorsque l'indicateur de divergence entre la population synthétique et les données agrégées décrivant la population réelle prend une valeur inférieure ou égale au seuil préfixé, la convergence est atteinte. L'optimisation combinatoire peut être opérée selon une variété de procédures : méthode de descente (hill climbing), recuit simulé (simulated annealing) ou en utilisant un algorithme génétique (Williamson et al., 1998).

2.5.5.2 Méthodes de synthèse de population par simulation

Les méthodes probabilistes visent à synthétiser une population en échantillonnant à partir d'une estimation de la distribution conjointe des variables contrôlées. Farooq et al. (2013) proposent une approche basée sur des simulations de Monte Carlo par Chaînes de Markov (Markov Chain Monte Carlo – **MCMC**) qui consiste à établir une estimation de la distribution conjointe des variables de contrôle à partir de différentes sources de données agrégées et désagrégées disponibles. En d'autres mots, ils proposent d'établir les probabilités conditionnelles complètes de chacune des catégories des variables en fonction des catégories des autres variables considérées. Ensuite, un échantillonneur de Gibbs est utilisé pour générer une population synthétique basée sur l'estimation établie de la distribution conjointe. Casati et al. (2015) introduisent le **MCMC hiérarchique**. Il s'agit d'une extension de l'approche MCMC à la synthèse multi-niveau moyennant une typologie d'individus intégrée dans le calcul des probabilités conditionnelles. Saadi et al. (2016) utilisent un modèle de Markov caché (Hidden Markov Model - **HMM**) dans l'exercice de synthèse de population.

Sun et Erath (2015) proposent d'utiliser un **réseau bayésien** pour établir une estimation de la distribution conjointe des variables de contrôle dans la population réelle. Le réseau bayésien est capable d'appréhender la structure de corrélation de la population à partir d'échantillons de données limités et par suite, de produire une estimation fiable de la distribution conjointe des variables considérées. L'utilisation du réseau bayésien évite alors à l'utilisateur le défi de définition des distributions conditionnelles des variables de contrôle ainsi que le piège d'un éventuel surajustement aux propriétés d'un échantillon utilisé (Sun & Erath, 2015). Sun et al. (2018) présentent un **réseau bayésien hiérarchique** permettant d'optimiser simultanément la qualité de la population aux niveaux des ménages et des individus.

Récemment, des méthodes relevant de l'apprentissage génératif profond telles que l'auto-encodeur variationnel (Variational Auto-Encoder - **VAE**) (Borysov et al., 2019), l'auto-encodeur variationnel conditionnel (Conditional Variational Auto-Encoder - **CVAE**) (Borysov et Rich, 2021), le réseau antagoniste génératif de Wasserstein (Wasserstein Generative Adversarial Network - **WGAN**) (Garrido et al., 2020), et le réseau antagoniste génératif conditionnel (Conditional Generative Adversarial Network - **CGAN**) (Johnsen et al., 2022), ont été adaptées à l'exercice de synthèse de population.

2.5.6 Méthodologie

Pour répondre à la question de recherche, des expérimentations sont menées sur les régions métropolitaines de Montréal, Toronto et Vancouver. Ainsi, les tendances peuvent être comparées et des conclusions généralisables émises.

2.5.6.1 Variables de contrôle

Les variables de contrôle considérées sont résumées dans le Tableau 2-11.

Tableau 2-11 Variables de contrôle

Variable	Définition	Catégories	Description
ppcount	Nombre de personnes	1	1 personne
		1	[0, 15]
ppage	Âge	2	[15, 65]
		3	65+
		1	Homme
ppsex	Sexe	2	Femme
		1	Jamais marié
ppmarst	Statut marital	2	Marié
		3	Vivant en union libre
		4	Séparé, divorcé ou veuf
		1	1 ménage
hhcount	Nombre de ménages	1	1 personne
		2	2 personnes
hhsize	Taille du ménage	3	3 personnes
		4	4 personnes
		5	5 personnes ou plus
		1	Sans famille de recensement
		2	1 famille de recensement sans enfants
hhtype	Type du ménage	3	1 famille de recensement avec enfants
		4	Plusieurs familles de recensement
		1	Moins de 30k\$
hhnetinc	Revenu après taxes du ménage	2	[30k\$, 60k\$]
		3	[60k\$, 100k\$]
		4	100k\$ ou plus

2.5.6.2 Système de zonage

Cinq strates ou résolutions géographiques sont considérées : la Région Métropolitaine de Recensement (RMR) (Figure 2-20), les Subdivisions de Recensement (SdR) (Figure 2-21), les Aires de Diffusion Agrégées (ADA) (Figure 2-22), les Secteurs de Recensement (SR) (Figure 2-23), et les Aires de Diffusion (AD) (Figure 2-24).

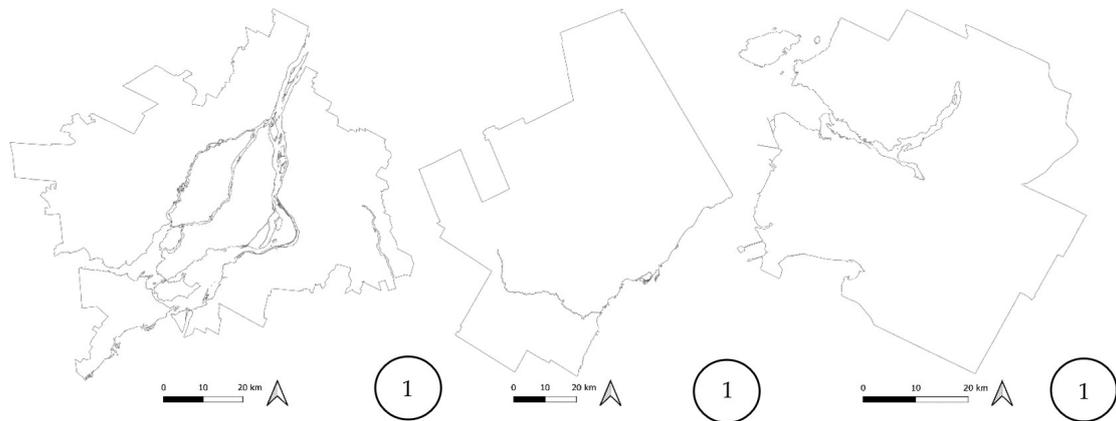


Figure 2-20. Limites des RMR de Montréal, Toronto et Vancouver (de gauche à droite)

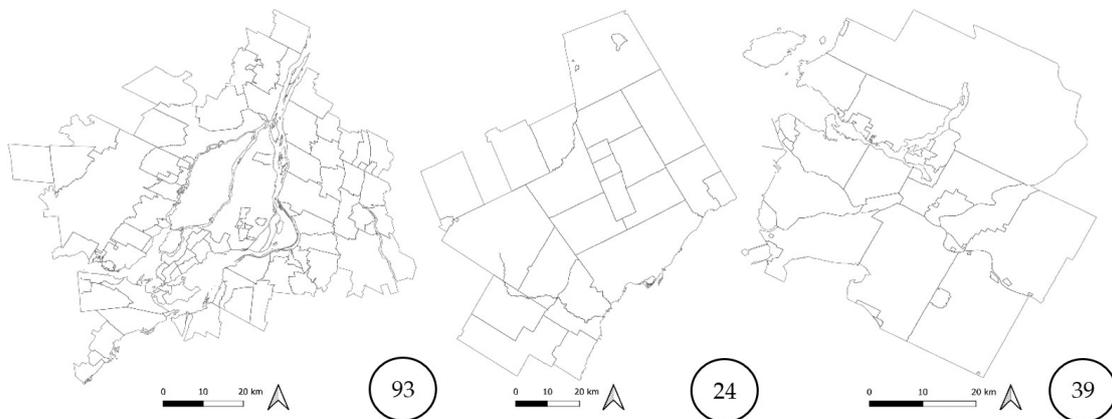


Figure 2-21. Limites des SdR de Montréal, Toronto et Vancouver (de gauche à droite)



Figure 2-22. Limites des ADA de Montréal, Toronto et Vancouver (de gauche à droite)



Figure 2-23. Limites des SR de Montréal, Toronto et Vancouver (de gauche à droite)



Figure 2-24. Limites des AD de Montréal, Toronto et Vancouver (de gauche à droite)

2.5.6.3 Scénarios

Les données utilisées proviennent du recensement canadien 2016. Les FMGD de Montréal, Toronto et Vancouver ont servi d'échantillons désagrégés. Pour ce qui est des données agrégées, nous avons extrait les totaux du recensement des trois RMR pour les variables de contrôle et résolutions géographiques considérées.

Incohérences du recensement

Les totaux du recensement présentent deux types d'incohérences. Les **incohérences intra-résolution** qui relèvent des différences entre les totaux des catégories de variables. Par exemple, la somme des hommes et des femmes d'une RMR est différente de la somme des effectifs des catégories d'âge de la même RMR. Les **incohérences inter-résolution** correspondent quant à elles aux différences entre l'effectif d'une catégorie de variable dans une zone géographique et la somme des effectifs de cette même catégorie dans les zones géographiques moins agrégées qui y sont incluses. Par exemple, la somme des hommes dans une RMR est différente de la somme des hommes des aires de diffusion qui composent cette RMR. Une méthodologie d'harmonisation basée sur l'IPF est proposée pour remédier à ces incohérences (Khachman et al., 2021).

Double contrôle

Le double contrôle, ou le contrôle simultané de deux résolutions géographiques, peut être appliqué pour deux raisons : 1) le besoin de contrôler des variables qui ne sont disponibles qu'à

des résolutions géographiques différentes, et 2) le besoin de représenter l'hétérogénéité spatiale de la population. Comme la désagrégation de la strate de référence est supposée améliorer la précision spatiale de la population synthétique, et l'agrégation de la strate de référence est supposée améliorer l'ajustement de celle-ci à la population réelle, des scénarios impliquant un double contrôle sont testés. L'idée est d'ajouter un contrôle au niveau de la strate la plus agrégée (RMR) tout en variant la strate de référence (qui sera alors moins agrégée) pour tester si cela permettrait d'optimiser à la fois l'ajustement et la précision spatiale de la population synthétique.

Scénarios

Ainsi, l'eIPU est utilisé pour synthétiser des populations pour les 3 RMR en considérant cinq strates de références, deux types de données (brutes et harmonisées) et deux types de contrôle géographique (simple et double). Le Tableau 2-12 récapitule les scénarios de synthèse exécutés pour chacune des RMR. Lorsque deux résolutions géographiques sont contrôlées, le terme « RÉGION » réfère à la strate la plus agrégée, et le terme « GÉO » à la strate la moins agrégée, en l'occurrence, la strate de référence.

Tableau 2-12 Scénarios

Scénario	Type de données	Résolutions contrôlées	RÉGION	GÉO
1	Brutes	1	-	RMR
2	Brutes	1	-	SDR
3	Brutes	2	RMR	SDR
4	Brutes	1	-	ADA
5	Brutes	2	RMR	ADA
6	Brutes	1	-	SR
7	Brutes	2	RMR	SR
8	Brutes	1	-	AD
9	Brutes	2	RMR	AD
10	Harmonisées	1	-	RMR
11	Harmonisées	1	-	SDR
12	Harmonisées	2	RMR	SDR
13	Harmonisées	1	-	ADA
14	Harmonisées	2	RMR	ADA
15	Harmonisées	1	-	SR
16	Harmonisées	2	RMR	SR
17	Harmonisées	1	-	AD
18	Harmonisées	2	RMR	AD

2.5.6.4 Indicateurs

Dans la littérature, les populations synthétiques sont généralement évaluées par la qualité de l'ajustement au niveau le plus agrégé. Par exemple, si une population synthétique est générée au niveau des AD, la somme des hommes des AD est calculée et comparée au nombre d'hommes au niveau de la RMR pour évaluer la qualité de synthèse, et ainsi de suite pour le reste des catégories de variables. Or, pour autant que nous sachions, la précision spatiale de la population synthétique, qui demeure un facteur important de la fiabilité des simulations basées sur celle-ci, n'a jamais été évaluée. Ainsi, l'une des contributions des travaux menés est la définition et l'intégration des erreurs de spatialisation dans l'évaluation de la population synthétique. Au total, trois indicateurs sont calculés pour comparer les scénarios de synthèse : les incohérences inter-résolution (α), les erreurs d'ajustement (β), et les erreurs de spatialisation (γ).

Incohérences inter-résolution du recensement (α)

Les incohérences inter-résolution reflètent les écarts entre la somme des effectifs recensés pour une catégorie de variable dans les ZdR et l'effectif recensé pour cette même catégorie au niveau le plus agrégé (RMR). Les effectifs recensés au niveau agrégé étant les plus fiables, tel qu'expliqué précédemment, ce niveau est considéré comme référence pour le calcul des erreurs α selon la formule suivante :

$$\alpha(\%) = \frac{\sum_{i=1}^m \left| \sum_{j=1}^n M_{i,ZdR_j} - M_{i,RMR} \right|}{hhcount_{RMR} + ppcount_{RMR}} \times 1000$$

Où

- 1 i réfère à la $i^{\text{ème}}$ catégorie de variable $\forall i = 1 \dots m$;
- 2 j réfère à la $j^{\text{ème}}$ ZdR $\forall j = 1 \dots n$;
- 3 M_{i,ZdR_j} réfère à l'effectif recensé pour la $i^{\text{ème}}$ catégorie de variable dans la $j^{\text{ème}}$ ZdR;
- 4 $M_{i,RMR}$ réfère à l'effectif recensé pour la $i^{\text{ème}}$ catégorie de variable dans la RMR;
- 5 $hhcount_{RMR}$ réfère au nombre de ménages dans la RMR;
- 6 $ppcount_{RMR}$ réfère au nombre d'individus dans la RMR.

Erreurs d'ajustement (β)

Les erreurs d'ajustement reflètent l'écart entre la population synthétique et le recensement au niveau le plus agrégé (RMR). Les effectifs recensés au niveau agrégé étant les plus fiables, tel qu'expliqué précédemment, ce niveau est considéré comme référence pour le calcul des erreurs β selon la formule suivante :

$$\beta(\%) = \frac{\sum_{i=1}^m \left| \sum_{j=1}^n m_{i,ZdR_j} - M_{i,RMR} \right|}{hhcount_{RMR} + ppcount_{RMR}} \times 1000$$

Où

- i réfère à la $i^{\text{ème}}$ catégorie de variable $\forall i = 1 \dots m$;
- j réfère à la $j^{\text{ème}}$ ZdR $\forall j = 1 \dots n$;
- m_{i,ZdR_j} réfère à l'effectif synthétisé pour la $i^{\text{ème}}$ catégorie de variable dans la $j^{\text{ème}}$ ZdR;
- $M_{i,RMR}$ réfère à l'effectif recensé pour la $i^{\text{ème}}$ catégorie de variable dans la RMR;
- $hhcount_{RMR}$ réfère au nombre de ménages dans la RMR;
- $ppcount_{RMR}$ réfère au nombre d'individus dans la RMR.

Erreurs de spatialisation (γ)

Les erreurs de spatialisation sont calculées à la résolution géographique la plus désagrégée puisque celle-ci constitue l'information spatiale la plus détaillée disponible. Ainsi, pour calculer les erreurs de spatialisation, la population synthétique est d'abord interpolée au niveau des AD depuis la strate de référence proportionnellement au nombre de ménages dans chaque AD (puisque le ménage est l'agent principal de calibration de l'eIPU). L'interpolation est exécutée au niveau de chacune des ZdR selon la formule suivante :

$$im_{i,AD_j} = m_{i,ZdR} \times \frac{hhcount_{AD_j}}{hhcount_{ZdR}}$$

Où

- i réfère à la $i^{\text{ème}}$ catégorie de variable;
- j réfère à la j^{am} AD incluse dans une ZdR;

- $m_{i,ZdR}$ réfère à l'effectif synthétisé pour la $i^{\text{ème}}$ catégorie dans une ZdR;
- im_{i,AD_j} réfère à fréquence synthétisée interpolée de la $i^{\text{ème}}$ catégorie de variable dans la $j^{\text{ème}}$ AD;
- $hhcount_{AD_j}$ réfère au nombre de ménages dans la $j^{\text{ème}}$ AD;
- $hhcount_{ZdR}$ réfère au nombre de ménages dans la ZdR.

Une fois les fréquences synthétisées interpolées calculées, les erreurs de spatialisation sont calculées par comparaison de celles-ci aux fréquences observées dans le recensement au niveau des AD selon la formule suivante :

$$\gamma(\%) = \frac{\sum_{i=1}^m \sum_{j=1}^n |im_{i,AD_j} - M_{i,AD_j}|}{hhcount_{RMR} + ppcount_{RMR}} \times 1000$$

Où

i réfère à la $i^{\text{ème}}$ catégorie de variable $\forall i = 1 \dots m$;

j réfère à la $j^{\text{ème}}$ ZdR $\forall j = 1 \dots n$;

im_{i,AD_j} réfère à l'effectif synthétisé interpolé pour la $i^{\text{ème}}$ catégorie de variable dans la $j^{\text{ème}}$ AD;

M_{i,AD_j} réfère à l'effectif recensé pour la $i^{\text{ème}}$ catégorie de variable dans la $j^{\text{ème}}$ AD;

$hhcount_{RMR}$ réfère au nombre de ménages dans la RMR;

$ppcount_{RMR}$ réfère au nombre d'individus dans la RMR.

2.5.7 Résultats

Différents groupes de scénarios sont comparés dans cette section pour évaluer les impacts (1) de l'agrégation de la strate de référence, (2) des incohérences inter-résolution du recensement, et (3) du contrôle additionnel au niveau de la RMR sur la qualité de la population synthétique. La qualité de la population synthétique englobe son ajustement aux données du recensement au niveau de la RMR, et sa précision spatiale qui reflète son respect des totaux recensés au niveau des AD.

2.5.7.1 Variations de α , β et γ en fonction de la strate de référence

Les variations des erreurs α , β et γ en fonction de la strate de référence sont d'abord évaluées pour les scénarios utilisant les données brutes du recensement et contrôlant une seule résolution géographique. Les trois RMR montrent généralement les mêmes tendances.

Sans surprise, la Figure 2-25 montre que les incohérences inter-résolution α augmentent lorsque la strate de référence est désagrégée puisque les effectifs recensés deviennent de plus en plus petits et il est de plus en plus nécessaire de recourir à des altérations des chiffres recensés pour protéger la vie privée. L'amplitude de α demeure quand-même basse avec un maximum de 3.5%.

En ce qui concerne les erreurs d'ajustement β , la Figure 2-26 montre les tendances attendues, i.e., une augmentation de β quand la strate de référence est désagrégée. Cela peut être expliqué par l'augmentation des erreurs α avec la désagrégation de la strate de référence, mais aussi par l'explosion de la complexité de la synthèse. Alors qu'un seul ensemble de contraintes doit être respecté par l'algorithme de synthèse au niveau de la RMR, 7525 ensembles de contraintes doivent être respectés au niveau des AD pour Toronto par exemple. Ainsi, l'augmentation de la complexité de la synthèse augmente les erreurs d'ajustement potentielles produites par l'algorithme. Alors que l'augmentation des totaux nuls arrondis (reflétée par l'augmentation d' α) explique aussi l'augmentation de β lorsque la strate de référence est désagrégée, il est important de mentionner que le problème des combinaisons d'attributs non observées dans l'échantillon n'explique pas la variation de β dans notre cas. En fait, puisqu'aucune information spatiale, à part la RMR d'appartenance, n'est fournie dans le FMGD, l'ensemble de l'échantillon de la RMR est

utilisé pour la synthèse à toutes les strates de référence. L'échantillon utilisé pour synthétiser dans une AD est donc le même que celui utilisé pour synthétiser dans la RMR. Cela n'empêche pas que lorsqu'un échantillon propre à une AD est disponible, sa taille réduite augmenterait les chances d'avoir des combinaisons d'attributs non observées et nuirait par conséquent à l'ajustement de la population synthétique.

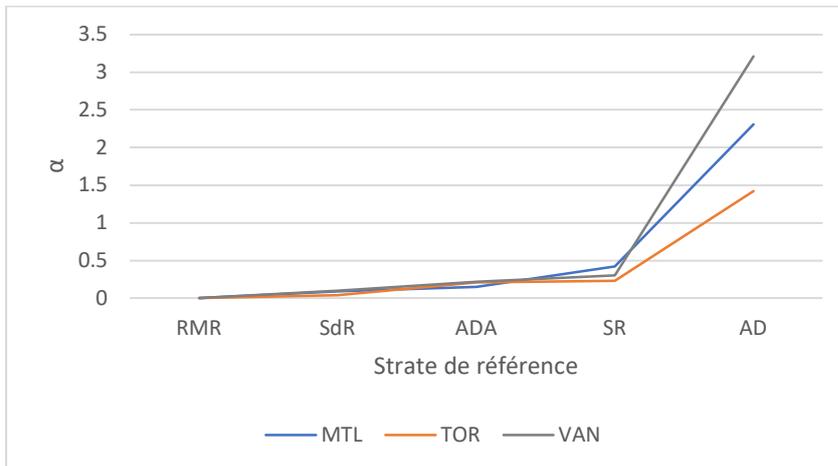


Figure 2-25. Variation de α en fonction de la strate de référence

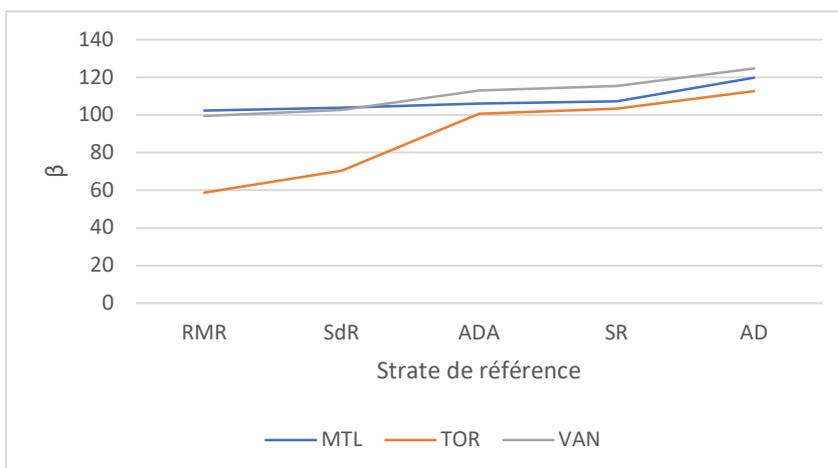


Figure 2-26. Variation de β en fonction de la strate de référence

Les erreurs de spatialisation γ (Figure 2-27) diminuent lorsque l'on désagrège la strate de référence. Cela confirme la précision spatiale améliorée de la population synthétique aux strates les plus désagrégées, donc la force de l'hypothèse d'homogénéité spatiale aux strates les plus agrégées. Les amplitudes de γ sont largement plus élevées que celles de β . Alors que le maximum atteint par β au niveau de l'AD est d'environ 120% pour les trois RMR, γ atteint environ 1000% au niveau de la RMR. Une population synthétique est donc plus susceptible de présenter des erreurs de spatialisation que des erreurs d'ajustement. De plus, le ratio entre le maximum et le minimum atteints par γ est d'environ 3, alors qu'il reste en-dessous de 2 pour β (1.2 pour Toronto). Ainsi, le gain de précision spatiale en désagrégeant la strate de référence est plus important que le gain en ajustement entraîné par son agrégation.

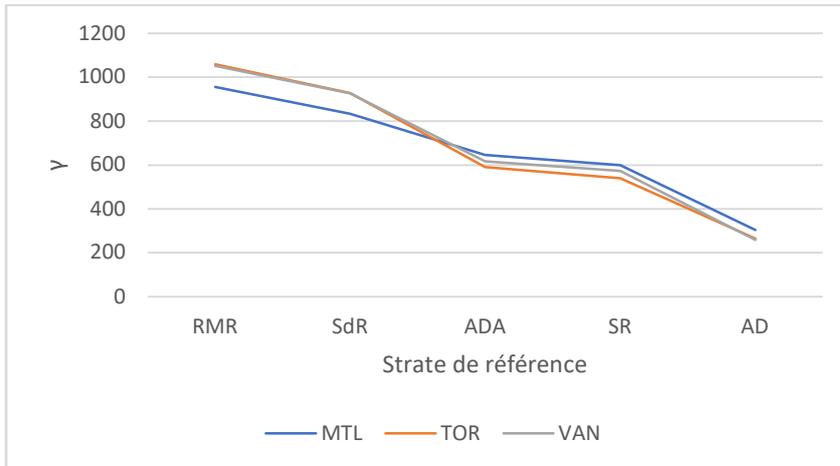


Figure 2-27. Variation de γ en fonction de la strate de référence

Comme nous définissons la qualité de la population synthétique comme une combinaison des erreurs β et γ , la variation de leur somme en fonction de la strate de référence est évaluée (Figure 2-28). L'erreur totale est réduite d'environ 64% au niveau de l'AD. Ainsi, la strate la plus désagrégée permettant de réduire drastiquement l'erreur totale pour les trois RMR, est confirmée comme le meilleur choix pour produire une population synthétique optimale.

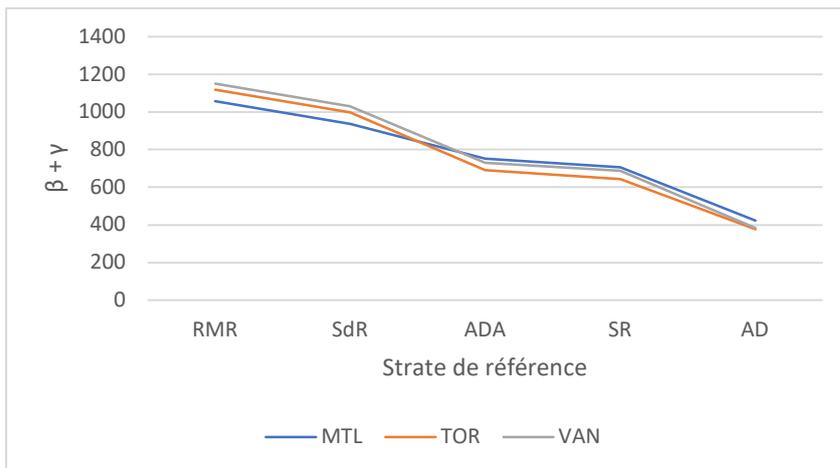


Figure 2-28. Variation de $\beta + \gamma$ en fonction de la strate de référence

2.5.7.2 Impact de l'harmonisation et du contrôle additionnel au niveau de la RMR

Afin d'évaluer l'impact du double contrôle et de l'harmonisation des données sur la qualité de la synthèse, les populations synthétiques produites au niveau de l'AD (fournissant l'erreur totale minimale) sont comparées. Il s'agit de quatre configurations de synthèse : 1) données brutes et une résolution géographique contrôlée (1R), 2) données brutes et deux résolutions géographiques contrôlées (2R), 3) données harmonisées et une résolution géographique contrôlée (1H), 4) données harmonisées et deux résolutions géographiques contrôlées (2H). Les résultats sont présentés sur la Figure 2-29.

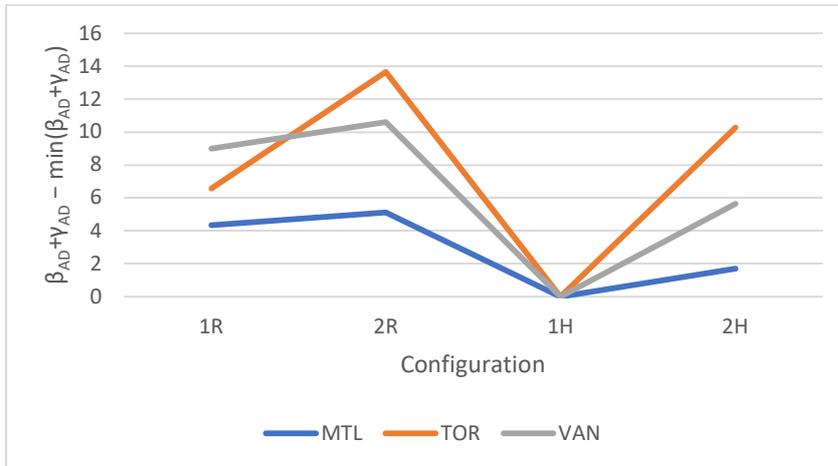


Figure 2-29. Variation de $\beta + \gamma$ au niveau de l'AD en fonction de la configuration utilisée

Konduri et al. (2016) trouvent que le double contrôle améliore la qualité de la population synthétique. Or, en comparant nos scénarios avec double contrôle à leurs homologues avec simple contrôle, le double contrôle s'avère détériorer la qualité de la population synthétique. Cette contradiction relève essentiellement de la différence dans la définition de la qualité. Alors que dans la littérature, la qualité d'une population synthétique reflète uniquement son ajustement à la strate la plus agrégée, notre définition inclut pour la première fois la précision spatiale dans la définition de la qualité. En fait, ajouter un contrôle au niveau de la RMR réduit effectivement β , mais il entraîne une augmentation plus importante de γ . Ainsi, pour minimiser l'erreur totale, le simple contrôle s'avère plus efficace. Enfin, tel que prévu, l'harmonisation permet de réduire l'erreur totale. La meilleure pratique est alors de contrôler une seule résolution géographique en harmonisant à l'avance les totaux du recensement pour qu'ils respectent les totaux correspondants au niveau de la RMR.

2.5.8 Conclusions et perspectives

Pour produire une population synthétique, il faut d'abord choisir une strate de référence. Or, l'impact de ce choix sur la qualité de la synthèse n'a jamais été mesuré ce qui ne permet pas d'effectuer un choix éclairé. Les travaux menés ont pour objectif principal d'évaluer l'impact de l'agrégation de la strate de référence sur la qualité de la population synthétique. Pour ce faire, des populations synthétiques sont générées pour les RMR de Montréal, Toronto et Vancouver en variant la strate de référence pour des fins de comparaison. L'impact des incohérences inter-résolution du recensement et de mesures potentielles de mitigation des erreurs (double contrôle et harmonisation) est aussi évalué.

Les contributions des travaux sont multiples. Sur le plan méthodologique, un indicateur permettant de mesurer l'erreur de spatialisation est défini et intégré pour la première fois dans la mesure de la qualité de la synthèse, et une méthode d'harmonisation des totaux du recensement est proposée. Les résultats montrent que lorsque l'on désagrège la strate de référence, α et β augmentent alors que γ diminue. Toutefois, l'amplitude et la sensibilité des erreurs de spatialisation γ à l'agrégation de la strate de référence se sont avérées plus importantes que celles des erreurs d'ajustement β . Ainsi, une population synthétique générée à la strate la plus désagrégée affiche une erreur totale ($\beta + \gamma$) minimale. Cette erreur peut être encore plus réduite en utilisant des totaux harmonisés du recensement pour la synthèse. Enfin, ajouter un contrôle correctif au niveau de la RMR s'est avéré contre-productif et demeure à éviter pour

optimiser la qualité de la population synthétique. En résumé, la meilleure pratique définie est celle d'une synthèse avec données harmonisées en contrôlant simplement la strate la plus désagrégée.

Cette meilleure pratique est appliquée dans la suite des travaux pour générer des populations synthétiques optimisées que l'on cherche alors à spatialiser au niveau du bâtiment. Le but est de proposer une méthodologie permettant de générer une population synthétique qui reflète les caractéristiques sociodémographiques, mais aussi la distribution spatiale de la population réelle. Celle-ci permettrait d'optimiser la modélisation des comportements de mobilité observés.

2.6 Multimodalité et scénarios post-pandémie

- Par : Hamed Malekzadehnooghabi, doctorat
- Supervision : Pre Catherine Morency
- Contexte : Chaire Mobilité
- Fin prévue : décembre 2024

2.6.1 Introduction

Ce projet de recherche porte sur la “Modélisation des Comportements Projetés Post-COVID (avec Enquêtes COVID 2020, 2021) et Scénarios Plausibles Post-COVID”. L'épidémie de COVID-19, amorcée en mars 2020 a radicalement modifié les habitudes de mobilité quotidienne. La COVID-19 a eu des impacts à court et fort possiblement à long terme sur la façon dont les gens se déplacent, télétravaillent, font leurs achats (en ligne ou en personne), et sur plusieurs autres activités. Par exemple, selon la littérature, quand on compare les données pré et pendant-COVID, le pourcentage de personnes qui prévoient faire du télétravail a augmenté de 13% à 26% aux États-Unis (Salon et al., 2021). De plus, 40 % des voyageurs aériens à des fins professionnelles anticipent voyager moins fréquemment après la COVID. Concernant les achats d'épicerie en ligne, il a été observé qu'ils ont été doublés. En plus de tout ce qui précède, les gens généralement marchent et font plus de vélo (Salon et al., 2021).

Un autre élément important de la recherche sur l'impact de la COVID est la formation d'habitudes. En moyenne, il faut 66 jours pour qu'une habitude se forme avec une amplitude de 18 à 254 jours (Lally et al., 2010). La pandémie a nettement dépassé la borne supérieure de cette plage ce qui permet de supposer que de nouveaux comportements, notamment de mobilité, se sont formés à travers la période COVID-19. Par conséquent, il est important de comprendre et anticiper les comportements plausibles qui suivront la pandémie.

2.6.2 Principales questions de recherche

Les questions de recherche les plus importantes de ce projet sont les suivantes:

- Quelles sont les évolutions des comportements de mobilité des personnes à travers trois périodes à savoir pré-COVID, pendant-COVID, et post-COVID (comportements anticipés)?
- Comment les gens envisagent-ils se déplacer dans un monde post-COVID?
- Comment et dans quelle mesure le comportement multimodal (utilisation de plusieurs modes et virtualisation des activités) des personnes changera-t-il en raison de la pandémie?

Il y a aussi deux questions concernant les attitudes des personnes:

- Comment les perceptions des personnes en regard de la COVID et les inquiétudes quant à leur santé affectent-elles leurs comportements de mobilité?
- Quelles variables attitudinales affectent les choix de mobilité notamment l'utilisation des différents modes de transport?

2.6.3 Approches de modélisation potentielles

Différents modèles peuvent être utilisés à des fins d'analyse. Par exemple on peut utiliser le modèle logistique basé sur la segmentation latente (LSL), les modèles d'équations structurelles (SEM), les modèles de choix hybrides (HCM), les modèles d'apprentissage automatique (GBM, RF, etc.), et d'autres modèles potentiels qui semblent appropriés.

Les modèles traditionnels tels que MNL, ont généralement une hypothèse d'indépendance des alternatives non pertinentes (IIA), ce qui signifie que les variables exploratoires ne sont pas autorisées à varier d'une population à l'autre. Ainsi, de tels modèles négligent l'hétérogénéité des préférences non observées qui peut se produire lorsque des choix sont faits. Cela peut entraîner des estimations incohérentes et biaisées (Provencher et Bishop, 2004). Pour surmonter ce problème, le modèle LSL capture l'hétérogénéité non observée en triant implicitement les individus en segments latents discrets. De plus, les modèles d'équations structurelles et modèles de choix hybrides sont utiles pour inclure des variables d'attitude ainsi que des caractéristiques sociodémographiques et de mobilité.

2.6.4 Données

Les données que nous allons utiliser sont l'enquête Origine Destination 2018 à des fins de comparaison et pour représenter les comportements pré-COVID ainsi que les résultats d'enquêtes visant à mesurer les impacts de la COVID sur les systèmes d'activités et la mobilité réalisées par l'équipe de la Chaire Mobilité (2020, 2021, et probablement 2022 qui inclura plus de variables attitudinales).

Certaines des variables attitudinales disponibles dans les enquêtes COVID de 2020 et 2021 sont liées aux raisons associées à un changement dans le nombre de véhicules possédés, à un déménagement, au fait de télétravailler ou pas, aux changements dans l'utilisation de différents modes de transport, et autres. D'autres variables pourront être incluses dans la version 2022 de l'enquête en lien avec les enjeux de sécurité perçue, fatalisme, inquiétudes de santé (tendance des individus à s'inquiéter pour leur santé), priorité de prévention face à la COVID lors du choix d'un mode de déplacement, niveau de productivité en télétravail, etc.

2.6.5 Résultats prévus

Ce projet vise à comprendre la transformation de différents comportements à travers différentes phases de la COVID ainsi qu'à développer des scénarios plausibles pour la période post-COVID:

- Fréquence des déplacements
- Utilisation du transport en commun
- Utilisation du vélo
- Utilisation de la marche
- Utilisation de la voiture privée

De plus, ce projet peut aussi aider à développer des stratégies pour mieux se préparer aux changements de comportement de déplacement après la COVID.

2.6.6 Schéma

Afin de mieux comprendre les grandes lignes du projet, la figure ci-dessous présente une vue schématique des composantes du projet. Les données, les modèles et les variables d'attitude ainsi que les résultats de l'enquête OD de 2018 fournissent des informations sur les comportements pré, pendant et post-COVID (comportements anticipés tels que déclarés par les répondants) en vue de mieux anticiper comment les gens se déplaceront dans un monde post-COVID?''.

QR: Comment les gens se déplacent-ils dans un monde post-COVID?

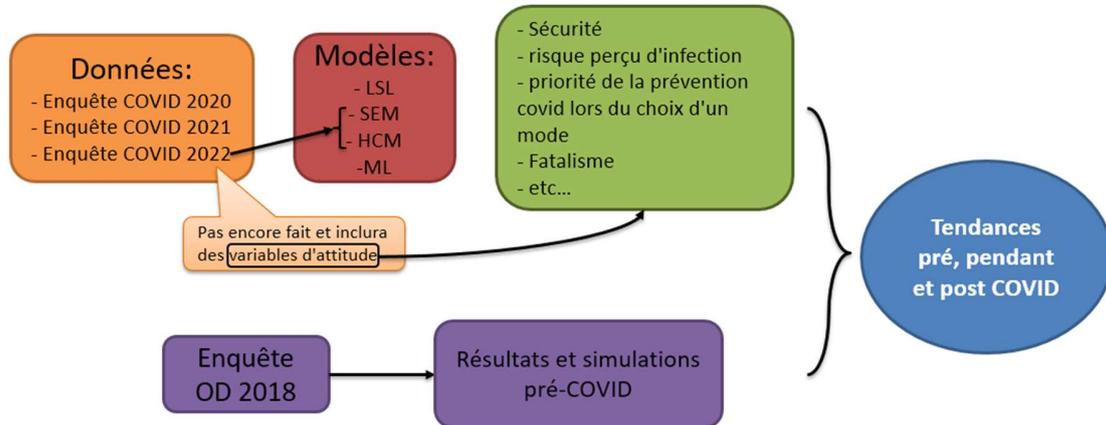


Figure 2-30. Vue schématique du projet de recherche de "multimodalité et scénarios post-pandémie"

2.7 Indicateurs de vulnérabilité des réseaux de transport en commun

- Par: Kaveh Rezvani Dehaghani, doctorat
- Supervision: Pre Catherine Morency
- Contexte: Chaire Mobilité / Transition

2.7.1 Introduction

La vie contemporaine dépend fortement de l'ensemble des systèmes connus sous le nom d'infrastructures pour soutenir le bien-être et le niveau de vie des personnes. Le principal inconvénient de cette dépendance est qu'une interruption ou une perturbation d'un système peut entraîner des conséquences négatives sur la société (Jenelius et Mattsson 2015). Les systèmes de transport sont un des systèmes vitaux, car les dommages qu'ils subissent imposent peuvent avoir des impacts importants sur d'autres systèmes. Cette question nécessiterait d'évaluer la vulnérabilité des systèmes de transport dans le monde d'aujourd'hui.

Compte tenu du rôle important du transport en commun dans les transports urbains et du large éventail de perturbations pouvant en affecter la performance, l'évaluation de la vulnérabilité des systèmes de transport en commun a suscité beaucoup d'attention au cours des dernières années. Les perturbations telles que les accidents, les incidents, les conditions météorologiques et les événements peuvent affecter la stabilité et la fiabilité d'un système de transport en commun. Les perturbations se propagent rapidement dans le réseau et, si elles ne sont pas traitées, une partie considérable du réseau pourrait être influencée. Ceci se traduit par des désagréments pour les usagers, affectant significativement la pérennité des activités urbaines et la vie quotidienne des habitants.

Deux grands modes de transport en commun sont le bus et le métro qui sont généralement complémentaires. Cette étude se concentre sur l'évaluation de la vulnérabilité d'un réseau (réseau de transport en commun de Montréal) composé de modes ferroviaires - y compris le métro et le train - et d'autobus face à différentes perturbations ainsi que les impacts de ces perturbations sur les usagers.

2.7.1.1 Questions de recherche

1. Comment mesure-t-on la vulnérabilité des réseaux de transport en commun ?
2. La variabilité de la vulnérabilité d'un réseau de transport public est-elle importante ? La vulnérabilité change-t-elle de manière significative dans le temps et dans l'espace et selon différents scénarios de défaillance ?
3. Les perturbations du réseau affectent-elles tous les utilisateurs du réseau de transport en commun et les paires Origine-Destination de la même manière ?

2.7.1.2 Objectifs de recherche

1. Introduire de nouveaux indicateurs pour quantifier la vulnérabilité des réseaux de transport en commun et éventuellement des réseaux multimodaux.
2. Étudier l'évolution de la vulnérabilité des réseaux de transport en commun à différents moments et par rapport à différents scénarios de défaillance.
3. Étudier l'étendue de l'influence des perturbations sur différentes catégories d'usagers, différentes paires Origine-Destination ainsi que d'autres composantes du réseau de transport en commun et d'autres réseaux de transport.

2.7.2 État des connaissances

2.7.2.1 Vulnérabilité en transport

La vulnérabilité des réseaux de transport trouve son origine dans (Berdica 2002). Elle a défini la vulnérabilité des transports comme une susceptibilité aux incidents pouvant entraîner des réductions considérables de l'état de service du réseau routier². Le concept de vulnérabilité a été développé par (Berdica et Mattsson 2007), (Jenelius et Mattsson 2012) et (Jenelius, Petersen et al. 2006). (Jenelius et Mattsson 2015) ont exprimé le concept de manière concise comme suit : La vulnérabilité du système de transport est... le risque de perturbation et de dégradation du système de transport³.

Un autre terme dans le domaine de l'ingénierie des transports qui a un concept assez lié à la vulnérabilité est la robustesse du réseau qui a été introduit par (Scott, Novak et al. 2006). Ces auteurs ont défini l'indice de robustesse du réseau comme la variation du coût du temps de trajet associée au réacheminement de toute la circulation dans le système si ce segment devenait inutilisable⁴. Ensuite, le concept de robustesse a été développé, des définitions plus précises ont été fournies par d'autres chercheurs :

- La robustesse du réseau est définie comme la mesure selon laquelle le réseau de transport peut fonctionner en présence de diverses perturbations de capacité sur les liens (Sullivan, Novak et al. 2010).
- La robustesse est la mesure selon laquelle, dans des circonstances prédéfinies, un réseau est capable de maintenir la fonction pour laquelle il a été initialement conçu (Snelder, Van Zuylen et al. 2012).
- La robustesse est la capacité du réseau à maintenir son niveau de service initial en présence d'incidents (de Oliveira, da Silva Portugal et al. 2016).

2.7.2.2 Méthodes d'évaluation de la vulnérabilité

Différentes classifications ont été proposées pour catégoriser les méthodologies d'analyse de la vulnérabilité des réseaux dans le domaine de l'ingénierie des transports au cours de la dernière décennie. L'une des premières catégorisations a été présentée par Mattsson et Jenelius (2015). Ces auteurs divisent les méthodologies d'analyse de vulnérabilité en méthodologies topologiques et en méthodologies basées sur le système. De même, Pan, Yan et al. (2021) ont classé les méthodologies de la recherche selon les mêmes catégories. Ils ont considéré les modèles mathématiques, les simulations et les méthodes basées sur les données comme les sous-catégories des méthodes de performance du système. Zhang et Zhang (2019) ont classé les méthodes des recherches antérieures en trois catégories de base appelées « méthodes basées sur la topologie du réseau », « méthodes basées sur l'accessibilité » et « méthodes basées sur des modèles mathématiques ». Cependant, l'une des catégorisations les plus complètes est présentée par Taylor (2017); il a classé les méthodologies d'analyse de la vulnérabilité des réseaux de transport comme suit :

² « a susceptibility to incidents that can result in considerable reductions in road network serviceability»

³ « The vulnerability of the transportation system is...the risk of disruption and degradation of the transportation system»

⁴ "the change in travel-time cost associated with rerouting all traffic in the system should that segment become unusable"

- **Méthodes basées sur la topologie** : méthodes qui tiennent compte de la structure du réseau, visant à identifier les nœuds ou les liens du réseau où une défaillance aurait les effets les plus importants sur le réseau.
- **Méthodes basées sur la facilité d'entretien** : méthodes qui tiennent compte simultanément des caractéristiques topologiques et des caractéristiques du réseau (offre), telles que la capacité des liens et des nœuds, la longueur des liens et les caractéristiques de temps de trajet, pour mesurer les effets probables des pannes de réseau sur les opérations et les déplacements.
- **Méthodes basées sur l'accessibilité** : méthodes qui considèrent la demande de déplacement de pointe (la plus élevée) et qui s'intéressent à la capacité de la population d'une région (et des différentes zones de cette région) à poursuivre leurs activités économiques et sociales normales lorsque le réseau de transport est dégradé.

2.7.2.3 Indicateurs de vulnérabilité topologique

La théorie des graphes, les réseaux complexes, la science des réseaux ou les méthodes topologiques sont les noms qui évoquent un concept similaire, considérant les réseaux comme un ensemble fini de nœuds et de liens. Lorsque ces méthodes sont utilisées pour modéliser le réseau, les indicateurs topologiques les plus fréquemment employés ((Gao, Liu et al. 2019);(Ma, Liang et al. 2020)), qui sont aussi appelés indicateurs structurels (Ma, Liang et al. . 2020), sont les suivants.

Intermédierité (betweenness)

Cet indicateur montre le rapport entre le nombre de chemins les plus courts à travers le lien ou le nœud et le nombre total de chemins les plus courts (Zhang, Zhao et al. 2021);(Chen, Zhang et al. 2021);(Zhang et Yao 2019);(Chaoqi, Ying et al. 2018).

$$B_i = \sum_{pq} \frac{\tau(i)_{pq}}{\tau_{pq}}$$

Où $\tau_{pg}(i)$ représente le nombre de chemins les plus courts qui traversent le nœud i entre les nœuds p et q , et τ_{pq} représente le nombre total de chemins les plus courts entre les nœuds p et q .

Degré

Cet indicateur affiche le nombre de tous les liens connectés au nœud i . Il est divisé en degrés entrants et sortants dans un graphe orienté (Zhang, Zhao et al. 2021); (Zhang, Wang et al. 2021).

$$D=k_i$$

Degré moyen

Cet indicateur montre la moyenne des degrés de tous les nœuds i du réseau (Ma, Shi et al. 2020); (Tran, Cheong et al. 2019).

$$k = \frac{1}{N} \sum_{i=1}^N k_i$$

Taille relative du composant géant (giant component)

Cet indicateur montre la taille relative du composant géant, noté S , qui caractérise la robustesse d'un réseau du point de vue de l'échelle du réseau, qui est calculé par :

$$S = \frac{N}{N'}$$

Où N' représente la taille du composant géant après certaines perturbations de nœuds, et N est la taille du réseau original. Une valeur élevée de S indique une forte stabilité du réseau face à des perturbations.

Coefficient d'agrégation

Cet indicateur montre le rapport du nombre de nœuds entre le nœud i et les points voisins sur le nombre total de nœuds possibles $k_i(k_i - 1)/2$ (Ma, Shi et al. 2020); (Tran, Cheong et al. 2019).

$$CC_i = \frac{2l_i}{k_i(k_i - 1)}$$

Coefficient d'agrégation global du réseau

Cet indicateur présente la moyenne des coefficients d'agrégation de tous les nœuds du réseau (Ma, Shi et al. 2020)

$$CC = \frac{1}{N} \sum_{i=1}^N CC_i$$

Longueur moyenne de trajet

Cet indicateur montre l'efficacité naturelle du réseau i.e. la distance moyenne entre chaque paire de nœuds (Li, Guo et al. 2018).

$$D(G) = \frac{1}{N(N-1)} \sum_{i \neq j \in G} d_{ij}$$

2.7.3 Contributions attendues

Le projet proposera différentes définitions et métriques pour mesurer la vulnérabilité des réseaux ainsi que les impacts de différentes perturbations. Le réseau de transport en commun montréalais sera d'abord utilisé comme étude de cas mais le projet vise à terme à intégrer d'autres modes de transport ainsi qu'à assurer la transférabilité des indicateurs développés à d'autres réseaux en vue d'une intégration dans la plateforme Transition.

2.8 Méthodologie d'évaluation de la qualité des accès actifs aux stations de métro

- Par : Eya Chékir, maîtrise
- Supervision : Pre Catherine Morency
- Contexte : Chaire Mobilité
- Fin projetée : décembre 2022

2.8.1 Introduction

Les modes de transport en commun sont connus pour être de grands générateurs de déplacements à pied. En effet, pour prendre le transport en commun, il faut d'abord accéder à un de ses points d'accès (tant à l'origine qu'à la destination). Et les conditions dans lesquelles se font les accès aux arrêts sont très déterminantes de l'attractivité globale du transport en commun et, de façon générale, du nombre d'utilisateurs et destinations qui seront accessibles à l'intérieur d'une distance et d'un temps considérés acceptables. Pourtant, il y a peu de travaux qui s'intéressent spécifiquement aux cheminements d'accès aux réseaux de transport, à leur impact sur l'attractivité du transport en commun et à l'efficacité des bassins d'attraction des nœuds de transport en commun structurants comme les stations de métro.

Ainsi, les conditions d'accès aux nœuds de transport en commun sont déterminantes de son attractivité. Pour cela nous nous intéressons au diagnostic de l'accessibilité piétonne aux différents édicules des stations de métro montréalaises à travers l'estimation d'indicateurs de couverture (bassin d'attraction), tortuosité (des parcours piétonniers) et caractéristiques des chemins d'accès aux stations. Ainsi, l'analyse des caractéristiques des différents niveaux de bassins d'attraction autour des stations de métro est le meilleur moyen pour comprendre les facteurs qui entrent en jeu lors d'un déplacement.

2.8.2 Objectifs

- Définir les bassins d'attraction autour des stations de métro.
- Développer des indicateurs pour déterminer la qualité du réseau piétonnier d'accès aux stations et de l'environnement bâti en général sur l'attractivité du transport en commun.
- Étendre la caractérisation sur l'ensemble des stations de métro montréalais puis au reste des nœuds d'accès au réseau de transport en commun

2.8.3 Revue de littérature

2.8.3.1 Complémentarité transport en commun et marche

Le transport en commun, du fait de l'implication de la marche qu'il entraîne pour se rendre d'une station de transport en commun à une destination (ou de l'origine à la station), peut être vu comme un mode de déplacement actif (Bennicelli, 2012). Vivre ou travailler près d'une station de transport en commun est un incitatif à la marche. Ainsi, les personnes qui utilisent les transports en commun pour aller au travail, ont tendance à marcher plus que celles qui utilisent uniquement le véhicule privé. Lorsque la desserte de transport en commun est efficace et rapide (métro, tram), les gens sont d'avantage prêts à s'y rendre à pied. De ce fait, on estime que les gens sont prêts à marcher une distance de 500 mètres pour accéder à un arrêt d'autobus. Cette distance peut atteindre un kilomètre quand il s'agit d'une station de métro ou d'une gare de train (Lachapelle et Noland, 2012; Collectivités viables, 2016).

2.8.3.2 La distance pour atteindre un point d'accès au transport en commun

La distance du transport en commun peut être vue comme la distance moyenne entre le domicile, ou lieu d'emploi, et la station la plus proche de transport en commun (collectivités viables, 2016). Ainsi, la marche est l'un des modes courants d'accès aux services urbains, y compris les transports publics, et peut donc largement influencer la dynamique des infrastructures de transport public. Une distance de marche considérée comme optimale jusqu'aux stations de transport en commun ou aux points d'accès, est une donnée importante pour essayer de caractériser la satisfaction des utilisateurs à l'égard du système de transport urbain global.

Pour un contexte de ville européenne, une distance de marche uniforme jusqu'à une station (5 minutes ou 477 mètres jusqu'à un bus ou un tramway) peut être prise pour mesurer l'accessibilité au réseau de transport en commun; cependant, ce seuil peut varier en fonction de la structure de la ville, sa dynamique et sa densité urbaine. Les activités physiques et l'usage de modes de transport actif (c.-à-d. la marche ou le vélo jusqu'aux gares) pour accéder aux transports en commun aident à réduire l'indice de masse corporelle rendant les usagers des transports en commun moins susceptibles d'être obèses. En fait, l'amélioration des conditions de marche vers les stations de transport en commun peut non seulement contribuer à réduire la circulation (en augmentant l'achalandage du transport en commun), mais peut également faciliter un mode de vie urbain sain. Cependant, la marche est souvent négligée non seulement en tant que mode de transport individuel (non motorisé), mais également en tant que partie des chaînes intermodales lors de la planification et du traitement de la mise en œuvre des transports publics. Souvent, les preuves empiriques ne sont pas claires ou il n'y a pas de consensus concernant les hypothèses de planification sur la distance de marche des stations. Ainsi, il existe une demande d'explorer si les habitudes de marche des usagers du transport en commun confirment le seuil supposé lors de la planification, y compris leur volonté de marcher pour accéder aux installations de transport public au sein de tissus urbains dynamiques (Rumana Islam Sarker, 1 March 2019).

2.8.4 Méthodologie

2.8.4.1 Méthodologie générale

L'étude des déplacements à pied est réalisée grâce à l'utilisation de deux ensembles de variables principales: les variables relatives à la demande et les variables relatives au voisinage – c'est-à-dire à l'environnement urbain. Les variables relatives à la demande regroupent les caractéristiques associées aux individus, aux ménages et à la structure des déplacements.

Ces variables peuvent être entièrement extraites des bases de données de déplacements et sont collectées lors des enquêtes sur la mobilité des personnes.

Les variables de voisinage caractérisent pour leur part l'environnement autour d'un point associé à un déplacement. Elles doivent être extraites de bases de données externes aux enquêtes de mobilité. En effet, l'unité spatiale de mesure peut varier tant dans sa localisation – à l'origine ou à la destination d'un trajet, ou encore tout au long de celui-ci.

**Est-ce que la qualité des accès a un effet sur l'intermodalité
marche et transport en commun**

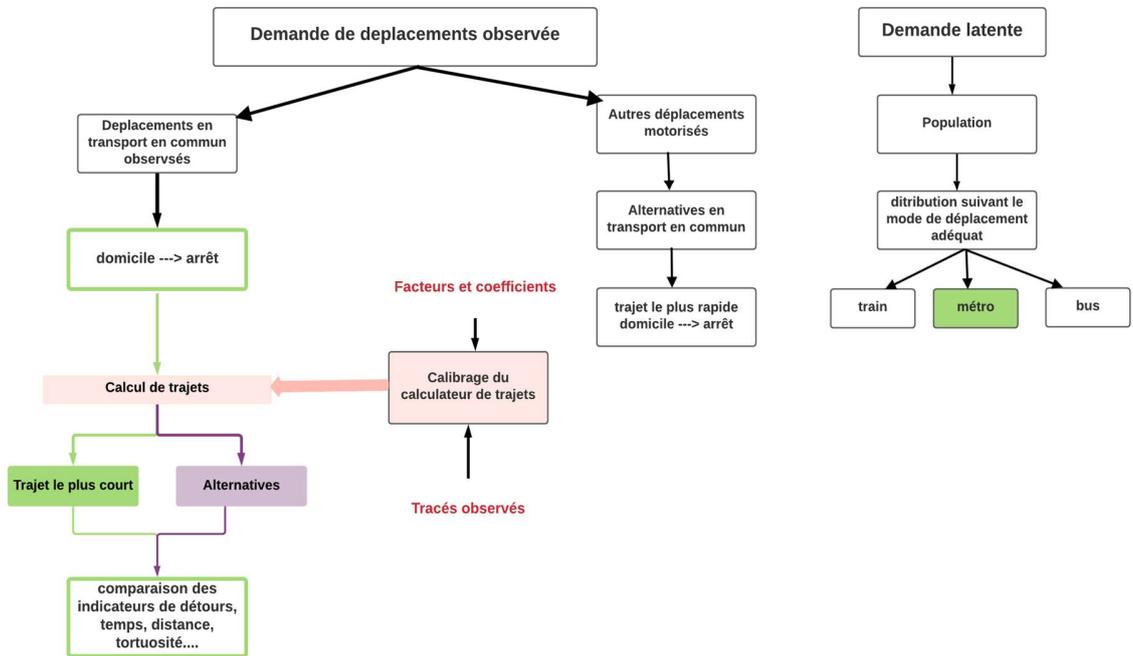


Figure 2-31 Méthodologie générale

2.8.4.2 Méthodologie de création des bassins d'attraction

Pour pouvoir délimiter le bassin d'attraction autour d'un accès de station de métro l'approche présentée à la Figure 2-31 est utilisée.

Au début, nous générons des points aléatoires près de chaque station. Ils peuvent ensuite être utilisés comme points d'origine et la station de métro comme destination. Les chemins de chaque origine à la destination choisie seront obtenus avec un calculateur de trajets. La distance de chaque parcours sera ensuite utilisée pour délimiter les zones où la distance de parcours est inférieure à 200 mètres, 400 mètres, 600 mètres, 800 mètres et 1000 mètres. Les surfaces obtenues seront considérées comme les bassins versants autour des stations. Si une station de métro possède plusieurs entrées, son bassin d'attraction sera le regroupement de toutes les zones de chalandise de ses accès. Une fois ces zones définies, nous allons développer différents indicateurs pour comparer les résultats entre les stations de métro et faire des hypothèses sur leur niveau d'accessibilité.

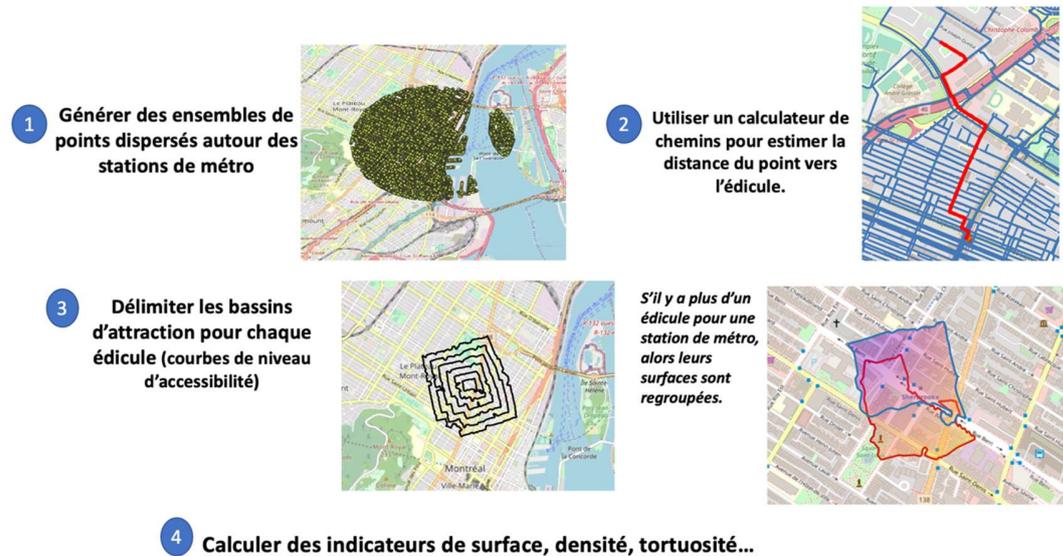


Figure 2-32 Création des bassins d'attraction

2.8.4.3 Sources de données utilisées

Afin de disposer des données nécessaires à l'élaboration de cette étude, plusieurs sources de données ouvertes ont été utilisées :

- Statistique Canada est l'agence nationale des statistiques. Elle rend accessible aux Canadiens des informations fiables sur différents sujets comme l'économie, l'environnement, la société et la démographie... Elle permet à travers ces données d'agir efficacement avec des entrées précises, que ce soit au niveau des citoyens ou des décideurs (Statistique Canada).
- Données Montréal : La Ville de Montréal s'engage à collecter et à gérer des données spécifiques à la ville fournissant des informations importantes et diversifiées pour le bien du développement de la ville et de ses citoyens. La Ville de Montréal s'engage à appliquer la politique des données ouvertes pour inciter ses unités administratives à fournir des données ouvertes.
- La Société de Transport de Montréal (STM) fournit des ensembles de données au format GTFS (General Transit Feed Specification). Ainsi, il devient facile pour chacun d'avoir la localisation des stations de transport en commun, des lignes de métro et d'autobus et leurs horaires (Société des transports de Montréal).
- OpenStreetMap (OSM) est une plateforme collaborative de cartographie en ligne qui fournit une base de données géographique riche et gratuite du monde, grâce à ses utilisateurs aux quatre coins du monde. OSM permet de travailler facilement avec des données très diverses telles que la localisation des magasins, restaurants, téléphones, écoles, banques, pharmacies, prisons, supermarchés....

6. Zone d'étude

Pour débiter cette étude, nous nous concentrons sur le métro de Montréal.

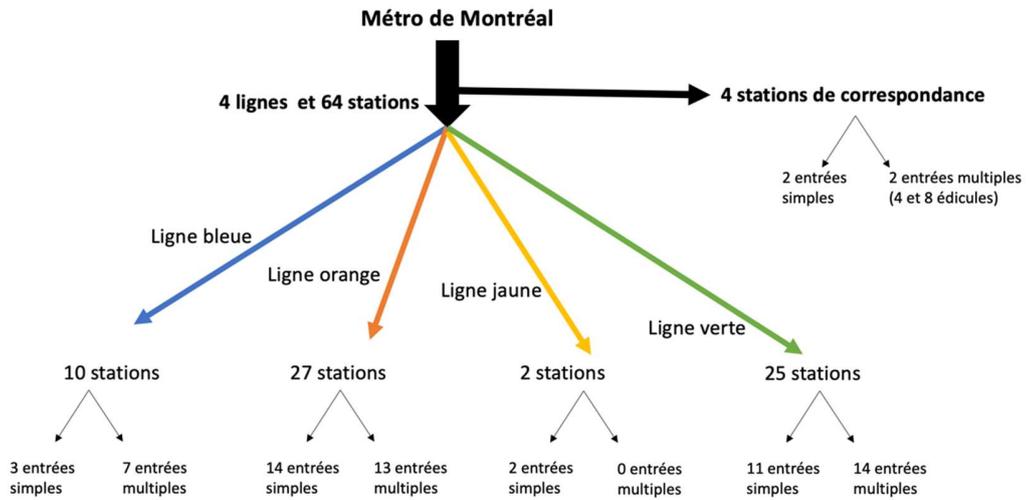


Figure 2-33 Décomposition du métro montréalais

La ligne orange, aussi appelée ligne 2 du métro de Montréal, est la ligne la plus importante du réseau de métro, que ce soit pour sa fréquentation, sa longueur ou son nombre de stations. Cette ligne permet également aux passagers de sortir du territoire de l'île de Montréal en le reliant à la ville de Laval. La Figure 2-34 présente les stations étudiées pour développer la méthodologie.

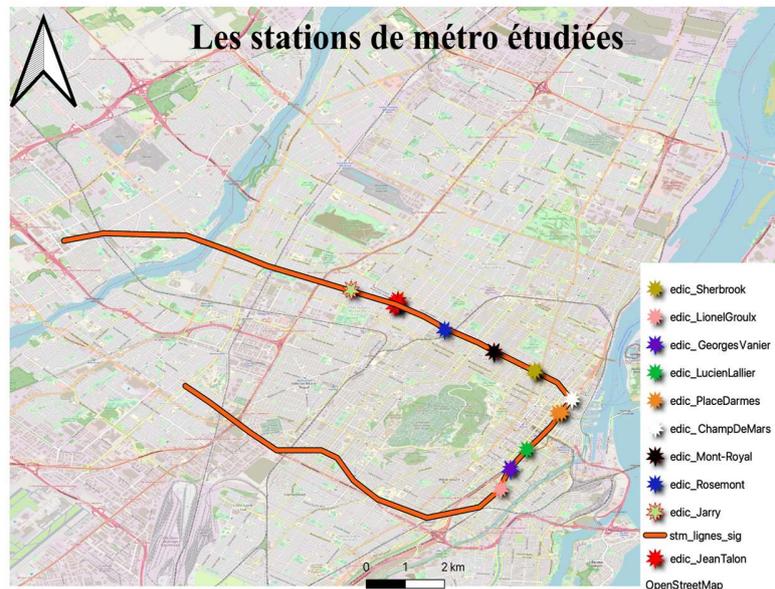


Figure 2-34 Stations actuellement étudiées dans le cadre du projet

2.8.5 Choix des indicateurs

2.8.5.1 *Indicateur de surface*

Cet indicateur est important dans la compréhension de ce qu'est une bonne accessibilité à une station de métro. En effet, si la surface du bassin d'attraction est grande cela peut être signe d'une bonne accessibilité et d'une attractivité plus importante vers les points d'accès, pouvant apporter de ce fait de potentiels nouveaux usagers du métro. Avant toute analyse de nos stations nous pouvons également supposer que plus une station comporte d'édicules, plus elle offre de possibilités d'accès aux usagers. Par conséquent, la surface du bassin d'attraction des stations à plusieurs édicules pourrait être plus importante comparée à celle des stations de métro ayant un édicule.

2.8.5.2 *Indicateur de périmètre*

Le périmètre est un indicateur important. Ce qui le différencie de la mesure de la surface c'est qu'il ne permet pas seulement d'avoir des conclusions sur la taille du contour d'un bassin d'attraction, mais il permet aussi d'avoir une idée de sa complexité. Son étude permet d'émettre des hypothèses sur la présence de détours et d'irrégularités dans la conception du réseau de trottoirs à l'intérieur du bassin étudié et peut être leur manque de praticité.

2.8.5.3 *Pourcentage de la superficie couverte par une courbe de niveau*

Cet indicateur permet de mettre en évidence la différence entre la distance de marche à vol d'oiseau et la distance de marche réelle calculée. Le pourcentage de surface couverte par une courbe de niveau permet d'appréhender la complexité du développement urbain d'un bassin d'attraction. On peut supposer qu'elle met en évidence la présence importante de détours lorsque ce pourcentage est faible, et un bon aménagement urbain lorsque ce pourcentage est à son maximum.

2.8.5.4 *Nombre total de dépanneurs et supermarchés*

La présence de supermarchés et de dépanneurs dans une région est un bon indicateur de son accessibilité piétonne. Car ces commerces doivent être accessibles au plus grand nombre et pas seulement aux automobilistes. On peut donc supposer que leur présence signifie une bonne infrastructure piétonne. Ainsi, nous supposons qu'une attention particulière au confort des piétons peut être apportée lorsque le nombre de commerces est élevé à l'intérieur d'un bassin d'attraction.

2.8.5.5 *Densité de population*

L'étude de cet indicateur part de l'hypothèse que les gens habitent là où tout est plus accessible à pied et où les transports en commun sont bien desservis. Ainsi, plus la station de métro est accessible à pied, plus il y a de monde. Ceci est d'ailleurs le cas au centre-ville où la densité de population est généralement la plus importante.

2.8.5.6 *Longueur du trottoir*

La longueur d'un trottoir peut être vue comme un investissement supplémentaire dans l'infrastructure piétonne, d'une attention particulière prêtée au confort des usagers et de la présence de plus d'espace pour l'accessibilité piétonnière. Ainsi, une longueur de trottoir importante peut signifier plus d'efforts et d'investissements de la part de la municipalité pour faciliter la marche et le transport actif. Mais cet indicateur peut se révéler important pour refléter la complexité de la configuration urbaine, la présence de détours et la pénibilité générale de la pratique de la marche dans l'environnement étudié.

2.8.5.7 Tortuosité moyenne

La tortuosité moyenne dans une zone peut nous aider à distinguer quelle infrastructure piétonne à côté de quel édicule à une station métro offre le moins de détours aux piétons.

2.8.5.8 Fréquence moyenne des bus dans une journée

Cet indicateur peut indiquer si la zone de chalandise est bien desservie par les transports en commun et si toutes les commodités sont offertes aux personnes pour pouvoir rejoindre un accès à une station de métro.

2.8.5.9 Fréquence moyenne des bus pendant la pointe du matin

La fréquence moyenne des bus pour une ligne tout autour d'une station de métro en pointe du matin (entre 5h et 9h) est très importante pour les usagers des transports en commun. Cet indicateur peut montrer si la zone délimitée par le bassin d'attraction étudiée, est bien desservie par les transports en commun lorsque la demande est intense et que toutes les commodités sont offertes aux personnes pour pouvoir rejoindre une station de métro en même temps pour se rendre au travail ou à l'école.

2.8.5.10 Fréquence moyenne des bus pendant la pointe de l'après-midi

La fréquence moyenne des bus pour une ligne tout autour d'une station de métro en pointe de l'après-midi (entre 15h et 18h) est très importante à étudier. Cet indicateur peut montrer si la zone du bassin d'attraction est bien desservie par les transports en commun lorsque la demande est intense et que toutes les commodités sont offertes aux personnes pour pouvoir rejoindre une station de métro en même temps pour rentrer chez elles.

2.9 Interactions entre chaînes de déplacements et choix modal

- Par : Mohamad Dabboussi, maîtrise
- Supervision : Pre Catherine Morency et Pre Geneviève Boisjoli
- Contexte : Chaire Mobilité

2.9.1 Introduction

Le développement durable, une approche qui garantit de répondre aux besoins humains actuels sans sacrifier ni mettre en danger l'avenir des autres sur 3 niveaux principaux (économique, environnemental et social), est l'un des objectifs principaux pour toutes les grandes villes et leaders du monde. À Montréal, l'atteinte de cet objectif nécessite de nombreuses étapes, dont la réduction des gaz à effet de serre et de l'utilisation de l'automobile en sont les principaux exemples. L'augmentation de la part modale du transport en commun est un moyen d'aider à atteindre ces objectifs. C'est d'ailleurs l'un des principaux objectifs du Plan métropolitain d'aménagement et de développement de Montréal élaboré en 2011, avec pour objectif d'augmenter la part modale du transport en commun aux heures de pointe du matin, actuellement à 26 % selon la dernière enquête Origine-Destination de 2018, à 35% d'ici 2031. Et bien que cet objectif puisse être irréaliste en raison des effets encore largement méconnus de la pandémie de COVID-19, l'objectif d'atteindre une meilleure part modale du transport en commun demeure un objectif à long terme à Montréal.

De ce point de vue, il est important de savoir quels facteurs affectent le plus le choix modal et le comportement de déplacement des personnes et quels types de scénarios pourraient nous aider à atteindre un transfert modal de l'utilisation de la voiture vers de meilleures alternatives telles que les transports en commun.

L'interaction entre la structure des chaînes de déplacements et le choix du mode est l'un des facteurs les moins abordés dans la littérature, en particulier avec des découvertes récentes qui suggèrent que c'est en fait l'enchaînement des déplacements qui vient en premier avant le choix du mode. Cela signifie que les gens sont susceptibles de décider quel mode emprunter en fonction de la chaîne de déplacements qu'ils ont en tête (Krygsman et al. 2007 ; Li et al. 2013 ; Yang et al. 2016 ; Ye et al. 2007, Currie et Delbosc, 2011). Par conséquent, les scénarios explorant l'effet de la modification de la structure de certaines chaînes de déplacements pourraient être significatifs.

Il est également essentiel de disposer des bons outils de planification pour tester l'effet de tels scénarios et prévoir avec précision leur impact (Efthymiou et al., 2014). Les modèles de choix modal sont un de ces outils, car ils permettent de prédire le mode de transport des personnes en fonction de différentes variables. Au fil des ans, différents modèles ont été utilisés avec des résultats et des approches différents. Ces dernières années, des algorithmes d'apprentissage automatique tels que les forêts aléatoires ont été utilisés comme outil de modélisation du choix modal avec des résultats impressionnants (Cheng et al 2019). Cependant, l'utilisation de l'algorithme d'apprentissage automatique comme outil de modélisation du choix de mode a pris en compte les trajets individuellement et aucune recherche, à notre connaissance, n'a encore été effectuée avec prise en compte des chaînes de déplacements.

2.9.1.1 Questions de recherche et objectifs

Les questions de recherche suivantes sont soulevées :

- Quels sont les effets de la structure de la chaîne de déplacements sur le choix modal?
- Comment peut-on appliquer les forêts aléatoires dans un contexte de choix modal tenant compte de la structure des chaînes?

Les objectifs de recherche sont :

1. Identifier et mesurer les variables qui ont un impact sur le choix modal; ces variables sont classées selon quatre niveaux : individu, ménage, chaîne de déplacement et environnement bâti.
2. Expérimenter la modélisation du choix modal des chaînes de déplacement et des systèmes d'activités en utilisant les forêts aléatoires et identifier les variables significatives les plus importantes.
3. Examiner les impacts de scénarios qui changent la structure des chaînes de déplacements (par exemple, une chaîne de déplacements complexe devient deux chaînes simples) sur les alternatives modales possibles.

2.9.2 État des connaissances

2.9.2.1 *Concept de chaîne de déplacements*

Le concept de chaîne de déplacement a beaucoup évolué au fil des ans. Plusieurs auteurs ont proposé différentes définitions, certaines très simplifiées et d'autres très complexes. La première mention d'une définition de chaîne de déplacement a été faite par Holzapfel (1986). Pour lui, le domicile d'une personne était considéré comme le point d'ancrage, or une chaîne de déplacements n'est considérée comme telle que lorsque 3 déplacements ou plus sont effectués. Cela signifie que pour lui que la chaîne la plus courante (Domicile-Travail-Domicile) n'est pas considérée comme une chaîne de déplacements à proprement dite.

Une autre définition, proposée par Thill et Thomas, (1987) est basée sur le fait qu'une chaîne est une séquence de déplacements entre différents lieux sur une période donnée. Selon leur définition, une chaîne de déplacements est fonction de la durée des activités concernées plutôt que du nombre d'activités.

Plusieurs autres définitions ont été faites au fil des années par différents auteurs (Goulias et Kitamura (1991), McGuckin et Murakami (1999), Vleugels (2005) etc.).

Primerano et al. (2008) utilisent plusieurs des concepts discutés par les auteurs précédents pour déterminer leur définition de la chaîne de déplacements. Une chaîne est définie comme la séquence des déplacements d'activité primaire et secondaire effectués entre le départ et le retour au domicile. Cette définition est encore la plus utilisée dans la littérature récente avec quelques exemples récents comme les travaux de Schneider et al. (2021) et les travaux de Y. Huang et al. (2021).

La définition et la typologie proposées par Valiquette (2010), puis reprises dans les travaux de Sicotte (2014), sont adaptées pour ce travail. La définition utilisée est presque identique à celle de Primerano et al. (2008), avec la légère différence de ne pas considérer l'ordre dans lequel les activités secondaires se produisent.

2.9.2.2 *Travaux précédents*

Cette recherche s'inscrit dans la continuité des travaux de Sicotte (2014). Dans son travail, il adopte également la typologie des chaînes de déplacements proposée par Valiquette (2010). Les variables explicatives ont été classées en 4 catégories principales : personne, ménage, déplacements et environnement bâti. Ces variables ont été utilisées dans un modèle logit multinomial, où un modèle basé sur les déplacements a été opposé à un autre qui considère les chaînes de déplacements. Une typologie de choix de mode a été adoptée qui divise les différents modes en 2 groupes principaux : avec ancrage et sans ancrage. Cela a permis de générer

différentes alternatives en fonction de la présence ou de l'absence de points d'ancrage (le domicile étant le plus important). Les résultats du modèle MNL ont montré que le modèle considérant les chaînes de trajet produisait une meilleure prédiction que celui considérant les trajets seuls (72,15 % de trajets simulés ont été observés vs 12,74 %), en particulier pour le cas de la conduite automobile (89 % vs 11 %). Dans ce travail, seules les chaînes de déplacements simples ont été considérées et l'étude a été réalisée à partir des déplacements de la ligne de train de banlieue Vaudreuil-Hudson dans la grande région de Montréal.

Cheng et al (2019) est l'une des recherches les plus récentes qui a exploré l'utilisation de la forêt aléatoire comme outil de modélisation pour le choix du mode de déplacement, ainsi que les effets des différents paramètres et variables du modèle sur les performances de prédiction. Dans ce travail, les variables explicatives ont également été regroupées en 4 grandes catégories : informations sur le ménage, l'individu, l'environnement bâti et les déplacements. Les résultats ont montré que les forêts aléatoires se sont mieux comportées en termes de précision de prédiction et de temps d'exécution de calcul par rapport à 3 autres modèles : MNL, machine à vecteurs de support (SVM), amplification adaptative (AdaBoost). L'étude a révélé que la forêt aléatoire est un outil efficace de modélisation du choix de mode, qui est efficace pour traiter la variabilité des données. Les déplacements ont toutefois été considérés individuellement dans cette étude.

2.9.3 Méthodologie

2.9.3.1 Données

Les données des enquêtes origine-destination de 1998, 2008 et 2018 réalisées dans la grande région de Montréal sont principalement utilisées pour ce travail. Les personnes âgées de 15 à 64 ans ont été prises en compte.

2.9.3.2 Typologie de chaînes

Chaînes simples

Ces chaînes ne consistent qu'en deux déplacements : un vers une activité, puis un autre pour le retour à la maison.

Chaînes complexes

Ces chaînes incluent toutes les chaînes qui consistent en plus de deux déplacements, c'est-à-dire plus d'une activité. Dans ce cas, on trouve une activité principale et une ou plusieurs activités secondaires pour compléter la chaîne.

Chaînes complexes mono-boucle

Ces chaînes incluent toutes les chaînes qui consistent en plus de deux déplacements, c'est-à-dire plus d'une activité. Dans ce cas, on trouve une activité principale et une ou plusieurs activités secondaires pour compléter la chaîne.

Chaînes complexes multi-boucle

Les chaînes à plusieurs boucles sont constituées de plusieurs boucles. Ainsi, le départ et le retour d'un individu vers un point d'ancrage autre que le domicile, qui est généralement le lieu de l'activité principale, crée une nouvelle boucle. Un exemple courant est celui d'une personne qui quitte son domicile pour aller travailler et quitte le lieu de travail à l'heure du dîner pour aller manger au restaurant, puis retourne au travail en début d'après-midi et rentre finalement chez elle en fin de journée.

Chaînes ouvertes

Ces chaînes sont en fait des chaînes de déplacements non connectées, c'est-à-dire que le dernier déplacement a un autre motif que le retour à la maison (ou que le premier trajet ne commence pas à la maison). Ces types de chaînes ne sont pas considérés dans ce travail.

L'évolution des types de chaînes au cours des 20 dernières années est présentée dans le Tableau 2-13. Les chaînes simples représentent plus de 80 % du total des chaînes au cours de toutes les années précédentes. Ils sont au plus bas en 2018 à 81% et au plus haut en 2008 à 85%. Aucune tendance claire n'est observée, à l'exception de l'augmentation constante des chaînes mono-boucle complexes lorsqu'elles sont calculées à partir de toutes les chaînes complexes.

Tableau 2-13 - Évolution des types de chaînes de 1998 à 2018

Type de chaîne	1998	2008	2018
Chaîne simple	82%	85%	81%
Chaîne complexe	18%	15%	19%
Mono-boucle complexe	75%	77%	81%
Multi-boucle complexe	15%	23%	19%

2.9.3.3 Évolution de la répartition modale des chaînes

Lorsque l'on considère les chaînes de déplacements, il devient de plus en plus problématique d'énumérer toutes les alternatives possibles. Par exemple, une chaîne de déplacements complexe avec 4 activités pourrait avoir plus de 5 000 alternatives possibles. Par conséquent, une liste simplifiée d'alternatives possibles a été créée pour ce travail qui regroupe les alternatives en fonction de leur fréquence d'apparition. Cette liste est :

- CD uniquement : Une chaîne dans laquelle tous les trajets ont été effectués en voiture uniquement
- PT uniquement : Une chaîne dans laquelle tous les déplacements ont été effectués en transport en commun uniquement
- CP uniquement : une chaîne dans laquelle tous les trajets ont été effectués en étant un passager de voiture uniquement
- Vélo/Marche : Une chaîne dans laquelle tous les déplacements ont été effectués soit à vélo seulement, soit en marchant seulement, soit en combinant les 2
- Bimodal : Une chaîne dans laquelle tous les déplacements ont été effectués par Kiss and Ride ou Park and ride ou une combinaison seulement
- Mixte avec voiture : Une chaîne dans laquelle les déplacements ont été effectués par une combinaison de modes, dont un ou plusieurs incluent l'utilisation d'une voiture (CD, CP, Bimodal)
- Mélange sans voiture : une chaîne dans laquelle tous les déplacements ont été effectués par une combinaison de modes, sans l'utilisation d'une voiture dans aucun déplacement.

- Autres uniques : Une chaîne dans laquelle tous les déplacements ont été effectués par 1 seul mode qui n'a pas été mentionné (taxi, autobus scolaire, etc. ...)

Le Tableau 2-14 montre l'évolution de la répartition modale de la chaîne au cours des 20 dernières années. Il a été démontré que l'année 2008 a été la meilleure car elle a les chaînes de voitures les plus basses et les chaînes les plus élevées utilisant uniquement le transport en commun. Une tendance observée dans le tableau est la diminution des alternatives « mixte avec voiture » et « mixte sans voiture ». Cela implique que les gens penchent vers l'utilisation d'un seul mode pour l'ensemble de leurs chaînes de déplacements au lieu de combiner différents modes ensemble.

Tableau 2-14 Évolution de la répartition modale des chaînes de déplacements de 1998 à 2018

Mode	1998	2008	2018
AC uniquement	61%	58%	60%
TC uniquement	12%	16%	14%
Vélo/Marche	9%	9%	9%
Mixte avec voiture	5%	4%	4%
Mixte sans voiture	4%	4%	3%
AP uniquement	6%	6%	6%
Bimodal	2%	2%	2%
Autre mode unique	1%	2%	2%

2.9.3.4 Évolution des systèmes d'activités

Les systèmes d'activité indiquent le nombre et le type d'activités qu'un individu exécute en une journée. Ils sont importants car ils nous donnent aussi un aperçu du type et du nombre de chaînes que les gens font tout au long de leur journée.

4 systèmes uniques représentent plus de 95 % de tous les systèmes d'activité différents :

- une chaîne simple seulement
- une chaîne complexe uniquement
- deux chaînes simples seulement
- une chaîne simple + une chaîne complexe

Ces quatre systèmes plus "autres" qui incluent tous les autres systèmes non mentionnés sont choisis pour l'étude.

Réaliser une seule chaîne simple par jour devient de plus en plus populaire au fil des années (de 61% à 63%), de même que réaliser une seule chaîne complexe (de 15% à 17%). Cela se fait au détriment des systèmes de deux chaînes simples seulement (de 16% en 98 à 12% en 18). Cela implique que les gens préfèrent réaliser une seule chaîne par jour, qu'elle soit simple ou complexe, au détriment de la réalisation de deux chaînes simples. Le Tableau 2-15 montre l'évolution du système de répartition des activités au cours des 20 dernières années.

Tableau 2-15 Évolution des systèmes d'activités de 1998 à 2018

Systèmes d'activités	1998	2008	2018
Une chaîne simple seulement	61%	62%	63%
Une chaîne complexe seulement	15%	16%	17%
Deux chaînes simples seulement	16%	14%	12%
Une chaîne simple et une chaîne complexe	4%	4%	4%
Autres	4%	4%	3%

2.10 Fardeau du répondant

- Par : Hedieh Keshavarz, maîtrise
- Supervision : Pre Catherine Morency
- Contexte : Chaire Mobilité / Mobilizing Justice

2.10.1 Introduction

Les enquêtes font partie des méthodes très anciennes de collecte de données dans le domaine des comportements de transport et de mobilité. De nos jours, grâce à l'évolution des technologies basées sur le Web, les enquêtes Web ont gagné en intérêt en étant moins coûteuses, moins chronophages et en étant capables de cibler plus d'audience. Cependant, le faible taux de réponse est l'un des défis les plus importants auxquels sont confrontés les sondages en ligne. "Le fardeau du répondant fait référence à l'effort requis par un répondant pour répondre à un questionnaire" (Briz-Redón. Á., 2021). L'augmentation de la charge de réponse réduira le taux de participation et la qualité des réponses recueillies par les enquêtes. Le taux de participation est considéré comme un facteur très important lorsqu'il s'agit d'enquêtes, car les résultats et l'analyse tirée de toute enquête ne seront pas fiables sans atteindre un taux de réponse minimum approprié. En général, le taux approprié est considéré comme étant d'environ 15 % à 30 %. Cependant, cela peut être inférieur lorsque la taille de l'échantillon est grande ou très supérieure lorsque la taille de l'échantillon est petite (Lund, B., 2021). Ainsi, l'objectif est de minimiser le fardeau de réponse afin d'atteindre un taux de participation plus élevé et de recueillir des réponses de meilleure qualité. Pour atteindre cet objectif, la première étape consiste à identifier les facteurs qui influencent le fardeau de réponse. Ensuite, une métrique doit être identifiée afin de quantifier le fardeau de réponse en fonction des caractéristiques de l'enquête. Enfin, un modèle sera développé pour estimer le taux de participation en fonction du fardeau de réponse mesuré. Ce modèle devrait être applicable à différentes enquêtes. L'enquête principale sur laquelle le modèle sera testé est l'enquête sur la justice en transport. Le projet « Mobilizing Justice » est financé par le Conseil de recherches en sciences humaines du Canada (CRSH) sous la direction de Steven Farber (Université de Toronto). Cette enquête vise à lutter contre les inégalités de transport et l'exclusion sociale en ciblant les personnes à faible revenu.

2.10.2 Principales questions de recherche

- Quels facteurs augmentent le fardeau de réponse ?
- Comment quantifier le fardeau de réponse ?
- Comment concevoir une enquête afin de minimiser le fardeau de réponse et d'augmenter le taux de participation ?
- Comment établir un lien entre le fardeau de réponse mesuré et le taux de participation ?

2.10.3 État des connaissances

Les caractéristiques de l'enquête sont les facteurs qui déterminent le fardeau de réponse. Moore, R. et al (2021), ont découvert que l'intégration d'un élément d'enquête dans le message d'invitation et l'envoi de rappels amusants augmenteront la probabilité d'engager les participants dans l'enquête. Dans certains cas, les personnes répondront à l'enquête, et certaines peuvent abandonner en cours d'enquête, mais dans l'ensemble, les participants répondront à plus de questions et davantage de données seront recueillies. Ceci est plus utile dans les enquêtes qui n'exigent pas qu'une réponse complète soit fournie.

Selon Groves et al, (2006), l'intérêt des participants pour le sujet de l'enquête est très déterminant sur la participation à l'enquête.

Plutôt que l'enquête elle-même et le fardeau qui lui est associé, le mode de contact s'avère efficace sur le taux de réponse. Sakshaug, J.W et al (2019), ont constaté qu'une invitation papier et un rappel papier sont corrélés à un taux de réponse plus élevé et à une quantité moindre d'éléments de non-réponse. L'envoi d'une invitation par courriel suivi d'un rappel papier est également associé à un taux de réponse élevé.

Revilla, M., (2017), a constaté que le format des questions et la longueur de l'enquête sont parmi les facteurs importants affectant le taux de complétion de l'enquête. Le format de grille, les questions ouvertes et les questions à réponses multiples dans lesquelles les répondants doivent sélectionner plusieurs options se révèlent plus contraignants. Smyth et al., (2006) ont recommandé d'éviter les questions à réponses multiples et de demander plutôt aux répondants de dire oui ou non à chaque élément.

Rudžionienė, J., et al, (2018), ont constaté que le sujet et la durée de l'enquête ont l'impact le plus significatif sur l'abandon d'une enquête. Son étude a également conclu que la personnalisation de l'invitation et la conception des enquêtes pour qu'elles soient plus adaptées aux téléphones mobiles et aux tablettes sont également très efficaces pour garder les répondants plus accros. Goritz (2014) et Yan et al. (2011) suggèrent également que la durée de l'enquête était positivement corrélée avec l'abandon des enquêtes. Toutefois, Rolstad, S., et al., (2011), estiment que bien que la longueur du questionnaire soit un facteur important du taux d'achèvement, un questionnaire court avec des options complexes peut parfois être plus contraignant.

Briz-Redón (2021), a examiné les caractéristiques sociodémographiques des répondants et leur association avec le taux de non-réponse et de réponse négligente indiquant qu'aux âges plus avancés (ceux âgés de 65 ans et plus), le taux d'éléments de non-réponse augmentera tandis que les répondants de moins de 40 ans semblent être davantage des répondants négligents.

Une étude précédente de Verzosa, (2020), a montré que l'ajout d'éléments ludiques et d'idées innovantes dans la conception de l'enquête gardera les répondants intéressés à répondre aux enquêtes.

Bradburn (1978), a développé le premier modèle associé au fardeau de l'enquête en tenant compte de quatre éléments : la durée de l'enquête, l'effort que les répondants doivent déployer pour remplir l'enquête, la fréquence d'être interrogé, ainsi que le stress et l'inconfort du répondant pendant l'enquête.

Un autre modèle associé au fardeau de réponse a été proposé par Fisher S., et al. (2001), qui considère 32 éléments regroupés en 3 catégories : le fardeau du répondant (associé aux attitudes des répondants), le fardeau de l'enquête (associé à la conception de l'enquête) et le fardeau de l'interaction (confrontation entre les répondants et l'enquête).

Schmid, B. et al. (2019), ont introduit un système pour mesurer le fardeau de réponse en fonction des caractéristiques de l'enquête telles que le type et la complexité des questions et le temps nécessaire pour répondre à chaque question. Ensuite, une relation est proposée pour estimer le taux de réponse en fonction du fardeau de l'enquête. Ces auteurs ont également constaté que le fardeau de réponse peut être atténué par l'existence d'incitatifs monétaires.

2.10.4 Objectifs de recherche

Compte tenu de tous les facteurs mentionnés ci-dessus, le premier objectif est de trouver des paramètres afin de quantifier la qualité de l'enquête en fonction de ses caractéristiques. En termes simples, une valeur associée au fardeau de réponse sera attribuée à une enquête en fonction de sa durée, du nombre de questions, du type de questions utilisées, de leur complexité et d'autres caractéristiques de l'enquête. Des valeurs plus élevées indiquent que plus d'efforts sont nécessaires pour répondre à l'enquête, de sorte que le fardeau du répondant sera plus élevé. Ensuite, un modèle peut être élaboré pour estimer le taux de participation en fonction du fardeau de réponse mesuré. Il vise à minimiser le fardeau de réponse et à augmenter la participation sans faire de compromis en ce qui concerne la qualité des données.

Comme mentionné précédemment, l'un des facteurs les plus efficaces sur le taux de participation aux enquêtes est le sujet de l'enquête et la manière dont les questions sont posées. Ainsi, le deuxième objectif ici est de réfléchir et de développer des idées sur la conception de l'enquête sur la justice en transport. Selon l'équipe de projet lors d'une séance de remue-méninges tenue en octobre 2021, l'enquête abordera 5 thèmes : "aspiration à la mobilité, obstacles et contraintes, résultats de vie et de bien-être, satisfaction à l'égard des déplacements, et déplacements supprimés et excessifs". Le modèle sera ensuite utilisé pour mesurer le fardeau et le taux de participation et afin d'assurer que l'enquête la plus efficace soit conçue à la fin. Différentes questions seront conçues pour répondre à chacun des critères mentionnés et une tentative sera entreprise pour minimiser le fardeau de cette enquête.

2.11 Impacts de scénarios de télétravail

- Par : Hubert Verreault, ing., associé de recherche
- Contexte : Chaire Mobilité

2.11.1 Introduction

Le transport est le plus grand contributeur des émissions de GES au Québec. Quand on s'intéresse aux comportements de mobilité de la population, différentes stratégies peuvent être mises de l'avant afin de diminuer les émissions, le nombre de déplacements et les distances motorisées. Ultiment, les stratégies possibles visent principalement à diminuer les passagers-km motorisés. La virtualisation des activités, particulièrement celles concernant le travail, permet de s'attaquer à cette problématique en visant la réduction du nombre de déplacements effectués pour le travail. Les impacts du télétravail ont été étudiés par plusieurs auteurs dans les dernières décennies. Bien que les externalités positives soient bien connues, le télétravail demeurait marginal au Québec avant la COVID-19. Bien qu'il soit difficile d'évaluer avec précision les impacts du télétravail dans les comportements de mobilité, il est évident qu'il a un impact important sur les véhicules-km motorisés sur les réseaux.

En ne sachant pas trop comment le télétravail va perdurer dans les prochaines années dans la population, il est intéressant de se questionner sur le développement d'une méthodologie afin d'évaluer les impacts du télétravail sur les comportements de mobilité et sur les émissions de GES. Cette étude vise donc à tester différents scénarios d'adoption du télétravail afin d'estimer la limite supérieure de ces impacts pour la grande région de Montréal (GRM).

2.11.2 État des connaissances

La littérature sur le télétravail est abondante et aborde différents sujets (taux d'adoption, fréquence, impacts et externalités). Nos expérimentations se concentrent sur les impacts du télétravail sur les comportements de mobilité et les conditions de circulation sur les réseaux. Bien que le télétravail fût en augmentation avant la pandémie dans plusieurs pays (Ojala. et Pyöriä (2018), Vilhelmson. et Thulin, (2016), Melo et Abreu (2017)), il demeurait assez peu fréquent au Canada (Lachapelle et al. (2018)).

Plusieurs auteurs se sont historiquement intéressés aux impacts du télétravail sur les conditions de circulation. Au niveau individuel, il semble y avoir un consensus dans la littérature que le télétravail diminue les véh-km motorisés dans les périodes de pointe (Andreev et al. (2010), Vanoutrive et al. (2010)). Plusieurs soulignent également que le télétravail entraîne une diminution dans la quantité de déplacements effectués et de la probabilité d'utiliser une automobile, entraînant une augmentation de l'utilisation des modes actifs (Lachapelle et al. (2018), Elldér (2020), Tavares (2017)). Cependant, Abreu et al. (2018) soulignent que les ménages contenant un seul travailleur entraînent une augmentation des distances parcourues quand ce seul travailleur effectue du télétravail. Le télétravail peut également entraîner plusieurs autres externalités positives sur le transport comme une augmentation de la sécurité et des bénéfices sur la santé (Tavares (2017), Chakrabarti (2018)).

Cependant, les impacts du télétravail sont moins consensuels au niveau collectif. Il semble ne pas y avoir de consensus auprès des chercheurs par rapport aux impacts du télétravail sur le transport : génération de déplacements supplémentaires, augmentation de la distance entre le domicile et le lieu de travail, etc. (Andreev et al. (2010), Abreu et al. (2018), Chakrabarti (2018)). Selon certains auteurs, les bénéfices de l'adoption du télétravail par la population sont moins importants que l'on pourrait penser (Kim (2017), Larson et Zhao (2017)). L'adoption du télétravail

peut également avoir un impact sur la quantité de GES émis (Menezes et al. (2017), Shabanpour et al. (2018)). Cependant, certains auteurs nuancent également les impacts du télétravail sur les émissions de GES (Zhu et Mason (2014)).

Bien que le télétravail devrait entraîner une réduction des véhicules-km parcourus, quelques études ont démontré que le télétravail peut entraîner une augmentation des déplacements (de Vos et al. (2018)) et des distances (Melo et Abreu (2017)). L'adoption du télétravail peut également entraîner une demande latente de déplacements étant donné que le télétravail libère du temps pendant la journée pour le travailleur (Eldér (2020), Hilbrecht et al. (2013)). Premièrement, selon la structure de la chaîne de déplacements, certains déplacements supplémentaires devront être effectués si des activités doivent toujours être effectuées malgré le télétravail (Andreev et al. (2010), Choo et al. (2005)). Dans ce contexte, il est important de considérer l'ensemble des activités journalières de l'ensemble des membres d'un ménage et non seulement l'activité travail (Melo et Abreu (2017), Abreu (2018), Kim et al. (2015)). Deuxièmement, le temps libéré pour le télétravailleur ainsi que l'amélioration générale des conditions des réseaux si un fort pourcentage de travailleurs adoptent le télétravail peut entraîner une augmentation des taux de déplacements et des activités pour l'ensemble de la population (Zhu (2012)). Finalement, la diminution dans le nombre de déplacements effectués pour le travail peut également avoir un impact sur le choix du mode de transport en rendant l'automobile plus attractive (Kitou et Horvath (2006)). Ce transfert modal, s'il se produit, va automatiquement réduire les bénéfices sur les réseaux de l'adoption du télétravail. Néanmoins, le télétravail peut également entraîner une augmentation de l'utilisation des modes actifs étant donné que les travailleurs ont plus de temps pendant la journée et que les déplacements restants dans les chaînes d'activités peuvent être de moins longues distances (Lachapelle et al. (2018)). Cependant Boell et al. (2013) soulignent que les modes actifs peuvent également être pénalisés selon la localisation de la résidence.

En plus des impacts directs de l'adoption du télétravail sur les comportements de mobilité, la localisation résidentielle peut aussi être impactée par le choix de faire du télétravail. Le télétravail peut donc amplifier le phénomène d'étalement urbain pour ces travailleurs (Zhu (2013, Rhee (2009)). Il est également démontré que les télétravailleurs travaillent, en moyenne, à une plus grande distance de leur lieu de travail habituel (Ravalet et Rérat (2019), Eldér (2014)). Il y a plusieurs constats conflictuels à propos des impacts du télétravail au niveau collectif. Litman (2011) suggère d'être prudent dans l'évaluation de ses impacts.

2.11.3 Méthodologie

Les données utilisées dans le cadre de cette étude se basent principalement sur l'enquête OD 2018. L'ensemble du territoire de l'enquête a été pris en considération. Les informations sur les déplacements incluent les lieux des activités effectuées pour le travail sont utilisées. En plus de ces informations, l'occupation principale est également utilisée.

La méthodologie proposée vise à évaluer les impacts, sur différents indicateurs, de différents scénarios de télétravail. Ces derniers sont estimés à partir de l'enquête OD en simulant un certain pourcentage des travailleurs qui effectuent du télétravail pour chaque journée de la semaine, soit du lundi au vendredi. Pour chaque journée de la semaine, les télétravailleurs sont sélectionnés aléatoirement parmi l'ensemble des travailleurs. Il est de plus supposé que ce télétravail s'effectue de la maison. En effectuant du télétravail, certains déplacements de ces travailleurs, principalement ceux reliés à la chaîne dont le travail est l'activité principale, n'ont plus de raisons d'exister. Cependant, certains nouveaux déplacements peuvent être nécessaires afin d'assurer la

mobilité de tous les membres du ménage. L'objectif de la méthodologie est d'évaluer les impacts de cette transformation des chaînes d'activité pour le travail sur le nombre de déplacements et les distances effectuées ainsi que sur les émissions de GES.

Afin d'évaluer les impacts du télétravail sur une semaine type d'activité, lundi au vendredi, la méthodologie de sélection aléatoire des télétravailleurs est répétée pour chaque jour de la semaine. La méthodologie proposée se décompose en différentes étapes:

- Extraction des travailleurs de l'enquête OD ;
- Pour tous les jours de la semaine (i de 1 à 5) ;
 - Sélection aléatoire des travailleurs assignés au télétravail pour le jour i selon un certain pourcentage parmi les travailleurs ;
 - Extraction des déplacements et des chaînes dont le travail est l'activité principale ;
 - Altération des déplacements et des chaînes afin d'éliminer le lieu de travail et les activités non-contraignantes ;
- Estimation des impacts du scénario de télétravail.

La méthodologie est également illustrée à la Figure 2-35.

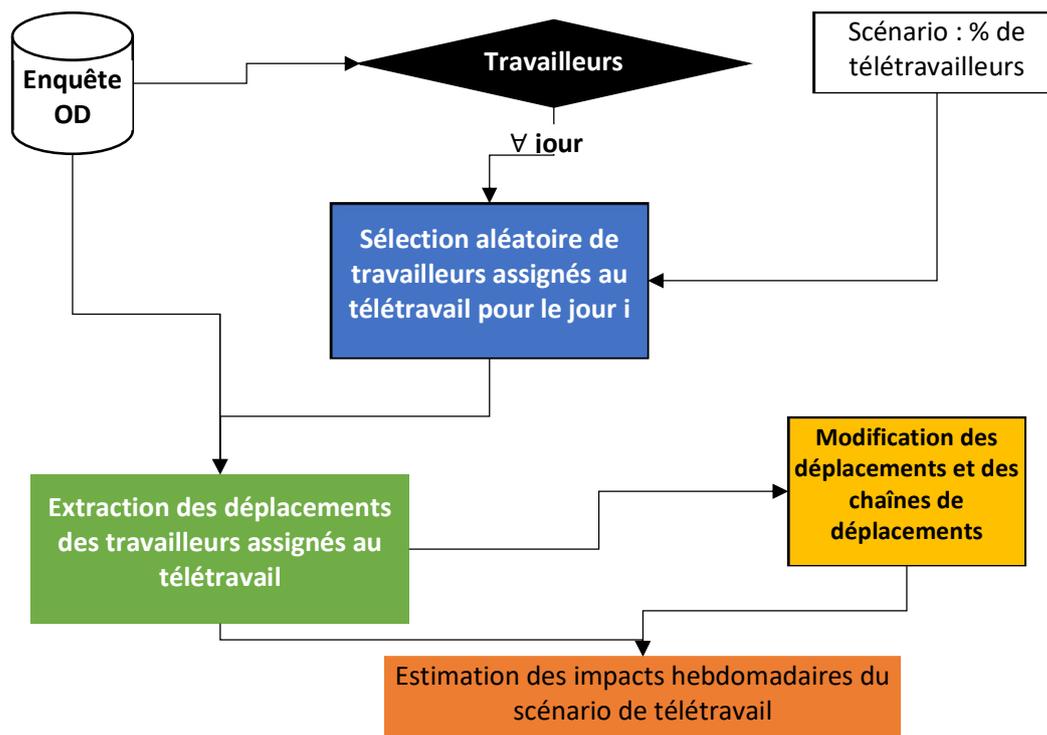


Figure 2-35 Étapes de la méthodologie

La méthodologie proposée se base sur plusieurs hypothèses simplificatrices:

1. Seulement les jours de semaine sont considérés ;
2. Outre les chaînes Travail effectuées en télétravail, la demande en transport demeure constante :

- a. Aucune modification des comportements de mobilité n'est appliquée à la suite de la réduction du nombre de déplacements sur les réseaux.
 - i. Les heures de départ des déplacements, les durées d'activités et les modes de transport ne seront pas modifiés.
 - b. Les déplacements contraints ne sont pas transférables entre les membres du ménage.
3. 100% de la population de travailleurs possède un emploi qui est propice au télétravail.
 4. Le télétravail est distribué uniformément pendant les jours de la semaine.

2.11.3.1 Définition

Le terme télétravail utilisé dans cette étude réfère à une activité travail effectuée à la maison. Bien qu'il soit possible de faire du télétravail dans un autre endroit que la résidence, cela n'a pas été considéré dans cette étude. Un travailleur est défini comme une personne dont l'occupation principale déclarée dans l'enquête OD est travailleur temps plein ou temps partiel. De plus, les personnes dont l'occupation principale n'est pas parmi celles-ci mais qui ont déclaré au moins un déplacement pour le travail ont été considérées comme des travailleurs. Une chaîne d'activité est définie comme une série de déplacements qui commence et finit à la résidence. Le point d'ancrage de la chaîne est donc le lieu de domicile. Seules les chaînes d'activités fermées ont été considérées dans les calculs. Le motif principal d'activité de la chaîne est le travail si au moins un déplacement pour ce motif a été effectué parmi les déplacements de la chaîne. Ce motif a préséance sur tous les autres motifs d'activité de la chaîne. Une activité contrainte est définie comme étant une activité qui ne peut pas être déplacée dans le temps. Ces activités ne peuvent donc pas être éliminées du programme d'activité d'un travailleur, même s'il effectue du télétravail. Il est supposé que les autres types d'activités puissent être déplacés à un autre moment dans la journée ou à une autre journée.

2.11.3.2 Scénarios de télétravail

Il est assez difficile de déterminer un scénario plausible de pourcentage de télétravail parmi les travailleurs. La COVID-19 a également démontré qu'une large partie des travailleurs peuvent désormais faire du télétravail de façon quasi-continue. Cette étude, dont les premiers calculs datent de 2019, vise donc à évaluer différents scénarios ayant chacun un pourcentage de télétravailleurs spécifique. Ces pourcentages varient de 10% à 50% par tranche de 10%. Il est également supposé que le télétravail est distribué uniformément pendant la semaine.

2.11.3.3 Altération des chaînes de déplacements

Effectuer une activité travail à la maison plutôt qu'au lieu de travail habituel va automatiquement changer la chaîne de déplacement habituellement effectuée. Plusieurs déplacements n'auront plus de raisons d'être effectués (activités non-contraintes) tandis que d'autres devront toujours être effectués (activités contraintes). De plus, certains nouveaux déplacements devront peut-être être effectués afin de pouvoir effectuer les activités contraintes. Par exemple, si un parent va porter son enfant à la garderie avant d'aller travailler, il devra quand même aller le porter et retourner à la maison ensuite lors d'une journée de télétravail. Il faut donc modifier les chaînes d'activités afin de prendre en considération ces activités contraintes. Les chaînes dont le motif principal n'est pas le travail ne sont pas altérées et demeurent donc identiques. La méthodologie proposée pour altérer les chaînes d'activité se décompose en plusieurs étapes :

Pour chaque travailleur assigné en télétravail :

- Extraire les chaînes d'activités dont le motif principal est le travail

Pour chaque déplacement extrait :

- Extraire les activités contraintes de la chaîne
- Créer des nouveaux déplacements permettant aux activités contraintes d'être effectuées.
 - Une nouvelle chaîne d'activité est créée avant l'activité travail si au moins une activité contrainte est effectuée avant l'activité travail.
 - Une nouvelle chaîne d'activité est créée après l'activité travail si au moins une activité contrainte est effectuée après l'activité travail.

À la fin du processus, si au moins une activité contrainte est effectuée, de nouveaux déplacements sont générés afin de remplacer les initiaux. Si aucune activité contrainte n'est effectuée dans la chaîne, la chaîne de déplacement est éliminée complètement. Chaque nouvelle chaîne d'activité commence et finit au domicile. S'il y a plus d'une activité contrainte effectuée avant ou après l'activité travail, la chaîne de déplacement créée permet d'effectuer l'ensemble de ces activités.

Ce processus se base sur plusieurs hypothèses. Premièrement, il est supposé que le mode initial utilisé pour faire les déplacements de la chaîne ne va pas changer et va être utilisé pour faire les nouveaux déplacements. En pratique, changer les paramètres de la chaîne de déplacements peut engendrer des changements sur les modes utilisés. Le mode initial utilisé dans la chaîne n'est plus nécessairement le mode optimal pour la nouvelle chaîne d'activité. Il est également supposé que l'heure de départ et les durées des activités contraintes ne peuvent pas être déplacées dans le temps. Il est aussi supposé que les autres membres du ménage ne puissent pas remplacer le télétravailleur pour ces activités contraintes. Cette hypothèse est importante, spécifiquement pour les déplacements de types reconduire/aller chercher quelqu'un.

La Figure 2-36 illustre un exemple de l'altération d'une chaîne de déplacement d'un travailleur assigné au télétravail.

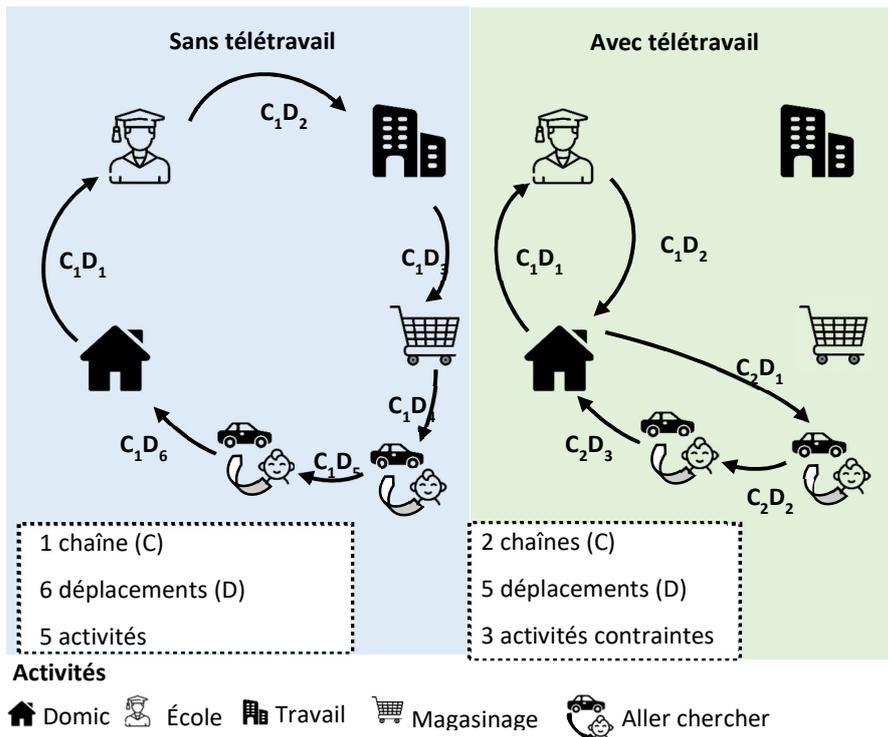


Figure 2-36 Exemple de la modification des déplacements et des chaînes de déplacements pour un travailleur assigné au télétravail

Dans cet exemple, le travailleur fait initialement une chaîne de déplacements lui permettant de faire son activité travail. Cette chaîne contient 6 déplacements et 5 activités. De ces activités, l'étude et aller chercher/reconduire sont des activités contraintes. Si ce travailleur faisait du télétravail, il n'aurait plus besoin de se rendre à son lieu de travail habituel et ces lieux d'activités non-contraints. Seuls les lieux d'activités contraints resteraient dans son horaire. Étant donné que ces activités contraintes se font avant et après avoir visité le lieu habituel d'activité, deux nouvelles chaînes d'activités seront construites.

2.11.3.4 Estimation des impacts

Afin d'évaluer les impacts des différents scénarios expérimentés, plusieurs indicateurs sont évalués en les comparant avec le scénario de référence qui correspond à l'état de 2018. Les indicateurs évalués incluent le nombre de déplacements ainsi que les distances effectuées selon le mode de transport. Des distances réseaux ont été calculées pour chaque déplacement. Pour les déplacements motorisés, les calculs de chemins sont effectués sur le réseau routier en corrigeant les temps de parcours sur les liens avec des données GPS. Ces calculs permettent également d'avoir accès à un itinéraire probable pour les déplacements et donc, d'évaluer la charge sur différentes infrastructures routières dont les ponts. Finalement, la variation des GES émis entre les scénarios est également estimée. À ce propos, la méthodologie suivante est appliquée :

- Pour chacun des modes de transports, un taux d'émission de GES par passager-km est estimé. Ce taux dépend de ces facteurs
 - Type de véhicule (dépend de la composition de la flotte)
 - Occupation des véhicules privés et publics
 - Conditions de circulation sur le réseau

- Estimation pour chaque déplacement des GES émis en fonction de la distance estimée et du taux d'émission estimé.

Cette procédure se base sur plusieurs hypothèses. Premièrement, la consommation d'essence moyenne des automobiles est estimée à 10,5 L/100 km. La transformation de ce taux en un taux par passager/km est obtenue en fonction du taux d'occupation moyen des automobiles en fonction de la période horaire (1,11 pour la pointe du matin par exemple). L'association du taux de consommation moyen et des taux d'occupation des véhicules en fonction de la période horaire donne un taux de 9,0 L/passager-km pour les déplacements en automobiles (auto-conducteur ou auto-passager)

Le taux de consommation moyen pour le transport en commun implique également certaines hypothèses sur l'occupation des véhicules. Il est supposé que l'occupation des véhicules est de 10 passagers par véhicule pendant les périodes hors-pointe et de 30 passagers par véhicule pendant les périodes de pointe. La consommation moyenne obtenue est donc de 2,2 L/100km. C'est certainement une surestimation du taux de consommation des services de transport en commun étant donné que le métro est considéré comme de l'autobus dans le cadre de ces calculs préliminaires.

2.11.4 Résultats

L'application de la méthodologie à l'échelle de la GRM permet d'évaluer les impacts potentiels d'une adoption importante du télétravail chez les travailleurs. Basés sur ces résultats, plusieurs types d'analyses peuvent être conduites. Premièrement, les impacts des scénarios sur les nombres des déplacements, sur les distances et sur les émissions de GES segmentés selon le mode sont présentés. Par la suite, la distribution horaire de la charge sur l'ensemble des ponts permettant d'accéder à l'Île de Montréal est présentée.

L'adoption du télétravail chez un certain pourcentage de travailleur implique une réduction du nombre de déplacements, des distances et des GES. Le Tableau 2-16 présente la réduction en pourcentage des nombres de déplacements et des distances parcourues selon le scénario de télétravail. Comme attendu, la réduction est de plus en plus importante en augmentant le pourcentage de télétravailleurs. Pour le scénario S10% (10% des travailleurs font du télétravail), le nombre de déplacements et les distances parcourues diminuent respectivement de 3,5% et de 5,0%. Si on s'intéresse uniquement aux travailleurs, ces diminutions sont respectivement de 6,1 et de 7,3%. En segmentant selon le mode utilisé, les diminutions les plus importantes sont le transport en commun. Cependant, si on calcule les impacts en nombre de déplacements et non en pourcentage, c'est l'automobile-conducteur qui a la plus forte diminution. Pour les diminutions des GES, le pourcentage qui est relié à l'utilisation de l'automobile est constant selon le scénario à 95,8%.

Tableau 2-16 Réduction en pourcentage des nombres de déplacements et des distances selon les scénarios

Élasticité (β)	Population	Mode	Scénarios de télétravail (X% de télétravailleurs)				
			S10%	S20%	S30%	S40%	S50%
Déplacements	Tous	Tous	-3,6%	-7,2%	-10,9%	-14,5%	-18,1%
		Tous	-6,1%	-12,2%	-18,4%	-24,6%	-30,6%
	Travailleurs	Auto-conducteur	-5,9%	-11,8%	-17,7%	-23,6%	-29,4%
		Auto-passager	-4,3%	-8,8%	-13,7%	-18,1%	-22,2%
		Marche ou vélo	-4,8%	-9,6%	-14,5%	-19,4%	-24,2%
		Transport en commun	-8,4%	-16,8%	-25,2%	-33,8%	-42,1%
Distances	Tous	Tous	-5,0%	-10,1%	-15,2%	-20,3%	-25,4%
		Tous	-7,3%	-14,7%	-22,2%	-29,7%	-37,0%
	Travailleurs	Auto-conducteur	-7,3%	-14,7%	-22,1%	-29,6%	-36,9%
		Auto-passager	-4,9%	-9,5%	-14,5%	-19,4%	-23,8%
		Marche ou vélo	-4,7%	-9,8%	-15,8%	-20,6%	-26,5%
		Transport en commun	-8,7%	-17,5%	-26,2%	-35,2%	-43,8%
GES	Tous	-5,2%	-10,5%	-15,8%	-21,2%	-26,4%	
	Travailleurs	-7,2%	-14,5%	-21,7%	-29,1%	-36,3%	
	% de la réduction de GES lié à l'utilisation de l'automobile	95,8%	95,8%	95,8%	95,8%	95,8%	

Comme attendu, plus le pourcentage de télétravailleurs est élevé, plus les diminutions des indicateurs sont importantes. Dans ce contexte, il est intéressant de comparer les ratios de diminution entre les scénarios de télétravailleur. Ce qui correspond à estimer l'élasticité de la demande en fonction de la variation dans le pourcentage de télétravailleurs.

$$\beta = \frac{\Delta (\text{Demande})}{\Delta (\% \text{ de télétravailleurs})}$$

Où la demande en transport correspond à la variation du nombre de déplacements ou de la distance parcourue selon les scénarios.

Les coefficients d'élasticité sont presque constants selon les scénarios étudiés. Ce résultat était attendu étant données les hypothèses utilisées dans la méthodologie. Premièrement, les télétravailleurs sont sélectionnés aléatoirement parmi les travailleurs et sont supposés uniformément distribués sur l'ensemble du territoire. Deuxièmement, il est supposé que la demande en transport, autre que celle liée aux chaînes d'activités travail, ne varie pas pendant la journée. Ces aspects expliquent partiellement l'élasticité constante observée. Cependant, cette

dernière varie en fonction du mode de transport utilisé. Le Tableau 2-17 présente les coefficients d'élasticité pour la GRM et pour les modes de transport évalués.

Tableau 2-17 Coefficient d'élasticité des nombres de déplacements et des distances parcourues en fonction des scénarios et de la région du domicile

Élasticité (β)	Population	Mode	Régions					
			GRM	Île de Montréal			Laval et Agglo. Longueuil	Couronnes
				Centre-Ville	Centre	Est et Ouest		
Déplacements	Tous	Tous	-0,36	-0,41	-0,37	-0,34	-0,34	-0,38
	Travailleurs	Tous	-0,61	-0,65	-0,60	-0,60	-0,60	-0,63
		Auto-conducteur	-0,59	-0,68	-0,55	-0,55	-0,56	-0,63
		Auto-passager	-0,44	-0,50	-0,43	-0,48	-0,40	-0,44
		Marche ou vélo	-0,48	-0,55	-0,45	-0,48	-0,50	-0,54
		Transport en commun	-0,84	-0,82	-0,81	-0,88	-0,91	-0,83
Distances	Tous	Tous	-0,51	-0,45	-0,47	-0,49	-0,48	-0,53
	Travailleurs	Tous	-0,74	-0,69	-0,69	-0,75	-0,72	-0,76
		Auto-conducteur	-0,74	-0,79	-0,67	-0,72	-0,69	-0,76
		Auto-passager	-0,48	-0,37	-0,41	-0,52	-0,42	-0,52
		Marche ou vélo	-0,51	-0,41	-0,55	-0,44	-0,43	-0,37
		Transport en commun	-0,87	-0,93	-0,84	-0,92	-0,91	-0,84
GES	Tous	-0,53	-0,53	-0,47	-0,50	-0,56	-0,44	
	Travailleurs	-0,72	-0,76	-0,64	-0,69	-0,68	-0,75	

Les coefficients d'élasticité sont intéressants parce qu'ils permettent de comparer la sensibilité des indicateurs et des modes. Basé sur le tableau précédent, il est possible de conclure que chaque addition de 1% de télétravailleurs va entraîner une diminution de 0,36% du nombre de déplacements effectués et de 0,51% des distances parcourues dans la GRM. Si on s'intéresse plus précisément aux travailleurs, une hausse de 1% de télétravailleur va entraîner une diminution de 0,61% du nombre de déplacements et de 0,74% des distances parcourues par les télétravailleurs. Le transport en commun est le mode qui a la plus grande sensibilité avec respectivement une diminution de 0,84% et de 0,87% pour le nombre de déplacements et les distances parcourues pour une augmentation de 1% de télétravailleurs.

Les coefficients d'élasticité en fonction de la région de résidence sont assez constants pour la variation du nombre de déplacements. L'hypothèse de distribution uniforme des télétravailleurs explique partiellement ce résultat. Cependant, pour les distances parcourues, les coefficients d'élasticité augmentent habituellement en fonction de la distance au Centre-Ville.

Comme mentionné précédemment, le mode auto-conducteur est celui qui entraîne la plus grande diminution dans le nombre de déplacements pour une augmentation de télétravailleurs. Il est donc intéressant de regarder l'impact du télétravail sur la charge observée sur les réseaux. Dans le cadre de cette étude, la charge sur l'ensemble des ponts permettant d'accéder ou de quitter l'Île de Montréal a été estimée pour chaque scénario.

La Figure 2-37 illustre la distribution horaire du taux d'utilisation des ponts pour accéder à l'Île de Montréal selon le pourcentage de télétravailleurs. La capacité des ponts a été estimée à partir de la charge maximale observée pendant la journée sur chacun des ponts.

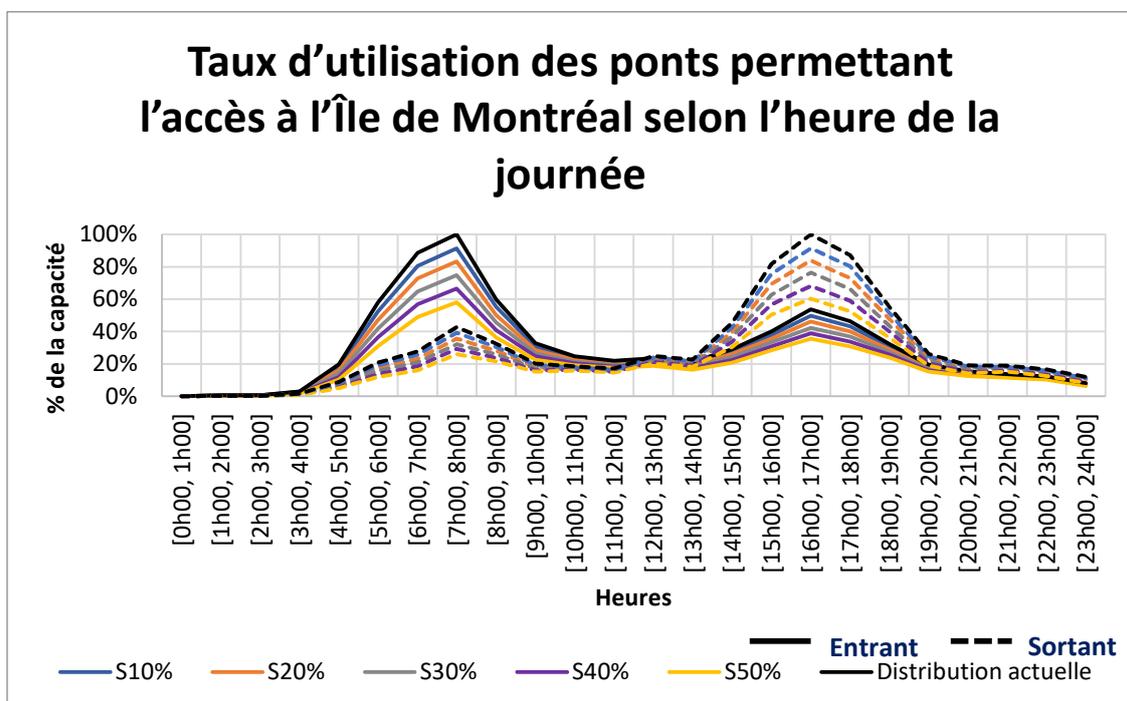


Figure 2-37 Taux d'utilisation par rapport à la capacité des ponts selon l'heure de la journée

Selon notre méthodologie, en 2018, 26,3% de la capacité des ponts est utilisé pour un jour moyen de semaine. Selon le scénario, le taux d'utilisation moyen des ponts diminue à 23,7% pour un pourcentage de télétravailleurs de 10% et à 17,5% pour un pourcentage de 50%. Comme attendu, les impacts les plus importants se retrouvent pendant les périodes de pointe. Pour le scénario avec 10% de télétravailleurs, il y a une diminution de 8,6% pendant l'heure de pointe du matin et de 8,5% pour celle de l'après-midi. En estimant les coefficients d'élasticité pour les heures de pointe, chaque augmentation de 1% de télétravailleurs entraîne une diminution de 0,84% et de 0,81% respectivement pour les directions entrantes et sortantes de l'Île. Ce résultat est étroitement lié avec l'hypothèse que la demande en transport demeure constante malgré les améliorations des conditions routières.

2.11.5 Conclusion

Cette étude visait à présenter une méthodologie apte à estimer la limite supérieure des impacts, dans une perspective de transport, d'une adoption importante du télétravail. La méthodologie présentée se base principalement sur la modification des chaînes d'activités qui incluent le travail chez les travailleurs. Ces chaînes d'activités sont modifiées afin de prendre en considération que l'activité travail est effectuée au domicile. S'il est supposé que la demande en transport demeure constante malgré la forte adoption du télétravail, les résultats suivants sont obtenus :

- La plus grande part des diminutions observées par l'adoption du télétravail proviennent des déplacements auto-conducteur.
- Une augmentation de 1% de télétravailleurs peut entraîner une diminution de 0,36% du nombre de déplacements effectués et de 0,51% des distances parcourues
- Une augmentation de 1% de télétravailleurs peut également entraîner une diminution de 0,53% des GES émis. 95,6% de cette diminution est reliée à l'utilisation de l'automobile.

2.11.5.1 Limitations et perspectives

La méthodologie proposée possède plusieurs limitations et perspectives. Premièrement, elle s'appuie sur plusieurs hypothèses simplificatrices qui ont toutes un impact sur les résultats obtenus. L'hypothèse la plus importante est que la demande en transport demeure constante. Il est donc supposé que les personnes ne modifient pas leurs comportements de mobilité malgré l'amélioration des conditions sur les réseaux et du temps libéré pendant leur journée. Cela soulève cependant un enjeu lié à l'adoption du télétravail. Si l'objectif est de maximiser les retombées, il est important de réduire parallèlement l'attractivité de l'automobile pour éviter un effet rebond à savoir un transfert modal vers l'automobile et la génération de nouveaux déplacements en automobile là et quand une capacité s'est dégagée. Il est également supposé que 100% des travailleurs pourraient effectuer du télétravail. Cette hypothèse est loin d'être vraie. De plus, la distribution spatiale des emplois pouvant se faire en télétravail n'est pas nécessairement la même que celle des emplois ne pouvant se faire en télétravail. De plus, afin de s'assurer de maximiser les retombées du télétravail, il faudrait également que ces épisodes de travail à la maison soient distribués uniformément pendant la semaine. Si l'ensemble des travailleurs décident d'effectuer du télétravail pendant la même la même journée, les impacts ne seront pas les mêmes.

2.12 Fusion inédite de données sur les véhicules et constats

- Par : Jean-Simon Bourdeau, ing., associé de recherche
- Contexte : Chaire Mobilité – collaboration avec Équiterre

2.12.1 Introduction

Le parc automobile a grandement évolué au Québec depuis les dernières décennies. Non seulement le nombre de véhicules est en augmentation, mais les propriétés de ceux-ci ont évolué avec les années.

Dans cette analyse, les données sur les véhicules immatriculés de la Société de l'assurance automobile du Québec (SAAQ) sont utilisées afin de quantifier l'évolution des propriétés des véhicules. Ces données comprennent certains attributs sur les véhicules comme la masse, la cylindrée ou le nombre d'essieux, mais d'autres informations comme les dimensions des véhicules et leur consommation énergétique sont manquantes. Les données de la SAAQ sont heureusement ouvertes depuis quelques années, cependant il peut s'avérer difficile de fusionner des ensembles de données provenant de propriétaires différents en l'absence d'uniformisation des données. Le défi réside donc dans l'association des données de la SAAQ avec d'autres sources de données qui contiennent les propriétés manquantes. Étant donné qu'il n'existe pas d'identifiant commun aux ensembles de données pour les associer facilement, il est nécessaire de développer une méthodologie de fusion de données qui permettra pour tous les véhicules immatriculés d'avoir accès à l'ensemble de leurs propriétés.

2.12.2 Contexte

À partir de 2004, le véhicule utilitaire sport (VUS) devient la classe de véhicule la plus vendue au Canada (Magnusson, 2005). Entre 2000 et 2020, la catégorie des camions légers (VUS, les camionnettes, fourgonnettes) immatriculés au Québec a augmenté de 169 % alors que le nombre total de véhicules de promenade a augmenté de 43 %, faisant passer la proportion de camions légers de 23,7 % de l'ensemble des véhicules de promenade à 44,6 % des véhicules de promenade (SAAQ, 2020).

Plusieurs chercheurs se sont intéressés à l'évolution du parc automobile dans différents pays du monde. Cependant les efforts de recherche concernent surtout la motorisation des véhicules (Taplin, 2005; Brand, 2013), les émissions polluantes (Musti, 2011; Fontaras, 2012) ou les véhicules autonomes (Hars, 2005; Sylva, 2021). Quelques auteurs se sont aussi intéressés à la modélisation de parcs de véhicules, notamment à Singapour (Wei, 2015) et en Suède (Hugosson, 2016).

En ce qui concerne la consommation énergétique des véhicules légers qui sont vendus au Canada, Ressources Naturelles Canada dispose d'un outil de recherche en ligne qui permet de télécharger l'ensemble des cotes de consommation de carburant (Ressources Naturelles Canada, 2022). En 2009, le Canadian Vehicle Survey estimait que la consommation d'essence moyenne des véhicules légers au Canada est passée de 10.6 à 10.7 L / 100 km entre 2005 et 2009 (Ressources Naturelles Canada, 2011).

2.12.3 Méthodologie de fusion de données

Le Tableau 2-18 montre les sources de données utilisées pour développer la méthodologie de fusion de données. La première source de données contient les véhicules de promenade immatriculés au Québec entre 1990 et 2019. La deuxième source de données est la Canadian Vehicle Specification (CVS) de Transports Canada et contient les dimensions physiques de tous les véhicules autorisés à être vendus au Canada depuis 1994. La troisième source de données

concerne les cotes de consommation énergétique et le propriétaire est Ressources naturelles Canada (RNCAN). La première source de données contient l'univers des véhicules immatriculés, et celle-ci sera bonifiée à l'aide des deux autres sources de données.

Tableau 2-18 Sources de données utilisées pour la fusion de données

Données	Propriétaire	Couverture temporelle	Nombre d'enregistrements
Véhicules de promenade possédés	SAAQ	1990-2019	113,4 M
CVS	Transports Canada	1994-2019	20 613
Consommation énergétique	Ressources naturelles Canada (RNCAN)	2005-2019	15 963

Le Tableau 2-19 montre la description des variables disponibles dans les données de véhicules immatriculés de la SAAQ. Pour les présentes analyses, seulement les véhicules de promenade (CLAS = PAU) appartenant à des personnes physiques (TYP_DOSS_PERS = P) sont considérés.

Tableau 2-19 Description des variables disponibles dans les données de véhicules immatriculés de la SAAQ (SAAQ, 2022)

Nom de la variable	Description
NO_SEQ	Numéro séquentiel du véhicule
CLAS	Code identifiant le type du véhicule de même que le type d'utilisation qui en est faite.
TYP_VEH_CATEG_USA	Type de véhicule
MARQ_VEH	Marque du véhicule
MODEL_VEH	Modèle du véhicule
ANNEE_MOD	Année du modèle du véhicule
MASSE_NETTE	Masse nette en kilogramme (kg) du véhicule
NB_CYL	Nombre de cylindres du moteur du véhicule
CYL_VEH	Capacité en centimètre cube (cm ³) des cylindres d'un moteur à explosion
NB_ESIEU_MAX	Nombre maximal d'essieux qui peuvent supporter un véhicule
COUL_ORIG	Couleur d'origine du véhicule
TYP_DOSS_PERS	Type de dossier : Personne physique (P) ou morale (M)
PHYS_SEX	Sexe du propriétaire du véhicule
REG_ADM	Région administrative de résidence du propriétaire du véhicule
MRC	Municipalité régionale de comté de résidence du propriétaire d'un véhicule.
CG_FIXE	Code géographique de la municipalité de résidence du propriétaire d'un véhicule
NIV	Numéro d'identification du véhicule (NIV ou VIN)

La Figure 2-38 montre les dimensions qui sont disponibles dans la base de données CVS. Les données qui seront utilisées sont celles qui sont encadrées en rouge, c'est-à-dire la longueur totale (OL), la largeur totale (OW) et la hauteur totale (OH).

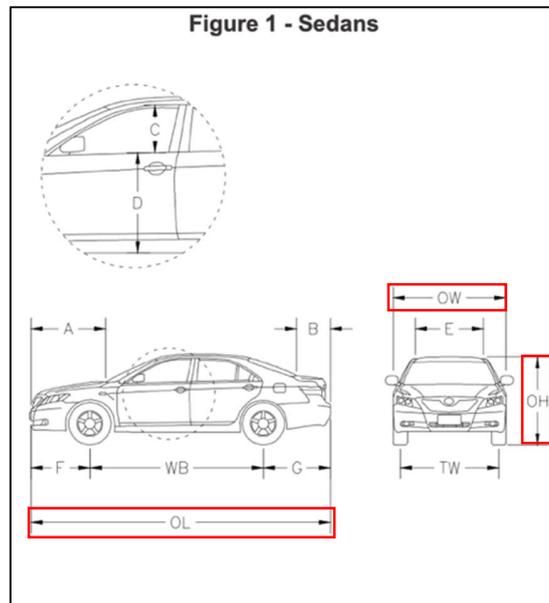


Figure 2-38 Dimensions des véhicules disponibles dans la base de données de la Canadian Vehicle Specification (CVS) (Transports Canada, 2012).

« Afin de mesurer le degré de similitude, les trigrammes sont utilisés. Un trigramme est une suite de trois caractères ou moins. La décomposition d'une chaîne de caractères en une série de trigrammes uniques permet d'estimer le degré de similarité entre deux chaînes de caractères même si elles n'ont pas le même nombre de caractères. Les trigrammes ne sont pas sensibles à la casse. De plus, seuls les caractères alphanumériques sont considérés. Finalement, deux espaces sont ajoutés en préfixe de chaque mot et un espace est ajouté en suffixe de chaque mot. La bibliothèque pg_trim du système de gestion de bases de données PostgreSQL a été utilisée pour créer les trigrammes et calculer le taux de similarité entre les chaînes de caractères⁵.

Par exemple, il peut arriver qu'un nom de modèle fasse référence à une gamme particulière du modèle, ou qu'une partie du nom du modèle soit manquante. L'exemple suivant a été illustré au Tableau 2-20 afin de montrer comment deux chaînes de caractères de longueurs différentes ont été associées à l'aide des critères définis précédemment. Le premier modèle « SANTA » est comparé au second modèle « Santa Fe Sport ». Le « Sport » pourrait indiquer une gamme d'options ou une motorisation différente du modèle de base, alors que le « Fe » est absent du premier modèle, sans doute parce que le premier modèle a été tronqué au premier espace. Le premier modèle comporte six trigrammes uniques, alors que le deuxième est décomposé en 14 trigrammes uniques. Comme les six trigrammes du premier modèle se retrouvent dans les 14 trigrammes du deuxième modèle, la similarité sera ici de 0,429, soit 6/14. » (Chaire Mobilité, 2021).

⁵ PostgreSQL 9.6.22 Documentation : <https://www.postgresql.org/docs/9.6/pgtrgm.html>

Tableau 2-20: Exemple de similarité entre deux séries de trigrammes (Morency, 2021)

Modèle	Trigramme
« SANTA »	{" s", " sa", ant, nta, san, "ta "}
« Santa Fe Sport»	{" f", " s", " fe", " sa", " sp", ant, "fe ", nta, ort, por, "rt ", san, spo, "ta "}

La Figure 2-39 montre la méthodologie de fusion de données qui a été développée. Premièrement, le Vehicle Identification Number (VIN) est utilisé afin d'associer les véhicules immatriculés avec une liste de VIN décodés. Le VIN est un code alphanumérique de 17 caractères, les 12^e à 17^e caractères étant uniques à chaque véhicule. La liste de VIN décodés sert seulement à avoir un meilleur nom de modèle, les noms de modèles dans les données de la SAAQ étant tronqués à 5 caractères. « Ensuite, la jointure entre les de la SAAQ et les données CVS est faite avec les critères suivants :

1. Même année-modèle;
2. Même marque;
3. Similarité d'au moins 30 % pour l'ensemble des trigrammes. Dans le cas où plusieurs candidats pour l'association rencontraient ce critère, celui ayant le plus haut pourcentage de similarité est retenu. » (Chaire Mobilité, 2021)

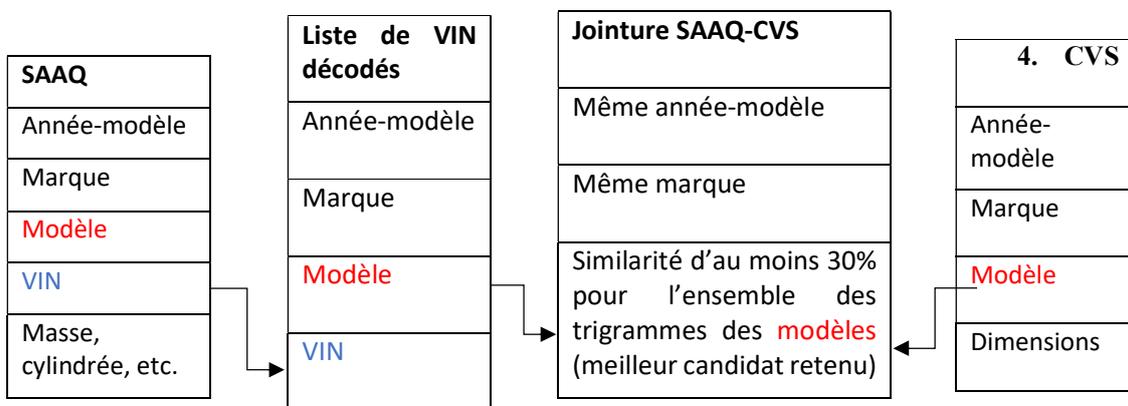


Figure 2-39 Mécanisme de fusion de données

« La Figure 2-40 montre l'évolution des taux d'association, soit le pourcentage des véhicules de la base de données de la SAAQ qui a été associé de manière fructueuse avec les données de CVS pour une année donnée. Il est important de noter que les données de CVS couvrent la période 1994-2019, le taux d'association entre 1990 et 1993 est donc nul. De plus, le faible taux d'association pour l'année 1994 peut être expliqué par le fait que plusieurs véhicules n'étaient pas dans les données de CVS à ce moment là puisqu'ils appartenaient à des années-modèles antérieures à 1994. À partir de 2001, le seuil de 50 % d'association est atteint, ce qui veut dire que la majorité des véhicules sont associés avec CVS à partir de cette année. » (Chaire Mobilité, 2021)

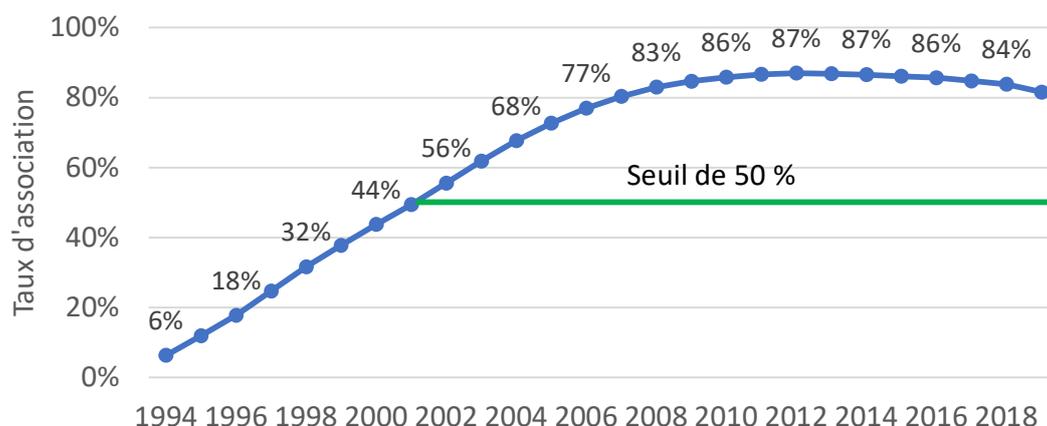


Figure 2-40 Évolution du taux d'association entre les données de la SAAQ et les données CVS entre 1994 et 2019

Également, la Figure 2-42 montre l'évolution du degré de similarité des candidats de la méthode des trigrammes entre 1994 et 2019. La ligne montre le pourcentage de modèles avec au moins un candidat à chaque année. Il est possible de voir que ce pourcentage est plus faible avant 2010, avec un minimum à 68.5 % en 2005. À partir de 2010, le taux de modèles avec au moins un candidat est plus élevé avec une valeur maximale de 89.2 % atteinte en 2018. En ce qui concerne le taux de similarité moyen, celui-ci varie entre 78.1 % en 2018 et 86.5 % en 2009 pour l'ensemble des candidats, c'est-à-dire ceux qui ont un degré de similarité supérieur ou égal à 30 %. Le taux de similarité semble avoir atteint un pic autour de l'année-modèle 2009, taux qui est redescendu depuis à des valeurs généralement entre 80 % et 85 %. En ce qui concerne le taux de similarité pour le meilleur candidat, c'est-à-dire le candidat ayant le plus haut taux de similarité pour chaque modèle, celui-ci varie entre 90.6 % en 1994 et 97.7 % en 2015. Ainsi, le choix d'associer les modèles de véhicules avec la méthode des trigrammes semble être une avenue prometteuse pour les prochaines années puisque le pourcentage de modèles associés et le taux moyen de similarité du meilleur candidat sont plus élevés dans les années les plus récentes.

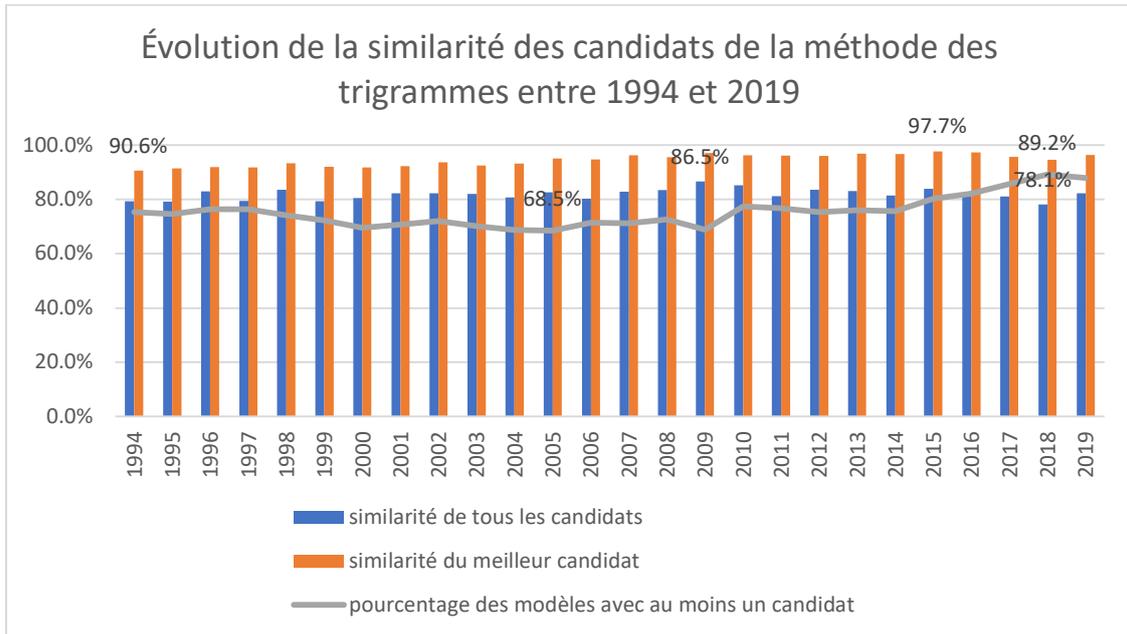


Figure 2-41 Évolution de la similarité des candidats de la méthode des trigrammes entre 1994 et 2019

Également, la Figure 2-41 montre la distribution de la similarité des candidats de la méthode des trigrammes entre 1994 et 2019. Tel que mentionné précédemment, aucun candidat ne possède un taux de similarité inférieur à 0.3 ou 30 % puisqu'il s'agit d'un seuil qui a été fixé lors de la sélection des candidats. Premièrement, il est possible de remarquer que très peu de candidats se retrouvent avec un taux de similarité entre 0.3 et 0.35. En ce qui concerne la similarité de l'ensemble des candidats, il est possible de remarquer que plus de 13 % des candidats ont un taux de similarité entre 0.35 et 0.40. Cependant, lorsque les meilleurs candidats sont regardés, seulement 1.8 % des candidats ont un taux de similarité entre 0.35 et 0.40. Ainsi, à peine plus de 10 % des candidats qui sont dans cette fourchette de taux de similarité seront retenus comme étant le meilleur candidat pour un modèle. Finalement, lorsque les meilleurs candidats (0.95 et plus) sont regardés, ceux-ci représentent plus de 67 % (2/3) de l'ensemble des candidats mais plus de 88 % (7/8) des meilleurs candidats. Cela veut donc dire que si le seuil pour le taux de similarité avait été monté à 0.95, 88 % des meilleurs candidats auraient été trouvés. Ainsi, le seuil de 0.3 initialement déterminé est très bas et ne génère pas énormément de candidats supplémentaires.

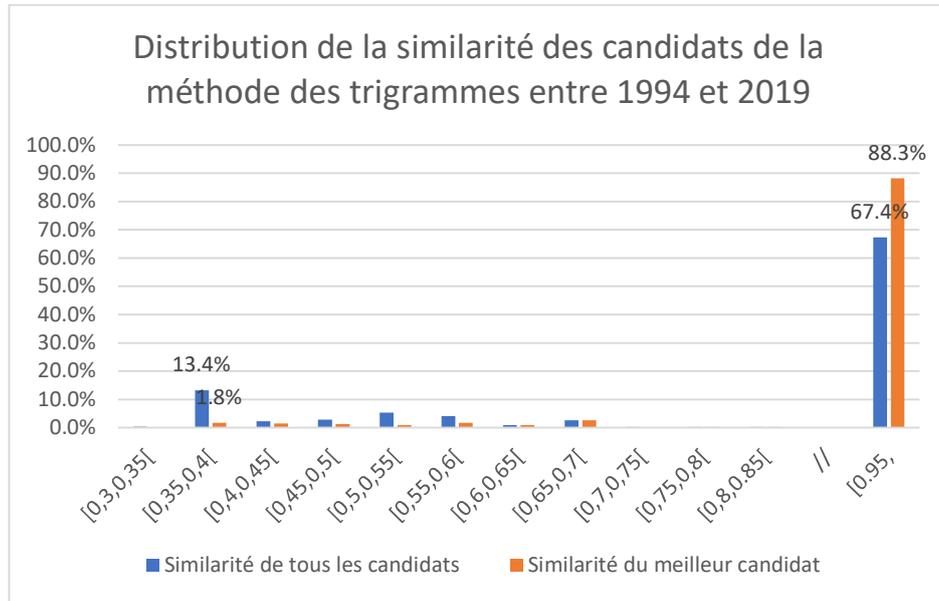


Figure 2-42 Distribution de la similarité des candidats de la méthode des trigrammes entre 1994 et 2019

2.12.4 Constats

Le premier constat qu'il est possible de faire est que les dimensions des véhicules sont en constante augmentation depuis les deux dernières décennies. Tel que montré au Tableau 2-21, la longueur moyenne des véhicules immatriculés a augmenté de 0.24% entre 2000 et 2019, tandis que l'empreinte au sol moyenne a augmenté de 3.26 % et l'empreinte volumétrique moyenne a augmenté de 10.5 %. La différence est encore plus importante lorsque les véhicules les plus récents du parc automobile de 2019 sont examinés, les augmentations sont de 1.08% pour la longueur moyenne, 5.40 % pour l'empreinte au sol moyenne et 14.25 % pour l'empreinte volumétrique moyenne.

Tableau 2-21 Évolution de la longueur moyenne, de l'empreinte au sol et de l'empreinte volumétrique des véhicules immatriculés au Québec entre 2000 et 2019

Année	Longueur moyenne - OL (cm)	Empreinte au sol - OL*OW moyenne (m ²)	Empreinte volumétrique moyenne - OL*OW*OH (m ³)
2000 – tous	460.9	8.15	12.00
2019 – tous	462.0	8.41	13.26
Variation 2000-2019	+ 0.24 %	+ 3.26 %	+ 10.5 %
2019 – véhicules de moins de 2 ans	464.5 (+ 1.08 % vs 2000 - tous)	8.59 (+ 5.40 % vs 2000 - tous)	13.71 (+ 14.25 % vs 2000 - tous)

Le Tableau 2-22 montre l'impact de la distance inter-véhiculaire sur la capacité estimée de stationnement sur rue dans un arrondissement central de l'île de Montréal, l'arrondissement

Outremont. Étant donné que la longueur moyenne d'un véhicule de promenade au Québec se situe à un peu plus de 4.6 mètres en 2019 tel que montré au Tableau 2-20, la distance inter-véhiculaire, qui tient compte de l'espace entre les véhicules stationnés sera plus grande, celle-ci a été estimée à 5.0 mètres, 5.5 mètres, 6.0 mètres et 6.5 mètres pour les fins de comparaison. La capacité de stationnement a été estimée à quatre moments différents dans la semaine, soit le mardi à 6h et 10h, le jeudi à 19h et le dimanche à 13h. En nombres absolus, la diminution de la capacité de stationnement varie entre 1 834 véhicules le mardi à 10h et 2 130 véhicules le mardi à 6h pour l'ensemble de l'arrondissement d'Outremont lorsque la distance inter véhiculaire passe de 5.0 à 6.5 mètres. En moyenne, lorsque la distance inter véhiculaire augmente de 5.0 mètres à 5.5 mètres, la capacité de stationnement diminue de 9.3 %, lorsque la distance inter véhiculaire passe de 5.5 à 6.0 mètres, la capacité de stationnement diminue de 7.1 % et lorsque la distance inter-véhiculaire passe de 6.0 à 6.5 mètres, la capacité de stationnement diminue de 10.6 %.

Tableau 2-22 Capacité estimée de stationnement sur rue dans l'arrondissement Outremont en fonction de la distance inter-véhiculaire

Capacité estimée de stationnement sur rue dans l'arrondissement Outremont	Distance inter-véhiculaire (m)			
	<u>5.0</u>	<u>5.5</u>	<u>6.0</u>	<u>6.5</u>
Mardi 6h	8 638	7 833	7 268	6 508
Mardi 10h	7 435	6 746	6 254	5 601
Jeudi 19h	7 844	7 107	6 607	5 907
Dimanche 13h	8 475	7 683	7 150	6 385

Le Tableau 2-23 montre l'impact de l'évolution du parc de véhicules sur la consommation moyenne combinée entre 2005 et 2019. Lorsque l'ensemble des véhicules sont comparés, la consommation moyenne diminue de 6.1 % ou 0.70 L / 100 km entre 2005 et 2019. Cependant, en supposant que le parc de véhicules de 2005 soit demeuré constant mais qu'il ait bénéficié de l'amélioration énergétique des véhicules de 2019, la diminution serait alors de 8.6 % ou -0.98 L / 100 km, ce qui signifie que selon ces estimations environ 0.28 L / 100 km de consommation d'essence seraient attribuables à la transformation du parc de véhicules, c'est-à-dire que les véhicules sont plus puissants, lourds et volumineux. Finalement, la consommation moyenne des véhicules les plus éco énergétiques est passée de 5.1 L / 100 km en 2005 à 4.3 L / 100 km en 2019, ce qui représente respectivement 44.4 % et 39.9 % de la consommation moyenne, ce qui signifie que les véhicules les plus éco énergétiques ont diminué leur consommation moyenne beaucoup plus que l'ensemble du parc automobile entre 2005 et 2019.

Tableau 2-23 Évolution de la consommation moyenne combinée d'essence de tous les véhicules et des 10 véhicules les plus éco énergétiques entre 2005 et 2019

Année	Consommation moyenne combinée (L / 100 km)	Consommation moyenne combinée des 10 véhicules les plus éco énergétiques (L / 100 km)
2005 – tous les véhicules	11.47	5.1 (- 6.37 L / 100 km)
2019 – tous les véhicules	10.77	4.3 (- 6.47 L / 100 km)
Variation 2005-2019	-0.70 L/100km (- 6.1%)	-
Véhicules de 2005 avec la consommation de 2019	10.48 (- 0.98 L/100km ou - 8.6 %)	-

2.12.5 Conclusion

En conclusion, la présente analyse a permis de fusionner les données sur les véhicules de promenade de la SAAQ avec les données sur les dimensions des véhicules provenant de la base de données CVS et les données sur les cotes de consommation énergétique. Une méthodologie de fusion de données a été développée qui permet d'associer les véhicules à partir des noms des modèles au moyen de l'utilisation des trigrammes. Lorsque plusieurs candidats répondent au critère minimal de similarité de 30 %, le meilleur candidat est sélectionné. Le pourcentage de similarité moyen et le pourcentage des modèles avec au moins un candidat augmentent avec les années.

Les données fusionnées ont permis de constater que les véhicules sont de plus en plus longs, larges et hauts, ce qui entraîne une augmentation de l'empreinte au sol et de l'empreinte volumétrique des véhicules. Cet accroissement de la longueur des véhicules a notamment comme impact de diminuer la capacité de stationnement sur rue. Également, les données fusionnées ont permis de constater que les véhicules consomment moins d'essence qu'auparavant mais que les gains ont été en partie annulés par la transformation du parc automobile. De plus, la consommation moyenne combinée est plus élevée dans le centre de la GRM que dans les municipalités périphériques.

Plusieurs limites doivent être soulignées. Premièrement, les bases de données ne sont pas exhaustives, ce qui rend la fusion de données imparfaite. Ensuite, les descriptifs utilisés dans les noms des modèles ne sont pas toujours les mêmes à travers les sources de données, ce qui rend la fusion plus difficile. Il a été observé que ces descriptifs sont plus souvent présents lorsque les véhicules ont des vocations particulières, qu'elles soient sportives, luxueuses ou utilitaires.

La transformation du parc automobile engendre de nombreux impacts. Certains de ces impacts, comme la consommation d'espace ou la consommation énergétique, ont été constatés dans la présente analyse. Cependant, plusieurs autres impacts devraient être étudiés, par exemple la capacité routière, la sécurité routière et l'impact sur les finances des ménages.

2.13 Qualité des informations géographiques et impacts sur les calculs

- Par : Pierre-Léo Bourbonnais, ing., associé de recherche
- Contexte : Chaire Mobilité

2.13.1 Introduction

Dans tout projet d'analyse, d'optimisation ou de simulation des systèmes de transport, les données permettant de modéliser les réseaux piéton, cyclable et routier représentent la base à laquelle s'arrime les données sur les réseaux de transport collectif, les générateurs de déplacements, les lieux d'activité et les lieux de résidence. De surcroît, du fait de la capacité des logiciels de modélisation à simuler des réseaux de plus en plus complexes et élargis, ils requièrent des données encore plus précises pour être en mesure de générer des résultats réalistes. Or, les géobases routières sont souvent incomplètes ou imprécises, ce qui limite les possibilités de réaliser des analyses rigoureuses et exhaustives. De plus, il est souvent difficile d'y avoir accès et de connaître leurs limitations et leur niveau de précision. C'est dans ce contexte que les plus importantes sources de données seront comparées et que l'impact du degré de complétion et de précision des données sera analysé et discuté.

En premier lieu, le contexte et les problématiques associées aux données actuelles seront présentés, suivis d'une description des données OpenStreetMap et de ses alternatives commerciales. L'impact de la qualité des données sur les temps de parcours piétons sera ensuite expliqué. Finalement, les ressources nécessaires à la complétion et la validation des données OpenStreetMap pour le Québec seront estimées.

2.13.2 Problématiques et contexte

Depuis plusieurs années, un grand nombre d'ensembles de données géographiques et statistiques sont générés et proposés aux différents intervenants dans le milieu du transport, de l'urbanisme et de la géomatique. Auparavant, ces données étaient rendues disponibles à certains groupes et à certains individus seulement (universités, décideurs, analystes, fonctionnaires, etc.), mais avec l'avènement des mouvements qui militent pour la diffusion de données ouvertes, les gouvernements et fournisseurs de données ont commencé à rendre accessible une partie de leurs ensembles de données.

La grande majorité de ces données sont colligées et diffusées par des institutions gouvernementales (municipales, provinciales et fédérales) ou des organismes publics. Toutefois, il existe des problématiques importantes associées en particulier au niveau de complétion des données et à leur précision, notamment sur le plan des données routières qui incluent les réseaux piéton et cyclable. En particulier, la connectivité des différents segments est quelque fois incomplète ou erronée et les attributs essentiels à la modélisation comme les limites de vitesse, la position des feux de circulation et des panneaux d'arrêt, l'emplacement des traverses piétonnes et cyclables, ainsi que le nombre de voies et les zones de transition des voies sont souvent manquantes ou incomplètes.

Le projet de développement de la plateforme Transition (Financement CRSNG, INNOVÉE, STM, STL, RTL, EXO, STO, STSherbrooke et MTQ) a amené plusieurs questionnements sur l'influence de la qualité des données sur l'analyse et les propositions d'optimisation des réseaux de transport collectif. La comparaison des temps de parcours entre les modes est au centre de cette optimisation et des simulations qui y sont associées. En ce sens, dans le cadre de projets d'optimisation au moyen d'un algorithme génétique (RTL, Drummondville, EXO), les constats suivants sont ressortis:

- Le réseau routier est quelquefois fragmentaire, incomplet, obsolète et souvent non validé
- Le réseau piéton (trottoirs, traverses et sentiers) est presque toujours inexistant ou très fragmentaire
- Le réseau cyclable est souvent incomplet ou mal connecté au reste du réseau
- Les trajets pour les bus sont difficiles à calculer : il est difficile de bien estimer les pénalités réelles aux feux de circulation, aux panneaux d'arrêt et lors des virages à droite et à gauche

2.13.3 OpenStreetMap

OpenStreetMap¹ est une plateforme collaborative de cartographie basée sur le même principe que Wikipédia. Les usagers inscrits (abonnement gratuit) peuvent ajouter ou modifier le contenu de la carte n'importe où sur la planète. Chaque changement (appelé *changeset*) est enregistré et daté pour que l'on puisse facilement retracer les changements effectués par chacun des usagers. Les données historiques sont toutes conservées et les données peuvent être exportées en tout temps au moyen d'un API (interface permettant de communiquer avec les données). La manière de coder les différents éléments cartographiques fait l'objet de discussions afin d'obtenir un consensus qui est présenté dans un WIKI disponible pour la communauté. Chaque communauté OpenStreetMap (ville, région ou un pays) peut proposer des paradigmes différents pour certaines données cartographiques pour mieux répondre aux particularités locales.

2.13.3.1 *Contributeurs actuels d'OpenStreetMap au Québec*

Au Québec, il existe déjà une communauté active d'usagers qui complètent, mettent à jour et discutent de l'évolution des données cartographiques de la province et de ses villes principales. Seulement pour la région de Montréal, il existe une dizaine d'usagers qui modifient et améliorent les données au moins une fois par semaine. Au Canada, 200 usagers actifs ont effectué au moins 1000 changements à la carte dans les deux derniers mois (juillet et août 2021). La Chaire Mobilité est déjà la plus grande contributrice de données OpenStreetMap au Canada en 2021 et 2022, avec plus de 4 millions de modifications apportées et plus de 17 000 *changesets* complétés en date d'avril 2022. La grande majorité des usagers actifs d'OpenStreetMap complètent les données bénévolement, bien que des compagnies comme Amazon, Lyft, Uber, Microsoft et Mapbox engagent des codificateurs pour corriger les problèmes et les erreurs rencontrées notamment dans les calculateurs de chemins utilisés pour les livraisons de colis et les services de taxi. La Chaire Mobilité complète les données de certaines régions en priorité selon les projets qu'elle mène auprès de ses partenaires de recherche.

2.13.3.2 *Utilisation d'OpenStreetMap en mode privé ou local*

Pour des besoins de modélisation et de planification, il est souvent nécessaire de projeter de nouvelles rues, de nouvelles configurations, de nouvelles pistes cyclables ou de nouveaux sentiers piétons qui n'existent pas en réalité. OpenStreetMap comprend un modèle de données public et ouvert pouvant être utilisé de manière locale sans publier les données pour la communauté. Cela permet par exemple d'importer les données existantes sur un serveur local et d'ensuite ajouter des éléments cartographiques projetés ou planifiés pour des besoins de simulation ou de modélisation.

2.13.3.3 *Problématiques associées à OpenStreetMap*

La problématique principale d'OpenStreetMap est associée à la variabilité de la qualité et du niveau de complétion des données d'une ville ou d'une région à l'autre. De manière générale, la quantité et la qualité des données cartographiques ajoutées et validées dans une région donnée

est influencée en grande partie par l'assiduité des usagers locaux. La deuxième problématique vient du fait qu'il existe, même si cela demeure rare, des cas de vandalisme². Certaines compagnies utilisent des stratagèmes pour faciliter la publicité de leur entreprise au moyen des données cartographiques, ce qui est plutôt facile à contrecarrer en filtrant certaines données lors de l'importation. Le deuxième cas vient des modifications de certains usagers qui ajoutent des données fictives ou suppriment des données réelles. Le troisième cas est relatif aux guerres d'édition³, qui surviennent lorsque deux ou plusieurs usagers ne s'entendent pas sur la manière de codifier un ou des éléments géographiques (souvent lorsqu'il n'y a pas de consensus). Ces usagers modifient les uns après les autres les éditions précédentes, jusqu'à ce qu'un administrateur impose un blocage temporaire ou permanent sur un ou plusieurs usagers. Heureusement, il existe de nombreux mécanismes de résolution des conflits et des problématiques de vandalisme. Une dernière problématique concerne les disputes territoriales qui apparaissent dans les données d'OpenStreetMap (notamment sur les frontières de certains pays comme la Chine, la Russie, l'Inde ou le Pakistan), qui ont des répercussions géopolitiques importantes, mais cette problématique est pratiquement inexistante au Canada à l'heure actuelle.

2.13.3.4 Utilisations officielles d'OpenStreetMap

OpenStreetMap est utilisé de manière officielle par certains gouvernements ou institutions, mais également par de nombreuses compagnies. Une liste d'utilisations officielles d'OpenStreetMap est disponible à cette adresse :

<https://welcome.openstreetmap.org/about-osm-community/consumers/>

Parmi ces utilisateurs, on compte des multinationales comme Amazon, Microsoft, Facebook, Apple, des compagnies de transport comme Air France, SNCF et Uber, ainsi que des compagnies offrant des services de données cartographiques comme ESRI, Garmin, Mapbox, Digital Globe et Telenav. De manière générale, la grande majorité des entreprises concurrentes de Google ou n'utilisant pas ses services utilisent les données d'OpenStreetMap. D'autre part, de plus en plus d'institutions gouvernementales utilisent OpenStreetMap comme source officielle de données et/ou partagent leurs données avec OpenStreetMap (Statistiques Canada, France, Lituanie, Italie, Brésil, New York City, *US National Park Service*, etc.) et de nombreuses agences de transport collectif les utilisent également.

2.13.4 Alternatives à OpenStreetMap

Le seul système de cartographie en ligne qui soit comparable à OpenStreetmap en termes de couverture et de qualité des données est Google Maps. Or, la licence d'utilisation de Google Maps étant très restrictive, un grand nombre d'opérations ne peuvent y être effectuées. Par exemple, aucun usager ne peut s'inscrire pour modifier ou compléter les données, ce qui limite grandement son utilisation potentielle dans une optique de développement des réseaux et de validation des données cartographiques. D'autre part, le modèle de données de Google Maps n'est pas public et ne peut donc être utilisé sur un serveur local. On ne peut donc pas l'utiliser pour des simulations ou de la modélisation. Néanmoins, l'avantage principal de Google Maps est le très grand nombre de commerces et de lieux d'activités qui, conscient de la portée de Google Maps auprès du grand public, ajoutent et mettent à jour leur adresse et leurs informations sur une base régulière, et ce, en négligeant souvent OpenStreetMap. Ce phénomène est moins présent en Europe, où les données d'OpenStreetMap sont de manière générale entretenues avec davantage de régularité, souvent en lien avec des lois et réglementations plus restrictives sur les activités commerciales de Google et des conglomérats commerciaux qui offrent des services de cartographie en ligne. De

son côté, Apple développe présentement son propre système cartographique et désire atteindre le niveau de complétion et de précision de Google dans les prochaines années. Microsoft possède également son système de cartographie appelé Bing Maps et collabore notamment avec OpenStreetMap pour l'accès à des photos aériennes pour le bénéfice des éditeurs.

2.13.5 Impacts de la qualité des données sur les temps de parcours piéton et l'accès au transport collectif

Dans le but de vérifier l'influence de la qualité des données sur les simulations, une comparaison des temps totaux de marche pour des déplacements de transport collectif avant et après la validation et la complétion des données a été réalisée sur le territoire de la ville de Drummondville, qui a été complété à 100% au cours de l'année 2020 par l'équipe de recherche. La Figure 2-43 montre les différences dans les temps de marche totaux (accès à l'origine, accès à la destination et marche lors des transferts) pour un échantillon aléatoire de 44 000 déplacements de transport collectif. Les points montrent l'origine des déplacements et seuls les déplacements dont le temps de marche est différent de plus de 5 minutes sont montrés. Les points verts représentent des déplacements dont la marche est au moins 5 minutes moins longue qu'avec le réseau non validé et non complété, alors que les points rouges représentent les déplacements dont la marche est au moins 5 minutes plus longue avec le réseau validé et complété. On remarque que la dispersion des différences de temps de marche n'est pas uniforme et que certains secteurs ressortent avec des temps moyens plus longs (cercles bleus) ou des temps moyens plus courts (cercles roses). La répartition du biais causé par la non-validité et la non-complétion des données de réseau préalables aux analyses et aux optimisations n'est donc pas aléatoire géographiquement, ce qui cause des problématiques importantes pour l'accessibilité aux arrêts. De plus, des temps de marche jusqu'à 15 minutes plus longs peuvent avoir une influence significative sur le choix modal d'une part, mais aussi sur le choix du trajet de transport collectif. De plus, un peu plus de 1% des déplacements qui pouvaient se faire avec un accès de 20 minutes et moins à l'origine et à destination ne peuvent plus se faire et vice-versa.

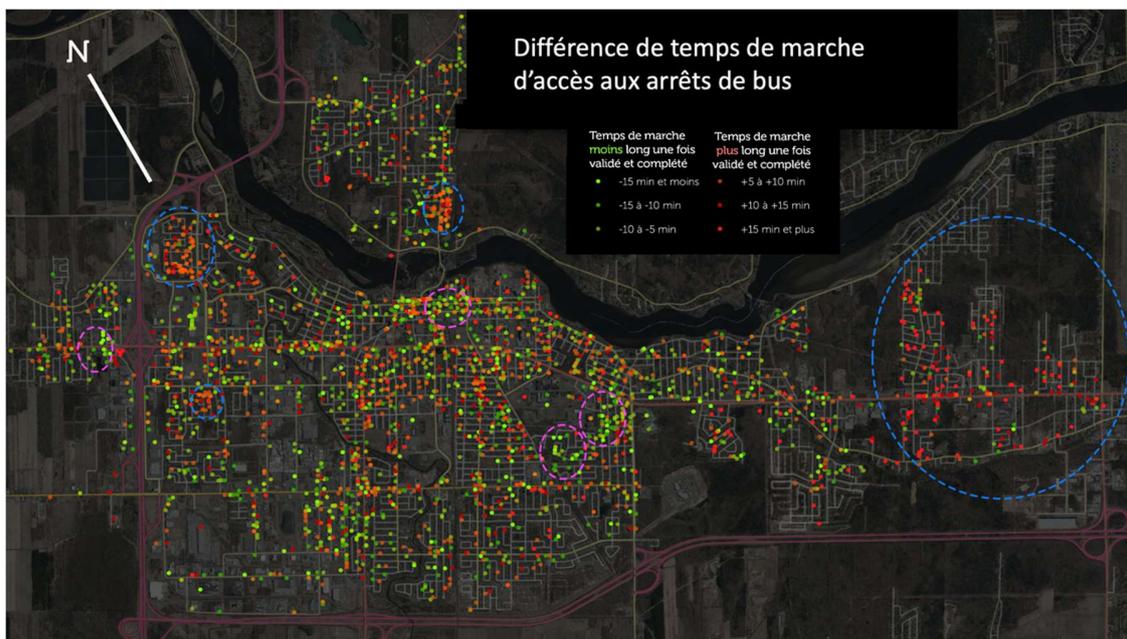


Figure 2-43 Comparaison des temps de marche pour un échantillon aléatoire de déplacements de transport collectif

Par ailleurs, pour déterminer l'influence du niveau de complétion et de précision des données sur les calculs de chemin, une analyse du nombre d'arrêts de bus accessibles à pied en 10 minutes a été réalisée avec le territoire complet de la Ville de Sherbrooke. Quatre niveaux de précision des données ont été comparés pour les déplacements à pied :

- Calcul à vol d'oiseau (5 km/h)
- Réseau routier seulement (5 km/h)
- Réseau routier + passages et sentiers piéton (5 km/h)
- Réseau routier + passages et sentiers piéton (5 km/h + de 20 secondes aux feux de circulation)

La Figure 2-44 montre la différence en pourcentage des arrêts qui sont accessibles en 10 minutes ou moins par rapport au niveau de référence (distance euclidienne avec vitesse de marche de 5 km/h). On remarque que les trois niveaux de précision sont comparables par rapport à la référence (perte d'environ 45 % des arrêts accessibles). Cependant, lorsqu'on observe les mêmes résultats sur la carte du territoire (Figure 2-45), on remarque que l'ajout des sentiers et traverses piétons et des pénalités aux feux amènent un biais considérable selon la position géographique. En effet, l'accessibilité dans les zones plus denses avec de nombreux feux de circulation provoquent une baisse marquée du nombre d'arrêt accessibles, alors que dans les zones moins denses, l'ajout des sentiers et des traverses permet à un plus grand nombre d'arrêts de bus d'être accessibles en 10 minutes ou moins.

Pourcentage des arrêts de bus dans un rayon de 10 minutes à vol d'oiseau encore accessibles en 10 minutes ou moins avec temps de marche réseau (5 km/h)

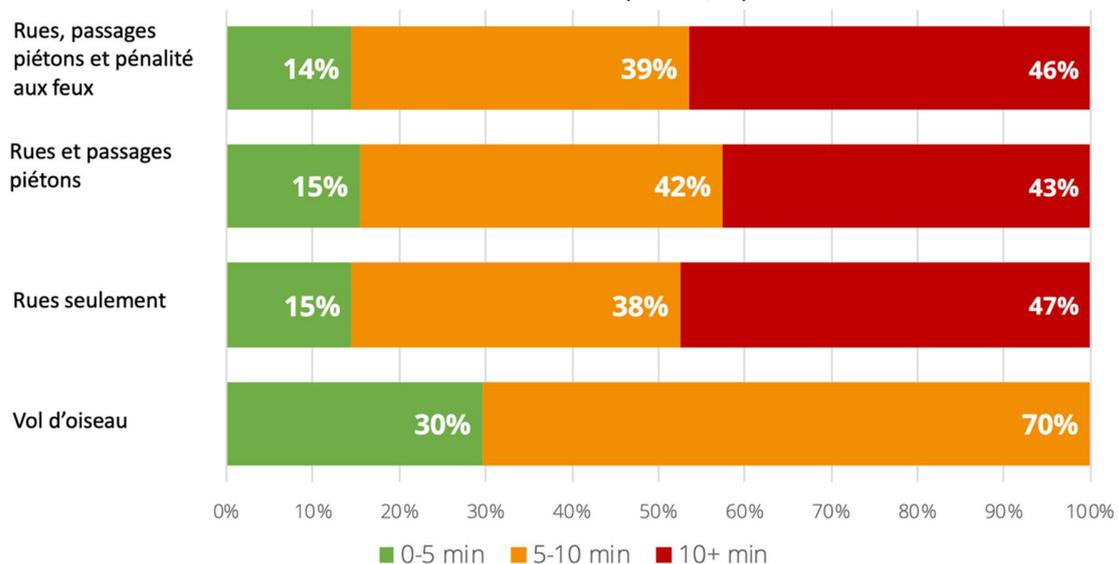


Figure 2-44 Pourcentage des arrêts de bus dans un rayon de 10 minutes à pied encore accessibles selon le niveau de précision des données

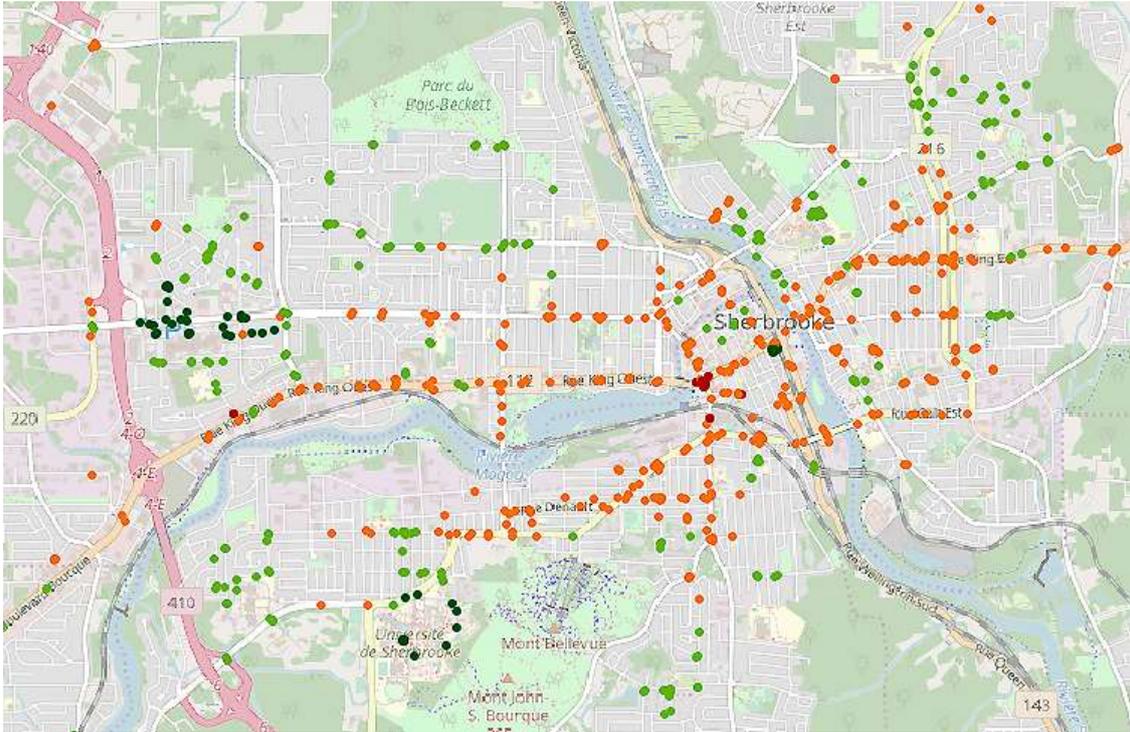


Figure 2-45 Comparaisons réseau routier seulement vs ajout du réseau piéton et pénalités aux feux. Orange/rouge: moins d'arrêts de bus accessibles à pied en 10 minutes qu'avec la marche sur le réseau routier seulement. Vert: plus d'arrêts de bus accessibles à pied qu'avec la marche sur le réseau routier seulement

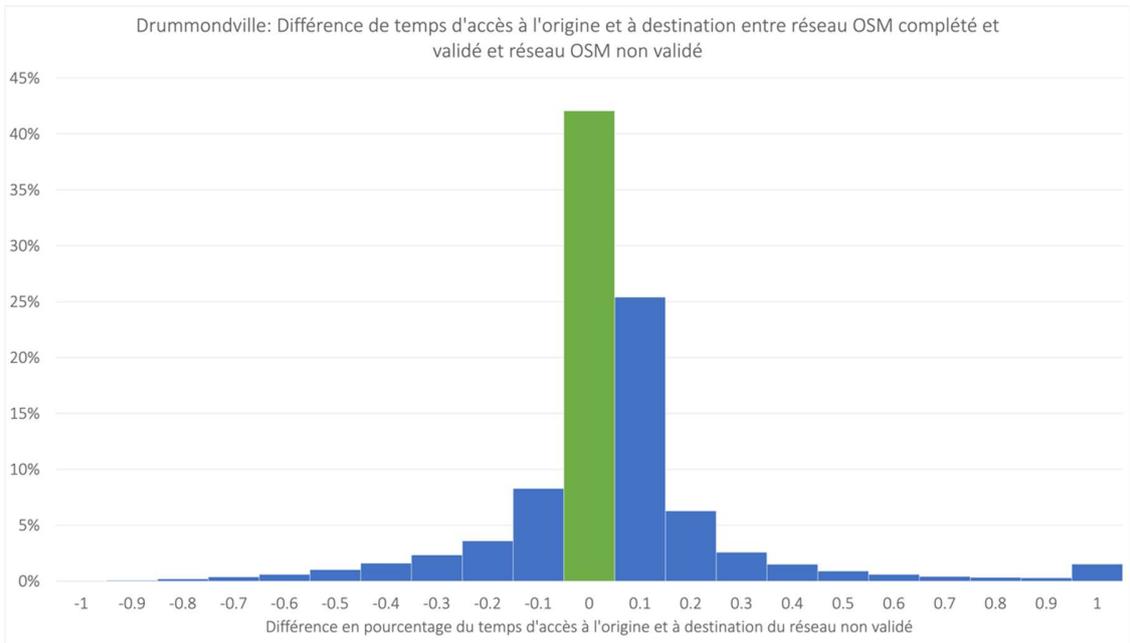


Figure 2-46 Différence de temps d'accès entre le réseau OpenStreetMap complété et validé et le réseau original non validé (Drummondville)

Enfin, lorsqu'on compare le niveau de complétion et de validation des données sur le réseau piéton pour Drummondville, on observe à la Figure 2-46 qu'environ 10% des trajets sont au moins

20% plus courts avec des données validées et complétées, alors que près de 15% des trajets sont au moins 20% plus longs. En conséquence, près de 25% des parcours d'accès au réseau de bus ont une différence d'au moins 20% dans leur temps de parcours. Aussi, un peu moins de 2% ont un accès au moins 100% plus long lorsqu'on utilise les données complètes et validées.

2.13.6 Ressources estimées requises pour valider et compléter les réseaux OpenStreetMap

Dans le but d'évaluer les ressources nécessaires pour compléter, d'une part, et pour mettre à jour en continu, d'autre part, les données d'OpenStreetMap pour le Québec, des statistiques ont été colligées sur les éditions effectuées par l'équipe de la Chaire depuis 2 ans.

2.13.6.1 Progression actuelle

Certaines régions ont déjà été complétées et mises à jour dans le cadre de projets réalisés par la Chaire Mobilité. Dans le cadre d'un projet de mise à niveau du réseau de transport collectif de la ville de Drummondville, l'ensemble des données cartographiques du territoire de la municipalité ont été complétées, validées et mises à jour entre 2020 et 2021. Aussi, une partie de la rive sud de Montréal a été complétée (notamment les villes de Sainte-Julie, La Prairie, Candiac, Delson, Saint-Constant et Saint-Catherine) et le secteur de la Presqu'île (Île Perrot et Vaudreuil-Dorion). L'ensemble du réseau autoroutier de la région de Montréal a également été validé et complété (réalignement, nombre de voies, vitesses, bretelles et voies de services).

2.13.6.2 Prévisions et statistiques

Le travail de codification, de validation et de complétion des données d'OpenStreetMap nécessaire à la réalisation de simulations précises et plausibles comprend plusieurs tâches distinctes qui demandent un certain temps pour être achevées. Le Tableau 2-24 Statistiques des ressources requises pour différentes tâches de codification dans OpenStreetMap (zone urbaine de densité moyenne) illustre les temps moyens requis pour effectuer les principales tâches de codification dans OpenStreetMap par km² en milieu urbain.

Tableau 2-24 Statistiques des ressources requises pour différentes tâches de codification dans OpenStreetMap (zone urbaine de densité moyenne)

Description	Variabiles affectées	Éléments comptés	Temps requis par élément (sec.)	Compte moyen par km ²	Temps total par km ² (min.)	Pourcentage du temps total requis
Routes de service, allées de stationnement et routes d'accès aux commerces, industries et bâtiments résidentiels de 4 logements et plus	Temps de parcours et accessibilité marche, vélo, auto, transport collectif	routes	70	80 routes/km ²	93	21%
Polygones des bâtiments résidentiels de 4 logements et plus	Synthèse de population Lieux d'activités (origines et destinations)	polygones	13	42 bâtiment/km ²	91	20%
Trottoirs et pistes cyclables avec traverses marquées et non marquées	Temps de parcours et accessibilité marche, vélo Sécurité des parcours en modes actifs	segments	66	46 segments/km ²	50	11%
Chemins d'accès piéton et vélo et portes des commerces, industries et bâtiments résidentiels de 4 logements et plus	Temps de parcours et accessibilité marche, vélo	segments	39	54 segments/km ²	35	8%
Écoles, collèges, parcs et autres institutions, incluant chemins d'accès, stationnements, nombre d'étudiants, nom et types	Temps de parcours et accessibilité marche, vélo Lieux d'activités (origines et destinations)	lieux	700	3 lieux/km ²	35	8%
Lieux d'activité (commerces, industries incluant type, nom et adresse)	Lieux d'activités (origines et destinations)	lieux	63	30 lieux/km ²	31	7%
Alignement des rues	Temps de parcours et accessibilité marche, vélo, auto, transport collectif	courbes	5	93 courbes/km ²	25	6%
Polygones des stationnements commerciaux, industriels et bâtiments résidentiels de 4 logements et plus	Disponibilité du stationnement	polygones	18	72 stationnements/km ²	22	5%
Attributs des rues (limites de vitesse, nombre de voies et voies de virage)	Temps de parcours et accessibilité marche, vélo, auto, transport collectif	segments	14	93 segments/km ²	22	5%
Polygones des zones résidentielles délimitées par les côtés de rues avec nombre total de logements	Synthèse de population Validation du nombre de logements bâtiments vs zones	polygones	13	300 courbes/km ²	21	5%
Panneaux d'arrêts et feux de circulation	Temps de parcours et accessibilité marche, vélo, auto, transport collectif Sécurité des parcours en modes actifs	arrêts/feux	8	105 arrêts/feux/km ²	14	3%
Polygones des bâtiments commerciaux et industriels	Lieux d'activités (origines et destinations) Attractivité et achalandage des lieux d'activité	polygones	39	12 bâtiments/km ²	8	2%
Total					448 (7.5 h)	100%

Le nombre d'heures requises pour compléter et valider les données d'OpenStreetMap représente donc environ près de 12 000 heures pour la grande région de Montréal et environ le double pour l'ensemble du Québec urbanisé. Le nombre d'heures requises par km² varie entre 1 et 27 heures selon la densité du secteur, avec une moyenne d'environ 7.5 heures.

2.13.7 Conclusion et perspectives

Bien que la complétion et la validation des données cartographiques représentent un défi considérable, la mise en commun des ressources de tous les intervenants du milieu des transports et de l'occupation du sol permet d'envisager la mise en place d'un ensemble de données précis et complet dans les années à venir. Le même processus de collaboration est déjà en cours dans plusieurs pays, notamment en Europe. En effet, les nouvelles lois sur la disponibilité des données ouvertes et l'augmentation de la tarification des services cartographiques en lignes de Google Maps dans les deux dernières années (augmentation de plus de 200% dans certains cas) accompagnée d'une licence commerciale très restrictive, encourage de nombreux décideurs à basculer vers des données ouvertes et publiques au bénéfice autant des professionnels du milieu du transport et de la géographie, des entreprises locales, ainsi que du grand public. Pour l'instant, seul OpenStreetMap est en mesure de répondre à ce défi.

2.14 Perspective économique de la transformation du parc de véhicules

- Par : Brigitte Milord, M.Sc. Économie, Associée de recherche
- Contexte : Chaire Mobilité

2.14.1 Contexte

La part de camions légers dans le parc de véhicules de promenade en circulation est en forte croissance au Québec, comme partout ailleurs dans le monde (Thomas and Walton, 2008; DfT, 2020; Opland, 2007). Les trois quarts des véhicules neufs vendus depuis quelques années sont des camions légers (Statistiques Canada, 2022). La transformation des habitudes de consommation des ménages, en faveur de véhicules plus gros et plus énergivores, suscite des inquiétudes grandissantes sur le plan environnemental, social et économique.

2.14.2 Publications et résultats associés

Pour inverser la tendance, il semble nécessaire d'identifier les causes et les conséquences de la transformation du parc de véhicules en circulation. Trois rapports ont ainsi émané des travaux de la Chaire Mobilité grâce à la contribution financière de Équiterre. Le premier rapport porte sur la définition des camions légers et l'évolution des propriétés des véhicules offerts par l'industrie automobile. Le second rapport porte sur la demande de camions légers ainsi que l'identification des causes possibles de la transformation du parc de véhicules (autant au niveau de l'offre que de la demande). Le troisième et dernier rapport porte sur les conséquences de la transformation du parc, en matière d'environnement, de sécurité routière, de retombées économiques, de répercussions financières pour les ménages et d'encombrement urbain (congestion, stationnement, occupation de l'espace). L'ensemble des publications peuvent être retrouvées sur le [site internet de Équiterre](#).

Dans une perspective de transfert et de partage de connaissances interdisciplinaire, les rouages de la théorie de la consommation (Nicholson et Snyder, 2008; Stiglitz, Walsh et Lafay, 2007) ont été présentés aux participants lors du colloque 2021. Tel qu'illustré ci-dessous, les effets de revenu peuvent contribuer à expliquer la croissance de la motorisation. Les effets de substitution peuvent quant à eux contribuer à expliquer l'appétit des consommateurs pour des véhicules plus gros et plus énergivores.

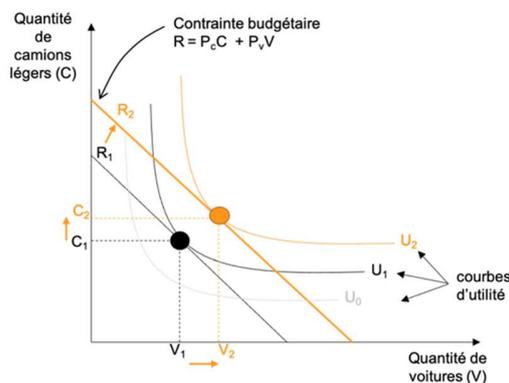


Figure 2-47 : Illustration d'un effet de revenu

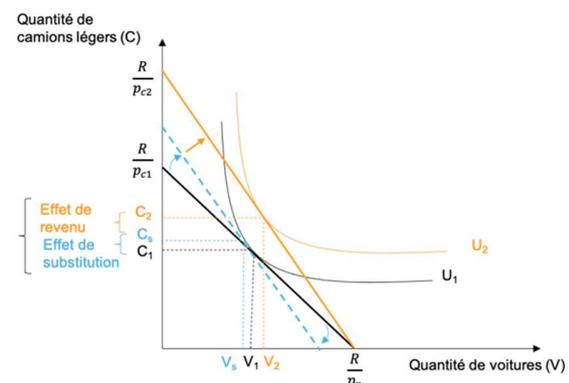


Figure 2-48 : Illustration d'un effet de substitution

Source : figures inspirées des travaux de vulgarisation économique de (Nicholson et Snyder, 2008; Stiglitz, Walsh et Lafay, 2007)

Les répercussions des effets de revenu et de substitution sur la motorisation et sur le choix de véhicules ont été abordées par quelques auteurs (Dargay 2001; Goodwin 1992; McCarthy et Tay 1998; Nolan 2010).

D'un point de vue économique, la plus importante contribution est d'avoir exploré les répercussions importantes que peut avoir l'allongement des prêts automobiles sur la demande de véhicules. L'Agence de la consommation en matière financière du Canada estime que les prêts à long terme encouragent les consommateurs à acheter des véhicules plus dispendieux et donc plus gros, plus énergivores ou plus luxueux (ACFC 2016). La figure ci-dessous poursuit cette réflexion en illustrant l'effet de l'allongement de la durée des paiements de 60 à 84 mois sur le choix de consommation d'un ménage qui dispose d'un budget mensuel pour son paiement auto d'au plus 450 \$ par mois. Si le terme de financement est fixé à 60 mois (en abscisse), alors le ménage est contraint à l'achat d'une voiture (voir le paiement mensuel correspondant sur l'axe des ordonnées). L'allongement de la durée du paiement améliore l'accessibilité financière des camions légers. Sous un financement de 84 mois, un ménage peut acquérir un véhicule 10 000 \$ plus cher, tout en respectant son budget mensuel, en supposant que le taux d'intérêt soit de 3 %. À des fins de simplification, l'effet des taxes de vente n'a pas été considéré.

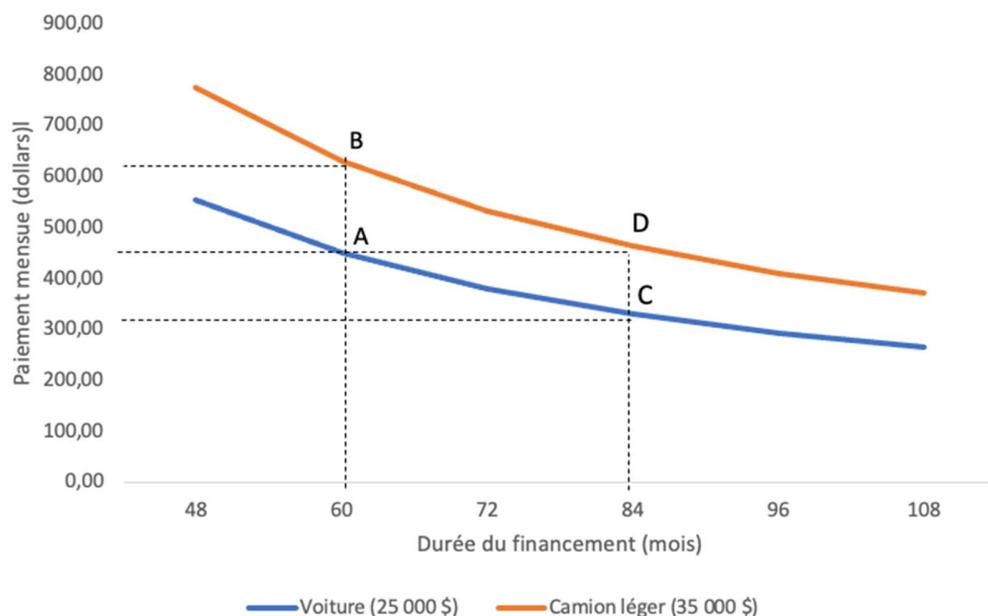


Figure 2-49 : Illustration théorique des options de financement sur le choix d'un véhicule, Source : figure inspirée des travaux de (ACFC 2016)

Une autre des contributions est d'avoir soulevé la présence d'un phénomène de transformation de la demande plus complexe connu sous le nom d'effet rebond. Aussi connu sous le nom de Paradoxe de Jevons, l'effet rebond découle d'une réduction des coûts énergétiques, qui produit un effet de revenu et qui se traduit par une augmentation de la demande énergétique. Ce paradoxe explique pourquoi les stratégies d'amélioration de l'efficacité énergétique sont souvent inefficaces pour favoriser l'atteinte d'objectifs environnementaux. En d'autres termes, il vaudrait mieux taxer la pollution que de subventionner l'efficacité énergétique (Stiglitz, Walsh et Lafay, 2007).

Plusieurs auteurs ont à cet égard noté que l'amélioration de l'efficacité énergétique des véhicules aurait été annulée par une augmentation du kilométrage parcouru (Greene et al., 1999; Small and Van Dender, 2007). Aux États-Unis, des analyses montrent que la croissance de la motorisation et la popularité croissante des VUS entraînent une dégradation des gains en matière d'efficacité énergétique (National Research Council, 2002). Les données canadiennes portent au même constat, comme le démontre la figure ci-dessous.

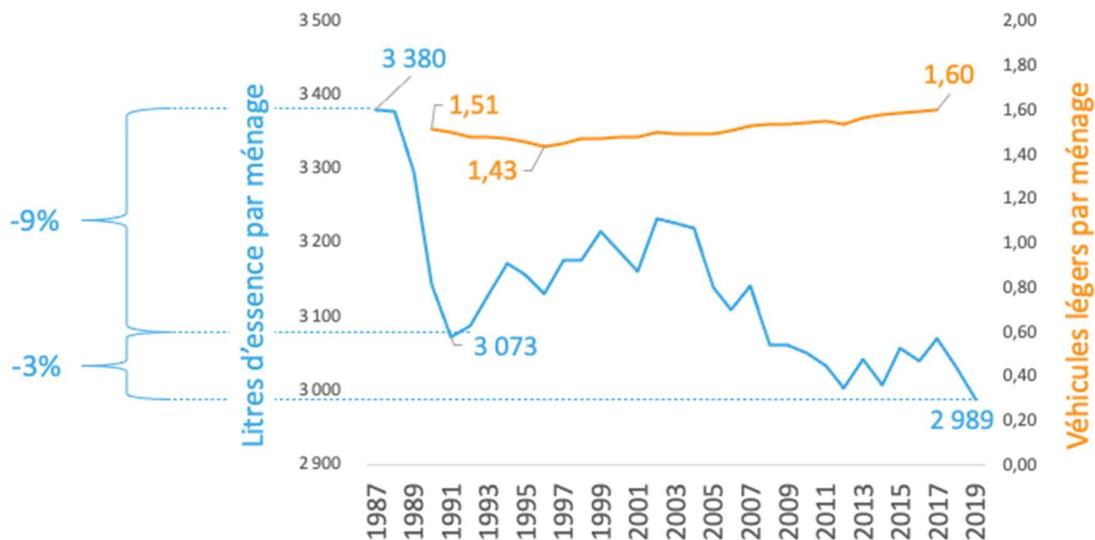


Figure 2-50 : Ventes nettes d'essence et véhicules immatriculés par ménage, Canada, 1987-2019

Sources: Statistique Canada, CANSIM, Tableaux 23-10-0067-01, 23-10-0066-01, Produit no 11-630-X au catalogue, Recensement 2016; Ressources naturelles Canada, Base de données nationale sur la consommation d'énergie

Des auteurs ont indiqué qu'une plus grande proportion de véhicules de faible efficacité énergétique, peut être retrouvée en banlieue et en régions plus éloignées (Lindsey et al., 2011). D'autres indiquent que la présence de véhicules énergivores tend à diminuer dans les environnements urbains offrant une mixité d'usage (Potoglou and Kanaroglou, 2007). Si les preuves scientifiques sont encore relativement peu nombreuses pour confirmer un impact de l'aménagement sur le choix de véhicules, il y a peut-être là un effet de revenu lui aussi peu étudié. L'étalement urbain pourrait permettre aux ménages de libérer de l'argent au niveau de l'habitation et de réallouer cet argent au profit d'une propriété plus grande, de l'achat de véhicules supplémentaires ou encore de véhicules plus gros.

Notre compréhension du phénomène à cet égard a continué d'évoluer en cours d'année grâce à l'obtention d'une subvention pour la réalisation d'une synthèse de connaissances sur l'abordabilité intégrée des localisations résidentielles co-financée par le CRSH et Infrastructure Canada (Morency, Tremblay-Racicot, Milord, 2022). Différentes théories économiques, notamment celle de la ville monocentrique, initialement développée par William Alonso (1964), puis enrichie par (Muth, 1969) et Mills (1972), semblent confirmer l'existence d'effets de revenu et de substitution.

2.14.3 Perspectives de recherche

Certaines études semblent indiquer que les objectifs des mécanismes d'incitation par les prix traditionnellement suggérés par les économistes pour réduire la demande pour les biens et

services qui génèrent des externalités négatives, comme la taxe sur les carburants, sont en fait largement incompris par le consommateur moyen. Contrairement à ce qu'on pourrait imaginer, les mécanismes coercitifs comme les règles et les lois pourraient bénéficier d'une plus grande acceptabilité sociale (Grimsrud et al. 2020). Pour contribuer scientifiquement à l'inversion de la tendance lourde en faveur de véhicules plus nombreux, plus volumineux et plus énergivores, il serait pertinent d'évaluer et de comparer les répercussions que pourraient avoir différents outils sur la demande québécoise de véhicules. Il serait ainsi très intéressant de comparer (1) une hausse de la taxe sur les carburants à une loi restreignant la durée des prêts automobiles, (2) l'imposition d'une taxe sur les véhicules volumineux à l'interdiction de l'usage de ces véhicules dans les milieux urbains denses ou encore (3) l'autofinancement du programme roulez-électrique par une taxe sur les véhicules énergivores à l'interdiction d'acheter un véhicule avec un moteur à combustion. L'évaluation et la comparaison pourraient notamment être effectuées au moyen d'une enquête sur les préférences déclarées.

En ce qui concerne l'effet rebond, un auteur mentionne l'existence théorique d'un effet rebond de nature psychosocial. Un bien de consommation perçu positivement par la société, que cette perception soit fondée ou non, favoriserait une consommation décomplexée de ce bien (Haan et al., 2006). L'augmentation de la présence de véhicules de plus en plus gros et énergivores sur les routes pourrait contribuer à décomplexer leur usage malgré des mises en garde de plus en plus importantes concernant les changements climatiques. Un auteur a mentionné que certaines réglementations, comme le *Corporate Average Fuel Agreement*, ont eu pour effet d'imposer des normes plus exigeantes aux voitures qu'aux camions légers et de détourner les consommateurs des voitures (Godek, 1997). Certains programmes d'aide financière à l'industrie automobile, comme le *Fonds d'innovation pour le secteur de l'automobile*, pourraient par ailleurs avoir contribué à améliorer davantage l'efficacité énergétique des camions légers que des voitures. Il est possible que tous les véhicules n'aient pas bénéficié de la même amélioration de l'efficacité énergétique et que l'effet rebond soit plus complexe qu'il n'y paraît. C'est-à-dire qu'il découlerait à la fois d'un effet de revenu attribuable à une diminution uniforme de la consommation de carburant, mais aussi d'un effet de substitution du fait du changement des coûts relatifs entre les véhicules. Cette question n'a pas été abordée sur le plan scientifique pour le moment. Il s'agit là de deux perspectives de recherche intéressantes à explorer. Il pourrait notamment être intéressant d'effectuer une analyse coûts-efficacité à savoir s'il vaut mieux subventionner l'amélioration de l'efficacité énergétique (ex. véhicules électriques, programme d'aide à l'industrie automobile) ou taxer la pollution.

En termes de perspectives de recherche, il apparaît incontournable de poursuivre les travaux sur l'abordabilité intégrée des localisations résidentielles. Les travaux devraient viser à produire des évidences empiriques québécoises afin de cerner certains des mécanismes économiques sous-jacents à l'étalement urbain. Étant donné la crise du logement, il apparaît essentiel d'établir des diagnostics plus exhaustifs des problèmes d'abordabilité, en tenant compte des coûts de transport, pour différents types de ménages à l'échelle du territoire. Il apparaît par ailleurs nécessaire de développer des outils d'aide à la localisation des ménages qui leur permettraient de mieux comprendre l'ensemble des répercussions financières des opportunités qui s'offrent à eux. Il apparaît finalement incontournable de mieux mesurer les coûts collectifs, incluant les coûts privés, les coûts publics (ex. infrastructures et services) et de coûts externes (ex. pollution), qui découlent d'un choix de localisation, ainsi que de qualifier leur distribution dans l'espace dans une perspective d'efficacité et d'équité.

3 Rayonnement

Les activités de l'année 2021 ont aussi été touchées par la pandémie. Plusieurs activités ont été annulées ou reportées et les activités de recherche ont été ralenties par les conditions exceptionnelles et limitantes.

3.1 Publications (2021)

Khachman, M., Morency, C., Ciari, F. (2021). Impact of the Geographic Resolution on Population Synthesis Quality, *ISPRS International Journal of Geo-Information* 10 (11), 790.

Laviolette, J., Morency, C., Waygood, O.D., Goulias, K.G. (2021). Car Ownership and the Built Environment: A Spatial Modeling Approach. *Transportation Research Record*, First Published October 21, 2021, <https://doi.org/10.1177/03611981211049409>

Mageau-Béland, J., Morency, C. (2021). Assessing Physical Activity Achievement by Using Transit, *Transportation Research Record*. 2675(8): 506-514

Lefebvre-Ropars, G., Morency, C., Negron-Poblete, P. (2021). Toward A Framework for Assessing the Fair Distribution of Space in Urban Streets, *Transportation Research Record*, First Published February 28, 2021, <https://doi.org/10.1177/0361198121995196>

Deschaintres, E. Morency, C., Trépanier, M. (2021). Measuring Changes in Multimodal Travel Behavior Resulting from Transport Supply Improvement, *Transportation Research Record*, First Published April 15, 2021, <https://doi.org/10.1177/03611981211003104>

Lefebvre-Ropars, G., Morency, C., Negron-Poblete, P. (2021). A Needs-gap analysis of street space allocation, *Journal of Transport and Land Use*, 14(1), 151-170. <https://doi.org/10.5198/jtlu.2021.1808>

Lepage, S., Morency, C. (2021). Impact of Weather, Activities and Service Disruptions on Transportation Demand, *Transportation Research Record* 2021, Vol. 2675(1) 294–304

Roy, F., Morency, C. (2021). Comparing Driving Cycle Development Methods Based on Markov Chains, *Transportation Research Record* 2021, Vol. 2675(3) 212–221

Bourbonnais, P.-L., Morency, C., Trépanier, M., Martel-Poliquin, E. (2021). Transit network design using a genetic algorithm with integrated road network and disaggregated O–D demand data, *Transportation* 48 (1), 95-130.

3.2 Principales participations à des congrès

3.2.1 Congrès annuel du Transportation research board 2020

Laviolette, J., Morency, C., Waygood, O., Goulias, K. (2021). Car Ownership and the Built Environment: A Spatial Modeling Approach, 100th Annual Meeting of the Transportation Research Board, Virtual

Lefebvre-Ropars, G., Morency, C., Negron-Poblete, P. (2021). Towards A Framework for Assessing the Fair Distribution of Space in Urban Streets, 100th Annual Meeting of the Transportation Research Board, Virtual

Garnier, C., Trépanier, M., Morency, C. (2021). Criteria to Transfer Paratransit Trips into a Transit Network, 100th Annual Meeting of the Transportation Research Board, Virtual

Grante, A., Morency, C., Bourdeau, J.-S. (2021). Impacts of Cyclability Features on Optimal Cycling Route, 100th Annual Meeting of the Transportation Research Board, Virtual

Deschaintres, E., Morency C., Trépanier, M. (2021). Measuring Changes in Multimodal Travel Behavior: What Is the Effect of Transport Supply Improvement? 100th Annual Meeting of the Transportation Research Board, Virtual

Morency, C., Verreault, H. (2021). Assessing the Metropolitan Impacts of Teleworking Scenarios, 100th Annual Meeting of the Transportation Research Board, Virtual

4 Références

- Abreu e Silva, J. and Melo, P.C. (2018). Does home-based telework reduce household total travel? A path analysis using single and two worker British households. *Journal of Transport Geography* 73, pp. 148-162
- ACFC. 2016. Financer une auto: Tendances du marché. Place of publication not identified: Financial Consumer Agency of Canada. <https://www.deslibris.ca/ID/10050040> (22 décembre 2020).
- Adnan, M., Pereira, F.C., Azevedo, C.M.L., Basak, K., Lovric, M., Raveau, S., Zhu, Y., Ferreira, J., Zegras, C. & Ben-Akiva, M. (2016). SimMobility: a multi-scale integrated agent-based simulation platform. 95th Annual Meeting of the Transportation Research Board. Washington D.C.
- Ampt, E. (2013). Workshop Synthesis: Longitudinal Methods: Overcoming Challenges and Exploiting Benefits. Dans J. Zmud, M. Lee-Gosselin, M. Munizaga, & J. A. Carrasco (édit.), *Transport Survey Methods* (p. 393-406): Emerald Group Publishing Limited.
- Andreev, P., Salomon, I., & Pliskin, N. (2010). Review: State of teleactivities. *Transportation Research Part C: Emerging Technologies*, 18(1), 3–20
- Anowar, S., Eluru, N. et Miranda-Moreno, L. F. (2014, 2014/07/04). Alternative Modeling Approaches Used for Examining Automobile Ownership: A Comprehensive Review. *Transport Reviews*, 34(4), 441-473. <https://doi.org/https://doi.org/10.1080/01441647.2014.915440>
- Anowar, S., Eluru, N. et Miranda-Moreno, L. F. (2016, May 01). Analysis of vehicle ownership evolution in Montreal, Canada using pseudo panel analysis [journal article]. *Transportation*, 43(3), 531-548. <https://doi.org/https://doi.org/10.1007/s11116-015-9588-z>
- Anowar, S., Yasmin, S., Eluru, N. et Miranda-Moreno, L. F. (2014, September 01). Analyzing car ownership in Quebec City: a comparison of traditional and latent class ordered and unordered models [journal article]. *Transportation*, 41(5), 1013-1039. <https://doi.org/https://doi.org/10.1007/s11116-014-9522-9>
- Balmer, M., Rieser, M., Meister, K., Charypar, D., Lefebvre, N., Nagel, K., & Axhausen, K. (2009). MATSim-T : Architecture and Simulation Times. Dans M. Balmer, M. Rieser, K. Meister, D. Charypar, N. Lefebvre, K. Nagel, K. Axhausen, A. L. Bazzan, & F. Klügl (Éd.), *Multi-Agent Systems for Traffic and Transportation Engineering* (pp. 57-78). Information Science Reference, Hershey, Penn.
- Bandivadekar, A., Cheah, L., Evans, C., Groode, T., Heywood, J., Kasseris, E., ... & Weiss, M. (2008). Reducing the fuel use and greenhouse gas emissions of the US vehicle fleet. *Energy Policy*, 36(7), 2754-2760.
- Bar-Gera, H., Konduri, K., Sana, B., Ye, X., & Pendyala, R. M. (2009). Estimating survey weights with multiple constraints using entropy optimization methods. 88th Annual Meeting of the Transportation Research Board. Washington D.C.
- Bayart, C., Bonnel, P., & Morency, C. (2009). Survey mode integration and data fusion: methods and challenges. Dans P. Bonnel, M. Lee-Gosselin, J. Zmud, & J.-L. Madre (édit.), *Transport survey methods: Keeping up with a changing world* (p. 587-611): Emerald Group Publishing Limited.
- Beckman, J. R., Baggerly, K. A., & McKay, M. D. (1996). Creating synthetic baseline populations. *Transportation Research Part A : Policy and Practice*, 30, 415-429.

- Bennicelli, C. (2012). Des transports actifs pour une ville active. Mémoire présenté à la Commission permanente sur la culture, le patrimoine et les sports, Vivre en ville, Montreal, 28 pages.
- Berdica, K. (2002). An introduction to road vulnerability: what has been done, is done and should be done, *Transport policy*, 9: 117-27
- Berdica, K., Mattsson, L.-G. (2007). Vulnerability: a model-based case study of the road network in Stockholm. in, *Critical infrastructure* (Springer)
- Bhat, C. R. et Guo, J. Y. (2007, 2007/06/01/). A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. *Transportation Research Part B: Methodological*, 41(5), 506-526. <https://doi.org/https://doi.org/10.1016/j.trb.2005.12.005>
- Bhat, C. R., Sen, S. et Eluru, N. (2009). The impact of demographics, built environment attributes, vehicle characteristics, and gasoline prices on household vehicle holdings and use. *Transportation Research Part B: Methodological*, 43(1), 1-18. <https://doi.org/https://doi.org/10.1016/j.trb.2008.06.009>
- Boell, S.K., Campbell, J., Cecez-Kecmanovic, D., & Cheng, J.E. (2013). The Transformative Nature of Telework: A Review of the Literature. *AMCIS*.
- Bonnel, P., & Munizaga, M. A. (2018). Transport survey methods - in the era of big data facing new and old challenges. *Transportation Research Procedia*, 32, 1-15. doi: 10.1016/j.trpro.2018.10.001
- Borysov, S.S. & Rich, J. (2021). Introducing synthetic pseudo panels : application to transport behaviour dynamics. *Transportation*. 48, 2493–2520.
- Borysov, S.S., Rich, J. & Pereira, F.C. (2019). How to generate micro-agents? A deep generative modeling approach to population synthesis. *Transportation Research Part C : Emerging Technologies*. 106, 73–97.
- Bradburn, N., 1978, August. Respondent burden. In *Proceedings of the Survey Research Methods Section of the American Statistical Association* (Vol. 35, pp. 35-40). Alexandria, VA, USA: American Statistical Association.
- Brand, C., Anable, J., & Tran, M. (2013). Accelerating the transformation to a low carbon passenger transport system: The role of car purchase taxes, feebates, road taxes and scrappage incentives in the UK. *Transportation Research Part A: Policy and Practice*, 49, 132-148.
- Briz-Redón, Á., 2021. Respondent Burden Effects on Item Non-Response and Careless Response Rates: An Analysis of Two Types of Surveys. *Mathematics*, 9(17), p.2035.
- Callegaro, M., & Yang, Y. (2018). The role of surveys in the era of “big data”. Dans *The Palgrave handbook of survey research* (p. 175-192): Springer.
- Cao, J. et Cao, X. (2014). The Impacts of LRT, Neighbourhood Characteristics, and Self-selection on Auto Ownership: Evidence from Minneapolis-St. Paul. *Urban Studies*, 51(10), 2068-2087. <https://doi.org/https://doi.org/10.1177/0042098013505887>.
- Cao, X., Mokhtarian, P. L. et Handy, S. L. (2009, 2009/05/01). Examining the Impacts of Residential Self-Selection on Travel Behaviour: A Focus on Empirical Findings. *Transport Reviews*, 29(3), 359-395. <https://doi.org/10.1080/01441640802539195>.

Casati, D., Müller, K., Fourrie, P. J., Erath, A., & Axhausen, K. W. (2015). Synthetic population generation by combining a hierarchical, simulation-based approach with reweighting by generalized raking. *Transportation Research Record: Journal of the Transportation Research Board*, 107-116.

Chaire Mobilité (2021). LES CAMIONS LÉGERS - Facteurs ayant contribué à la transformation du parc de véhicules légers. Tiré de https://legacy.equiterre.org/sites/fichiers/divers/rapport2_chairemobilite.pdf.

Chaire Mobilité (2021). Les camions Légers. Facteurs ayant contribué à la transformation du parc de véhicules légers, rapport déposé à Équiterre, 117 pages, https://legacy.equiterre.org/sites/fichiers/divers/rapport2_chairemobilite.pdf

Chakrabarti, S. Does telecommuting promote sustainable travel and physical activity? *J. Transp. Health* 2018, 9, 19–33.

Chaoqi, Fu, Wang Ying, Wang Xiaoyang, and Gao Yangjun. (2018). Multi-node attack strategy of complex networks due to cascading breakdown, *Chaos, Solitons & Fractals*, 106: 61-66.

Chen, Hongyu, Limao Zhang, Qiong Liu, Hongtao Wang, and Xiaosong Dai. (2021). Simulation-based vulnerability assessment in transit systems with cascade failures, *Journal of Cleaner Production*, 295: 126441.

Chen, T. et He, T. (2021). xgboost: eXtreme Gradient Boosting. <https://cran.r-project.org/web/packages/xgboost/xgboost.pdf>

Cheng Long, Xuewu Chen, Jonas De Vos, Xinjun Lai, Frank Witlox (2019), Applying a random forest method approach to model travel mode choice behavior, *Travel Behaviour and Society*, Volume 14.

Cherchi, E., & Bhat, C. (2018). Workshop Synthesis: Data analytics and fusion in a world of multiple sensing and information capture mechanisms. *Transportation Research Procedia*, 32, 416-420. doi:10.1016/j.trpro.2018.10.059

Choo, S., Mokhtarian, P. L., & Salomon, I. (2005). Does telecommuting reduce vehicle-miles traveled? An aggregate time series analysis for the U.S. *Transportation*, 32(1), 37–64.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: a seasonal-trend decomposition. *Journal of official statistics*, 6(1), 3-73.

Collectivités viables. (2016, octobre 19). Récupéré sur collectivitesviables.org: <https://collectivitesviables.org/articles/potentiel-pietonnier.aspx>

Consumer Agency of Canada. <https://www.deslibris.ca/ID/10050040> (22 décembre 2020).

Currie, G., Delbosc, A (2011). Exploring the trip chaining behaviour of public transport users in Melbourne. *Transp. Policy*.

Dargay, Joyce. 2001. « The effect of income on car ownership: evidence of asymmetry ». *Transportation Research Part A: Policy and Practice* 35(9): 807-21.

Dargay, Joyce. 2001. « The effect of income on car ownership: evidence of asymmetry ». *Transportation Research Part A: Policy and Practice* 35(9): 807-21.

- de Oliveira, Eduardo Leal, Licínio da Silva Portugal, and Walter Porto Junior. (2016). Indicators of reliability and vulnerability: Similarities and differences in ranking links of a complex road system, *Transportation research part A: policy and practice*, 88: 195-208
- de Vos, D., Meijers, E., & van Ham, M.(2018). Working from home and the willingness to accept a longer commute. *The Annals of Regional Science*, 61(2), 375–398.
- DfT, 2020. *Vehicle Licensing Statistics: Annual 2019 (National Statistics)*. Department for Transport, United Kingdom.
- Ding, C., Cao, X. et Wang, Y. (2018, 2018/12/01/). Synergistic effects of the built environment and commuting programs on commute mode choice. *Transportation Research Part A: Policy and Practice*, 118, 104-118. <https://doi.org/https://doi.org/10.1016/j.tra.2018.08.041>
- Efthymiou, D., Farooq, B., Bierlaire, M., & Antoniou, C. (2014). Multidimensional Indicators Analysis for Transport Policy Evaluation. Paper presented at the Transportation Research Board 93rd Annual Meeting.
- El Faouzi, N.-E., & Klein, L. A. (2016). Data Fusion for ITS: Techniques and Research Needs. *Transportation Research Procedia*, 15, 495-512. doi:10.1016/j.trpro.2016.06.042
- Eldér, E. (2014). Residential location and daily travel distances: The influence of trip purpose. *Journal of Transport Geography*, 34, 121–130.
- Eldér, E. (2020). Telework and daily travel: New evidence from Sweden. *Journal of Transport Geography*, 86, 102777.
- Ewing, R. et Cervero, R. (2010). Travel and the Built Environment. *Journal of the American Planning Association*, 76(3), 265-294. <https://doi.org/https://doi.org/10.1080/01944361003766766>
- Farooq, B., Bierlaire, M., Hurtubia, R., & Flötteröd, G. (2013). Simulation based population synthesis. *Transportation Research Part B : Methodological*, 243-263.
- Fisher, S. and Kydoniefs, L., 2001. Using a theoretical model of response burden (RB) to identify sources of burden in surveys. In 12th International Workshop on Household Survey Nonresponse, Oslo, Norway, September (pp. 12-14).
- Fontaras, G., & Dilara, P. (2012). The evolution of European passenger car characteristics 2000–2010 and its effects on real-world CO2 emissions and CO2 reduction policy. *Energy Policy*, 49, 719-730.
- Fournier, N., Christofa, E., Akkinpally, A.P. & Azevedo, C.L. (2021). Integrated population synthesis and workplace assignment using an efficient optimization-based person-household matching method. *Transportation*.
- Friedman, J. H. (2001, 10/1). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189-1232. <https://doi.org/https://doi.org/10.1214/aos/1013203451>
- Friedman, J. H. (2002, 2002/02/28/). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4), 367-378. [https://doi.org/https://doi.org/10.1016/S0167-9473\(01\)00065-2](https://doi.org/https://doi.org/10.1016/S0167-9473(01)00065-2)
- Gao, Lei, Xingquan Liu, Yu Liu, Pu Wang, Min Deng, Qing Zhu, and Haifeng Li. (2019). Measuring road network topology vulnerability by Ricci curvature, *Physica A: Statistical Mechanics and its Applications*, 527: 121071.

- Gärling, T. et Steg, L. (2007). Threats from car traffic to the quality of urban life: problems, causes and solutions. Emerald Group Publishing Limited. <https://doi.org/10.1108/9780080481449>
- Garrido, S., Borysov, S.S., Perreira, F.C. & Rich, J. (2020). Prediction of rare feature combinations in population synthesis : Application of deep generative modelling. *Transportation Research Part C: Emerging Technologies*, 120, 102787.
- Giraud, A., Trépanier, M., Morency, C., & Légaré, F. (2016). Data Fusion of APC, Smart Card and GTFS to Visualize Public Transit Use. CIRRELT. Tiré de <https://www.cirrelt.ca/documentstravail/cirrelt-2016-54.pdf>
- Godek, Paul E. 1997. « The Regulation of Fuel Economy and the Demand for “Light Trucks” ». *The Journal of Law and Economics* 40(2): 495-510.
- Goodwin. 1992. « A Review of New Demand Elasticities with Special Reference to Short and Long Run Effect of Price Changes ». *Journal of Transport Economics and Policy* 26(2): 155-69.
- Goodwin. 1992. « A Review of New Demand Elasticities with Special Reference to Short and Long Run Effect of Price Changes ». *Journal of Transport Economics and Policy* 26(2): 155-69.
- Görizt, A.S., 2014. Determinants of the starting rate and the completion rate in online panel studies1. *Online Panel Research: Data Quality Perspective*, A, pp.154-170.
- Grapperon, A., Farooq, B., & Trépanier, M. (2016). Information fusion of smart card data with travel survey. CIRRELT. Tiré de <https://www.cirrelt.ca/documentstravail/cirrelt-2016-59.pdf>
- Greene, David L., James R. Kahn, et Robert C. Gibson. 1999. « Fuel Economy Rebound Effect for U.S. Household Vehicles ». *The Energy Journal* 20(3). <http://www.iaee.org/en/publications/ejarticle.aspx?id=1314> (2 février 2021).
- Grimsrud, Kristine M., Henrik Lindhjem, Ingvild Vestre Sem, et Knut Einar Rosendahl. 2020. « Public acceptance and willingness to pay cost-effective taxes on red meat and city traffic in Norway ». *Journal of Environmental Economics and Policy* 9(3): 251-68.
- Groves, R.M., Couper, M.P., Presser, S., Singer, E., Tourangeau, R., Acosta, G.P. and Nelson, L., 2006. Experiments in producing nonresponse bias. *International Journal of Public Opinion Quarterly*, 70(5), pp.720-736.
- Guo, J.Y., & Bhat, C.R. (2007). Population synthesis for microsimulating travel behavior. *Transportation Research Record : Journal of the Transportation Research Board*, 92-101.
- Guo, Z. (2013). Does residential parking supply affect household car ownership? The case of New York City. *Journal of Transport Geography*, 26, 18-28. <https://doi.org/https://doi.org/10.1016/j.jtrangeo.2012.08.006>
- Hars, A. (2015). Self-driving cars: The digital transformation of mobility. In *Marktplätze im Umbruch* (pp. 539-549). Springer Vieweg, Berlin, Heidelberg.
- Hastie, T., Tibshirani, R. et Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media. <http://web.stanford.edu/~hastie/ElemStatLearn/>
- Hensher, D., Louviere, J., & Swait, J. (1998). Combining sources of preference data. *Journal of Econometrics*, 89(1), 197-221. doi:10.1016/S0304-4076(98)00061-X

- Hilbrecht, M., S.M. Shaw, L.C. Johnson et J. Andrey (2013). Remixing work, family and leisure: teleworkers' experiences of everyday life. *New Technology, Work and Employment*, 28(2), 130-144.
- Holzapfel, H. (1986). *Trip relationships in urban areas*. Aldershot: Gower.
- Hugosson, M. B., Algers, S., Habibi, S., & Sundbergh, P. (2016). Evaluation of the Swedish car fleet model using recent applications. *Transport Policy*, 49, 30-40.
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice* (2e éd.). Melbourne, Australia: OTexts.
- Islam, M. T. (2010). Unraveling the relationship between trip chaining and mode choice using structural equation models (pp. [12], 122 p.).
- Islam, M. T., & Habib, K. M. N. (2012). Unraveling the relationship between trip chaining and mode choice: evidence from a multi-week travel diary. *Transportation Planning and Technology*, 35(4), pp 409-426. 138
- Jenelius, E., Mattsson, L.-G. (2012). Road network vulnerability analysis of area-covering disruptions: A grid-based approach with case study, *Transportation research part A: policy and practice*, 46: 746-60
- Jenelius, E., Petersen, T., Mattsson, L.-G. (2006). Importance and exposure in road network vulnerability analysis, *Transportation research part A: policy and practice*, 40: 537-60.
- Jiang, Y., Gu, P., Chen, Y., He, D. et Mao, Q. (2017, 2017/05/01/). Influence of land use and street characteristics on car ownership and use: Evidence from Jinan, China. *Transportation Research Part D: Transport and Environment*, 52, 518-534. <https://doi.org/https://doi.org/10.1016/j.trd.2016.08.030>
- Johnsen, M., Brandt, O., Garrido, S. & Pereira, F. (2022). Population synthesis for urban resident modelling using deep generative models. *Neural Computing and Applications*. 34:4677–4692.
- Khachman, M., Morency, C. & Ciari, F. (2021). Impact of the Geographic Resolution on Population Synthesis Quality. *ISPRS International Journal of Geo-Information*, 10(11):790.
- Kim, S.-N.(2017). Is telecommuting sustainable? An alternative approach to estimating the impact of home-based telecommuting on household travel. *International Journal of Sustainable Transportation*, 11(2), 72–85.
- Kim, S.-N., Choo, S., & Mokhtarian, P. L.(2015). Home-based telecommuting and intra-household interactions in work and non-work travel: A seemingly unrelated censored regression approach. *Transportation Research Part A: Policy and Practice*, 80, 197–214.
- Kitamura, R., Chen, C., Pendyala, R.M. & Narayanan, R. (2000). Micro-simulation of daily activity-travel patterns for travel demand forecasting. *Transportation*, 27, 25–51.
- Kitou, E., & Horvath, A.(2006). Transportation Choices and Air Pollution Effects of Telework. *Journal of Infrastructure Systems*, 12(2), 121–134.
- Konduri, K. C., You, D., Garikapati, V. M., & Pendyala, R. (2016). Enhanced synthetic population generator that accomodates control variables at multiple geographic resolutions. *Transportation Research Record: Journal of the Transportation Research Board*, 40-50.

- Krygsman, S., Arentze, T., Timmermans, H (2007). Capturing tour mode and activity choice interdependencies: a co-evolutionary logit modelling approach. *Transp. Res. Part A Policy Pract.* 41(10), 913–933.
- Kusakabe, T., & Asakura, Y. (2014). Behavioural data mining of transit smart card data: A data fusion approach. *Transportation Research Part C: Emerging Technologies*, 46, 179-191. doi:10.1016/j.trc.2014.05.012
- Lachapelle, U, Noland, R.B. (2012). Does the commute mode affect the frequency of walking behavior? *The public transit link*, *Transport Policy*, Volume 21, May 2012, Pages 26-36
- Lachapelle, U., Tanguay, G. A., & Neumark-Gaudet, L. (2018). Telecommuting and sustainable travel: Reduction of overall travel time, increases in non-motorised travel and congestion relief? *Urban Studies*, 55(10), 2226-244.
- Lally P, Van Jaarsveld CH, Potts HW, Wardle J. How are habits formed: Modelling habit formation in the real world. *European journal of social psychology*. 2010 Oct;40(6):998-1009.
- Larson,W., Zhao, W.,(2017). Telework: urban form, energy consumption, and greenhouse gas implication. *Economic Inquiry* 55 (2), 625–1166
- Lavolette, J., Morency, C. et Waygood, E. O. D. (2022). A kilometer or a mile? Does buffer size matter when it comes to car ownership? 2022 Transportation Research Board Annual Meeting, Washington DC.
- Lavolette, J., Morency, C. et Waygood, E. O. D. (2022, en révision). A kilometer or a mile? Does buffer size matter when it comes to car ownership? *Journal of Transport Geography*.
- Li, Xianghua, Jingyi Guo, Chao Gao, Zhen Su, Deng Bao, and Zili Zhang. (2018). Network-based transportation system analysis: A case study in a mountain city, *Chaos, Solitons & Fractals*, 107: 256-65.
- Lindsey, Marshall, Joseph L. Schofer, Pablo Durango-Cohen, et Kimberly A. Gray. 2011. « The Effect of Residential Location on Vehicle Miles of Travel, Energy Consumption and Greenhouse Gas Emissions: Chicago Case Study ». *Transportation Research Part D: Transport and Environment* 16(1): 1-9.
- Litman, T. (2011). *If Health Matters: Integrating Health Objectives in Transportation Planning*, Victoria Transport Policy Institute:1-42.
- Litman, T. (2009). *Transportation cost and benefit analysis*. Victoria Transport Policy Institute, 1-19.
- Lovelace, R., Birkin, M., Cross, P., & Clarke, M. (2016). From Big Noise to Big Data: Toward the Verification of Large Data sets for Understanding Regional Retail Flows. *Geographical Analysis*, 48(1), 59-81. doi:10.1111/gean.12081
- Lund, B., 2021. The questionnaire method in systems research: an overview of sample sizes, response rates and statistical approaches utilized in studies. *VINE Journal of Information and Knowledge Management Systems*.
- Ma, Fei, Fei Liu, Kum Fai Yuen, Polin Lai, Qipeng Sun, and Xiaodan Li. (2019). Cascading failures and vulnerability evolution in bus–metro complex bilayer networks under rainstorm weather conditions, *International Journal of Environmental Research and Public Health*, 16: 329.

- Ma, Fei, Wenjing Shi, Kum Fai Yuen, Qipeng Sun, Xiaobo Xu, Yongjie Wang, and Zuohang Wang. (2020). Exploring the robustness of public transportation for sustainable cities: A double-layered network perspective, *Journal of Cleaner Production*, 265: 121747.
- Ma, Fei, Yuan Liang, Kum Fai Yuen, Qipeng Sun, Yujie Zhu, Yixuan Wang, and Wenjing Shi. (2020). Assessing the vulnerability of urban rail transit network under heavy air pollution: A dynamic vehicle restriction perspective, *Sustainable Cities and Society*, 52: 101851.
- Mackie, P. J., Wardman, M., Fowkes, A. S., & Whelan, G. (2003). Values of Travel Time Savings in the UK: Institute for Transport Studies, University of Leeds.
- Magnusson, E. (2005). Véhicules utilitaires sport: moteur de changement (Documet analytique No. No 11-621-MIF au catalogue — No 020). Statistique Canada.
- Martel-Poliquin, É. (2012). Mieux comprendre les déterminants du choix modal. (M.Sc.A Mémoire), École Polytechnique de Montréal, Montréal, Canada.
- Mattsson, Lars-Göran, and Erik Jenelius. (2015). Vulnerability and resilience of transport systems—A discussion of recent research, *Transportation research part A: policy and practice*, 81: 16-34
- McCarthy, Patrick, et Richard Tay. 1998. « New Vehicle Consumption and Fuel Efficiency: A Nested Logit Approach ». *Transportation Research Part E: Logistics and Transportation Review* 34(1): 39-51.
- McCarthy, Patrick, et Richard Tay. 1998. « New Vehicle Consumption and Fuel Efficiency: A Nested Logit Approach ». *Transportation Research Part E: Logistics and Transportation Review* 34(1): 39-51.
- Melo, P. C., & de Abreu e Silva, J. (2017). Home telework and household commuting patterns in Great Britain. *Transportation Research Part A: Policy and Practice*, 103, 1–24. <https://doi.org/10.1016/j.tra.2017.05.011>
- Menezes, E., Gori, A., Carvalho, C.S.,(2017). Effectiveness of low-carbon development strategies: Evaluation of policy scenarios for the urban transport sector in a Brazilian megacity. *Technological Forecasting and Social Change*, Volume 114, Pages 226-241
- Miller, E. J., Srikukenthiran, S., & Chung, B. (2018). Workshop Synthesis: Household travel surveys in an era of evolving data needs for passenger travel demand. *Transportation Research Procedia*, 32, 374-382. doi:<https://doi.org/10.1016/j.trpro.2018.10.067>
- Miller, E., Roorda, M., & Carrasco, J. (2005). A tour-based model of travel mode choice. *Transportation*, 32(4), 399-422. doi: 10.1007/s11116-004-7962-3
- Moeckel, R. et Yang, D. (2016). Car Ownership and the Impact of Built Environment, Demographics and Transport Systems. *Journal of Civil Engineering and Architecture*, 10(5). <https://doi.org/https://doi.org/10.17265/1934-7359/2016.05.008>
- Montréal, C. M. d. (2021). Plan Métropolitain d'Aménagement et de Développement.
- Moore, R. and Varghese, L., 2021. An Examination of the Role of Funny Reminders and Embedded Survey Items on Response Rate. *Social Science Computer Review*, 39(1), pp.162-172.
- Morency, Bourdeau, Milord (2021). Les camions légers: définitions et évolution de l'offre, 68 p.

Morency, Bourdeau, Verreault, Milord (à venir). Les camions légers : impacts de la transformation du parc de véhicules, 200 p. <https://www.equiterre.org/fr/ressources/actualite-comprendre-la-hausse-des-camions-legers-au-canada-afin-de-renverser-la-tendance>

Morency, C., Trépanier, M., Saunier, N., Verreault, H., & Bourdeau, J.-S. (2018). Using 5 parallel passive data streams to report on a wide range of mobility options. *Transportation Research Procedia*, 32, 82-92.

Morency, Catherine, Tremblay-Racicot, Fanny, Milord, Brigitte (2022). L'abordabilité intégrée des localisations résidentielles: État des connaissances scientifiques, des données, des indicateurs et des outils, CIRRELT-2022-06, 57 pages.

Morency, Milord, Bourdeau (2021). Les camions légers : facteurs ayant contribué à la transformation du parc de véhicules légers, 177 p.

Morency, Tremblay-Racicot, Milord (2022). L'abordabilité intégrée des localisations résidentielles : état des connaissances scientifiques, des données, des indicateurs et des outils, 57 p. <https://www.polymtl.ca/mobilite/publications>

Moreno, A.T. & Moeckel, R. (2018). Population Synthesis Handling Three Geographical Resolutions. *International Journal of Geo-Information*.

Müller, K., & Axhausen, K. W. (2011). Hierarchical IPF : Generating a synthetic population for Switzerland. 51st Congress of the European Regional Science Association. Barcelona.

Munizaga, M., Devillaine, F., Navarrete, C., & Silva, D. (2014). Validating travel behavior estimated from smartcard data. *Transportation Research Part C: Emerging Technologies*, 44, 70-79. doi:10.1016/j.trc.2014.03.008

Musti, S., & Kockelman, K. M. (2011). Evolution of the household vehicle fleet: Anticipating fleet composition, PHEV adoption and GHG emissions in Austin, Texas. *Transportation Research Part A: Policy and Practice*, 45(8), 707-720.

National Research Council. 2002. Effectiveness and Impact of Corporate Average Fuel Economy (CAFE) Standards. https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/162944_web.pdf (1 novembre 2020).

Nicholson, Walter, et Christopher Snyder. 2008. *Microeconomic Theory: Basic Principles and Extension*. 10th éd. Mason: South-Western.

Nicholson, Walter, et Christopher Snyder. 2008. *Microeconomic Theory: Basic Principles and Extension*. 10th éd. Mason: South-Western.

Nolan, Anne. 2010. « A dynamic analysis of household car ownership ». *Transportation Research Part A: Policy and Practice* 44(6): 446-55.

Nolan, Anne. 2010. « A dynamic analysis of household car ownership ». *Transportation Research Part A: Policy and Practice* 44(6): 446-55.

Ojala, S., & Pyöriä, P.,(2018). Mobile knowledge workers and traditional mobile workers: Assessing the prevalence of multi-locational work in Europe. *Acta Sociologica*, 61(4), 402–418.

Opland, L., 2007. Size classification of passenger cars: Pre-study on how to size classify passenger cars by inventorying the existing classification models. Chalmers University of Technology, Göteborg.

- Ortúzar, J. de D., Armoogum, J., Madre, J. L., & Potier, F. (2011). Continuous Mobility Surveys: The State of Practice. *Transport Reviews*, 31(3), 293-312. doi:10.1080/01441647.2010.510224
- Pan, Shouzheng, Hai Yan, Jia He, and Zhengbing He. (2021). Vulnerability and resilience of transportation systems: A recent literature review, *Physica A: Statistical Mechanics and its Applications*, 581: 126235
- Pirdavani, A., Bellemans, T., Brijs, T., Kochan, B., & Wets, G.(2014). Assessing the road safety impacts of a teleworking policy by means of geographically weighted regression method. *Journal of Transport Geography*, 39, 96–110.
- Potoglou, D. et Kanaroglou, P. S. (2008). Modelling car ownership in urban areas: a case study of Hamilton, Canada. *Journal of Transport Geography*, 16(1), 42-54. <https://doi.org/https://doi.org/10.1016/j.jtrangeo.2007.01.006>
- Potoglou, Dimitris, et Pavlos S. Kanaroglou. 2007. « Household Demand and Willingness to Pay for Clean Vehicles ». *Transportation Research Part D: Transport and Environment* 12(4): 264-74.
- Prelec, Drazen, et Duncan Simester. 2001. « Always Leave Home Without It: A Further Investigation of the Credit-Card Effect on Willingness to Pay ». : 8.
- Primerano, F., Taylor, M. A. P., Pitaksringkarn, L., & Tisato, P. (2008). Defining and Understanding Trip Chaining Behaviour. *Transportation: Planning, Policy, Research, Practice*, 35(1), pp 55-72.
- Provencher B, Bishop RC (2004). Does accounting for preference heterogeneity improve the forecasting of a random utility model? A case study. *Journal of Environmental Economics and Management*. 2004 Jul 1;48(1):793-810.
- Ravalet, E., Rérat, P.,(2019). Teleworking: Decreasing Mobility or Increasing Tolerance of Commuting Distances?. *Built Environment*, Volume 45, Number 4, December 2019, pp. 582-602
- Report from a brainstorming session held as part of the Mobilizing Justice project
- Ressources Naturelles Canada. (2011). Canadian Vehicle Survey - Summary Report 2009. Ottawa: Office de l'Énergie - Ressources Naturelles Canada. Tiré de <http://oe.nrcan.gc.ca/publications/statistics/cvs09/pdf/cvs09.pdf>.
- Ressources Naturelles Canada. (2022). Outil de recherche pour les cotes de consommation de carburant. Tiré de <https://fcr-ccc.nrcan-rncan.gc.ca/fr>.
- Revilla, M., 2017. Analyzing survey characteristics, participation, and evaluation across 186 surveys in an online opt-in panel in Spain. *methods, data, analyses*, 11(2), p.28.
- Rhee, H.-J. (2009). Telecommuting and urban sprawl. *Transportation Research Part D: Transport and Environment*, 14(7), 453–460.
- Rolstad, S., Adler, J. and Rydén, A., 2011. Response burden and questionnaire length: is shorter better? A review and meta-analysis. *Value in Health*, 14(8), pp.1101-1108.
- Roorda, M. J., Passmore, D., & Miller, E. J. (2009). Including Minor Modes of Transport in a Tour-Based Mode Choice Model with Household Interactions. *Journal of Transportation Engineering*, 135(12), pp 935-945.
- Rudžionienė, J., Grigas, V., Enwald, H. and Kortelainen, T., 2018. Drop out factors in data literacy and research data management survey: experiences from Lithuania and Finland. *Informacijos mokslai*, 82, pp.115-130.

Saadi, I., Mustafa, A., Teller, J., Farooq, B. & Cools, M. (2016). Hidden Markov model-based population synthesis. *Transportation Research Part B: Methodological*, 90, 1–21.

SAAQ. (2021). Nombre de véhicules en circulation selon le type d'utilisation, le type de véhicule et l'âge du véhicule, Québec et régions administratives. Tiré de http://www.bdso.gouv.qc.ca/pls/ken/ken213_afich_tabl.page_tabl?p_iden_tran=REPERAU78HS14-47196194527m46H&p_lang=1&p_m_o=SAAQ&p_id_ss_domn=718&p_id_raprt=3372.

SAAQ. (2022). Véhicules en circulation - Documentation des données ouvertes. Tiré de <https://saaq.gouv.qc.ca/donnees-ouvertes/vehicules-circulation/vehicules-circulation-documentation.pdf>.

Sakshaug, J.W., Vicari, B. and Couper, M.P., 2019. Paper, e-mail, or both? Effects of contact mode on participation in a web survey of establishments. *Social Science Computer Review*, 37(6), pp.750-765.

Salon D, Conway MW, da Silva DC, Chauhan RS, Derrible S, Mohammadian AK, Khoeini S, Parker N, Mirtich L, Shamshiripour A, Rahimi E. The potential stickiness of pandemic-induced behavior changes in the United States. *Proceedings of the National Academy of Sciences*. 2021 Jul 6;118(27).

Salvini, P. A., & Miller, E. J. (2005). An Operational Prototype of a Comprehensive Microsimulation Model of Urban Systems. *Networks and Spatial Economics*, 217-234.

Sarker, R.I., Mailer, M., Sikder, S.K. (1 March 2019). Walking to a public transport station: Empirical evidence on willingness and acceptance in Munich, Germany, *Smart and Sustainable Built Environment*, Volume 9, Issue 1.

Schmid, B. and Axhausen, K.W., 2019. Predicting response rates of all and recruited respondents: A first attempt. *Transport Findings*.

Scott, D. M., Novak, D. C., Aultman-Hall, L., Guo, F. (2006). Network robustness index: A new method for identifying critical links and evaluating the performance of transportation networks, *Journal of Transport Geography*, 14: 215-27.

Shabanpour, R. Golshani, N. Tayarani, M.; Auld, J. Mohammadian. (2018). A. Analysis of telecommuting behavior and impacts on travel demand and the environment. *Transport. Res. Part D Transp. Environ.* 2018, 62, 563–576.

Sicotte, Gabriel & Morency, Catherine & Farooq, Bilal. (2017). Comparison Between Trip and Trip Chain Models: Evidence from Montreal Commuter Train Corridor.

Sioui, L., Morency, C. et Trépanier, M. (2013, 2013/01/01). How Carsharing Affects the Travel Behavior of Households: A Case Study of Montréal, Canada. *International Journal of Sustainable Transportation*, 7(1), 52-69. <https://doi.org/https://doi.org/10.1080/15568318.2012.660109>

Small, Kenneth A., et Kurt Van Dender. 2007. « Fuel Efficiency and Motor Vehicle Travel: The Declining Rebound Effect ». *The Energy Journal* 28(1). <http://www.iaee.org/en/publications/ejarticle.aspx?id=2176> (2 février 2021).

Smyth, J.D., Dillman, D.A., Christian, L.M. and Stern, M.J., 2006. Comparing check-all and forced-choice question formats in web surveys. *Public opinion quarterly*, 70(1), pp.66-7

- Snelder, M, Van Zuylen, HJ, Immers, LH. (2012). A framework for robustness analysis of road networks for short term variations in supply, *Transportation research part A: policy and practice*, 46: 828-42.
- Spurr, T., Chu, A., Chapleau, R., & Piché, D. (2015). A Smart Card Transaction “Travel Diary” to Assess the Accuracy of the Montréal Household Travel Survey. *Transportation Research Procedia*, 11, 350-364. doi:10.1016/j.trpro.2015.12.030
- Statistiques Canada (2022). Tableau 20-10-0001-01. Ventes de véhicules automobiles neufs, <https://doi.org/10.25318/2010000101-fra>
- Stevens, M. R. (2017, 2017/01/02). Does Compact Development Make People Drive Less? *Journal of the American Planning Association*, 83(1), 7-18. <https://doi.org/10.1080/01944363.2016.1240044>
- Stiglitz, Joseph, Carl Walsh, et Jean-Dominique Lafay. 2007. *Principes d'économie moderne*. 3e édition. Bruxelles: De Boeck.
- Stiglitz, Joseph, Carl Walsh, et Jean-Dominique Lafay. 2007. *Principes d'économie moderne*. 3e édition. Bruxelles: De Boeck.
- Sullivan, J. L., Novak, D.C., Aultman-Hall, L., Scott, D.M. (2010). Identifying critical road segments and measuring system-wide robustness in transportation networks with isolating links: A link-based capacity-reduction approach, *Transportation research part A: policy and practice*, 44: 323-36
- Sun, L., & Erath, A. (2015). A Bayesian network approach for population synthesis. *Transportation Research Part C*, 49-62.
- Sun, L., Erath, A., Cai, M. (2018). A hierarchical mixture modeling framework for population synthesis. *Transportation Research Part B: Methodological*, 114, 199–212.
- Tavares, A. I. (2017). Telework and health effects review. *International Journal of Healthcare*, 3(2), 30.
- Taylor, Michael. (2017). *Vulnerability analysis for transportation networks* (Elsevier).
- Templ, M., Meindl, B., Kowarik, A., Dupriez, O. (2017). Simulation of Synthetic Complex Data: The R Package SimPop. *Journal of Statistical Software*, 79, 1-38.
- Thomas, J.A., Walton, D., 2008. Vehicle Size and Driver Perceptions of Safety. *International Journal of Sustainable Transportation* 2, 260–273. <https://doi.org/10.1080/15568310701359015>
- Tran, Vu Hieu, Siew Ann Cheong, and Ngoc Dung Bui. 2019. Complex network analysis of the robustness of the Hanoi, Vietnam bus network, *Journal of Systems Science and Complexity*, 32: 1251-63.
- Transport Canada (2012). *Guide to the Canadian Vehicle Specifications Database*.
- Valiquette, F. (2010). *Typologie des chaînes de déplacements et modélisation descriptive des systèmes d'activités des personnes*. (M.Sc.A Mémoire), École Polytechnique de Montréal, Montréal, Canada.
- Vanoutrive, T., L. Van Malderen, B. Jourquin, I. Thomas, A. Verhetsel et F. Witlox,(2010). Mobility Management Measures by Employers: Overview and Exploratory Analysis for Belgium. *European Journal of Transport and Infrastructure Research*, 10(2), 121-141.

- Venigalla, M. (2004). Household travel survey data fusion issues. Communication présentée à Resource Paper, National Household Travel Survey Conference: Understanding Our Nation's Travel (vol. 1).
- Verreault, H., & Morency, C. (2011). Transcending the Typical Weekday with Large-Scale Single-Day Survey Samples. *Transportation Research Record*, (2230), 38-47. doi:10.3141/2230-05
- Verreault, H., & Morency, C. (2018). Integration of a phone-based household travel survey and a web-based student travel survey. *Transportation*, 45(1), 89-103. doi:10.1007/s11116-016-9726-2
- Verzosa, N.M., 2020. Let's Play for the Future of Transport: The Gamification Effect on Travel Survey Participation, Acceptance, and Data Quality (Doctoral dissertation, University of Sydney).
- Vilhelmson, B. and Thulin, E., (2016), Who and where are the flexible workers? Exploring the current diffusion of telework in Sweden. *New Technology, Work and Employment*, 31:77-96.
- Vleugels, I. V., Hans; Steenberg, Thérèse; Vande Walle, Stefan; Cornelis, Eric; Castaigne, Marie. (2005). Déterminants des choix modaux dans les chaînes de déplacements. (CP/B8/421). Belgique: Politique scientifique fédérale.
- von Behren, S., Bönisch, L., Niklas, U. et Chlond, B. (2020). Revealing Motives for Car Use in Modern Cities—A Case Study from Berlin and San Francisco. *Sustainability*, 12(13). <https://doi.org/https://doi.org/10.3390/su12135254>
- Waddell, P. (2002). UrbanSim: Modeling Urban Development for Land Use, Transportation, and Environmental Planning. *Journal of the American Planning Association*, 297-314.
- Wang, K., Hossain, S., & Habib, K. N. (2021). A hybrid data fusion methodology for household travel surveys to reduce proxy biases and under-representation of specific sub-group of population. *Transportation*. doi:10.1007/s11116-021-10228-x
- Wang, X. Q., Shao, C. F., Yin, C. Y. et Zhuge, C. X. (2018, Sep). Exploring the Influence of Built Environment on Car Ownership and Use with a Spatial Multilevel Model: A Case Study of Changchun, China [Article]. *International Journal of Environmental Research and Public Health*, 15(9), 14, article no 1868. <https://doi.org/https://doi.org/10.3390/ijerph15091868>
- Wang, X., Smith, K., & Hyndman, R. (2006). Characteristic-Based Clustering for Time Series Data. *Data Mining and Knowledge Discovery*, 13(3), 335-364. doi:10.1007/s10618-005-0039-x
- Williamson, P., Birkin, M., & Rees, P. H. (1998). The estimation of population microdata by using data from small area statistics and samples of anonymised records. *Environment and Planning Part A*, 785-816.
- Xianyu, J. (2013). An Exploration of the Interdependencies between Trip Chaining Behavior and Travel Mode Choice. *Procedia - Social and Behavioral Sciences*, 96(0), 1967-1975. doi:
- Yan, T., Conrad, F.G., Tourangeau, R. and Couper, M.P., 2011. Should I stay or should I go: The effects of progress feedback, promised task duration, and length of questionnaire on completing web surveys. *International Journal of Public Opinion Research*, 23(2), pp.131-147.
- Yang, L., Shen, Q., Li, Z (2016). Comparing travel mode and trip chain choices between holidays and weekdays. *Transp. Res. Part A Policy Pract.* 91, 273–285.
- Ye, C., Konduri, K., Pendyala, R. M., Sana, B., & Waddell, P. (2009). A methodology to match distributions of both household and person attributes in the generation of synthetic populations. *Transportation Research Board*.

- Ye, X., Pendyala, R. M., & Gottardi, G. (2007). An exploration of the relationship between mode choice and complexity of trip chaining patterns. *Transportation Research Part B: Methodological*, 41(1), 96-113.
- Yin, C., Shao, C. et Wang, X. (2018). Built Environment and Parking Availability: Impacts on Car Ownership and Use. *Sustainability*, 10(7). <https://doi.org/https://doi.org/10.3390/su10072285>
- Zhang, Hong, and Yangang Yao. (2019). An integrative vulnerability evaluation model to urban road complex network, *Wireless Personal Communications*, 107: 193-204
- Zhang, Jianhua, Ziqi Wang, Shuliang Wang, Wenchao Shao, Xun Zhao, and Weizhi Liu. (2021). Vulnerability assessments of weighted urban rail transit networks with integrated coupled map lattices, *Reliability Engineering & System Safety*, 214: 107707.
- Zhang, Jing, and Zhi-Hui Zhang. (2019). A Review of the Research Methods on Vulnerability of Transportation System, *CICTP 2019*: 4294-305
- Zhang, Lu, Yannan Zhao, Dongli Chen, and Xinhuan Zhang. (2021). Analysis of Network Robustness in Weighted and Unweighted Approaches: A Case Study of the Air Transport Network in the Belt and Road Region, *Journal of Advanced Transportation*, 2021
- Zhang, W., Zhao, Y., Cao, X., Lu, D. et Chai, Y. (2020, 2020/09/01/). Nonlinear effect of accessibility on car ownership in Beijing: Pedestrian-scale neighborhood planning. *Transportation Research Part D: Transport and Environment*, 86, 102445. <https://doi.org/https://doi.org/10.1016/j.trd.2020.102445>
- Zhu, P. (2012). Are telecommuting and personal travel complements or substitutes? *The Annals of Regional Science*, 48(2), 619–639.
- Zhu, P. (2013). Telecommuting, Household Commute and Location Choice. *Urban Studies* 50(12), 2441-2459.
- Zhu, P., & Mason, S. G. (2014). The impact of telecommuting on personal vehicle usage and environmental sustainability. *International Journal of Environmental Science and Technology*, 11(8), 2185–2200.
- Zhu, S., Amirjamshidi, G., & Roorda, M. J. (2018). Data Fusion of Commercial Vehicle GPS and Roadside Intercept Survey Data. *Transportation Research Record*, 2672(44), 10-20. doi:10.1177/0361198118768516
- Zhuge, C., Li, X., Ku, C.-A., Gao, J., & Zhang, H. (2017). A Heuristic-Based Population Synthesis Method for Micro-Simulation in Transportation. *KSCE Journal of Civil Engineering*, 2373-2383.