



Titre: Modélisation des interactions entre les modes de transport par
Title: l'intégration de sources diversifiées de données

Auteur: Elodie Deschaintres
Author:

Date: 2023

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Deschaintres, E. (2023). Modélisation des interactions entre les modes de
Citation: transport par l'intégration de sources diversifiées de données [Thèse de doctorat,
Polytechnique Montréal]. PolyPublie. <https://publications.polymtl.ca/53378/>

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/53378/>
PolyPublie URL:

**Directeurs de
recherche:** Catherine Morency, & Martin Trépanier
Advisors:

Programme: Génie civil
Program:

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

**Modélisation des interactions entre les modes de transport par l'intégration de
sources diversifiées de données**

ELODIE DESCHAINRES

Département des génies civil, géologique et des mines

Thèse présentée en vue de l'obtention du diplôme de *Philosophiæ Doctor*

Génie civil

Février 2023

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Cette thèse intitulée :

Modélisation des interactions entre les modes de transport par l'intégration de sources diversifiées de données

présentée par **Elodie DESCHAINES**

en vue de l'obtention du diplôme de *Philosophiæ Doctor*

a été dûment acceptée par le jury d'examen constitué de :

Francesco CIARI, président

Catherine MORENCY, membre et directrice de recherche

Martin TRÉPANIÉ, membre et codirecteur de recherche

Bruno AGARD, membre

Marco DIANA, membre externe

DÉDICACE

À ma grand-mère paternelle, pour m'avoir inspiré l'envie d'un meilleur dessein.

À mes parents, pour m'avoir donné les moyens de prendre ce chemin.

À mon conjoint, sans qui cette thèse n'aurait jamais eu de fin.

REMERCIEMENTS

Je tiens tout d'abord à remercier mes deux directeurs de thèse, Catherine Morency et Martin Trépanier. Merci de m'avoir fait confiance pour ce si beau sujet, de m'avoir offert autant d'opportunités de développement professionnel, et merci de m'avoir fait bénéficier de votre grande expertise et réputation dans le domaine du transport. En effet, cette recherche n'aurait pas été possible sans toutes vos relations dans le milieu.

Cela m'amène à remercier les partenaires et collaborateurs de la Chaire Mobilité et du CIRRELT pour m'avoir donné accès à de multiples bases de données. Je remercie tout particulièrement Kim Mc Donough et Élisabeth Bouchard d'avoir répondu aussi rapidement à mes nombreuses requêtes sur les données de la STM.

Je suis également très reconnaissante envers les trois programmes de bourse qui ont financé cette thèse : la Chaire de recherche du Canada (CRC) sur la mobilité des personnes, les Fonds de recherche du Québec - Nature et technologie (FRQNT), ainsi que l'Institut de valorisation des données (IVADO). Ce soutien financier a grandement facilité la poursuite de mes études.

De plus, je remercie tous les associés de recherche pour leur aide dans la réalisation de ce projet, notamment Jean-Simon Bourdeau et Amaury Philippe pour le prétraitement de différentes sources de données passives, Hubert Verreault pour la manipulation des données des enquêtes Origine-Destination, ainsi que Pierre-Léo Bourbonnais et Geneviève Bastien pour l'utilisation et le développement de la plateforme Transition.

Je salue également tous mes collègues de Poly, anciens et actuels, qui ont agrémenté ces dernières années de tant de bons souvenirs. En particulier, je remercie mes deux acolytes de doctorat, Gabriel Lefebvre-Ropars et Jérôme Laviolette : la belle dynamique de notre trio de « superstars » a grandement participé à mon épanouissement professionnel et personnel.

Je remercie évidemment mes parents, Michel et Christine Deschaintres, pour leurs encouragements et leur soutien sans faille tout au long de mon parcours académique. Merci de m'avoir ouvert les portes vers d'autres horizons, malgré le sacrifice de devoir me laisser partir de plus en plus loin.

Pour terminer, merci infiniment à mon conjoint, Alexis Csejtei, de m'avoir soutenue moralement durant toutes ces années d'études supérieures. Sa présence à mes côtés dans les bons comme dans les mauvais moments restera une des plus belles preuves d'amour qu'il m'ait faites.

RÉSUMÉ

Cocktail transport, mobilité intégrée et multimodalité/intermodalité sont des termes de plus en plus répandus dans le domaine de la mobilité urbaine. En outre, la diversification des modes de transport est généralement présentée comme un moyen de réduire la dépendance à l'automobile, et donc de tendre vers une mobilité plus durable. Cependant, les bénéfices de cette plus grande disponibilité de modes restent à clarifier. Chaque mode étant habituellement analysé de manière indépendante, son rôle et ses interactions avec les autres moyens de transport sont encore mal compris. En particulier, des questions se posent sur le sens de la relation (complémentaire ou concurrentielle) entre les nouveaux services de mobilité partagée et le transport en commun.

De même, peu d'outils ont été développés pour modéliser plusieurs modes simultanément. L'effet anticipé des stratégies de planification sur la répartition modale repose ainsi souvent sur des hypothèses simplificatrices. Ces lacunes sont notamment dues à des enjeux méthodologiques. En effet, plusieurs ensembles de données indépendants doivent être combinés pour pouvoir considérer tous les modes en même temps. De plus, des analyses longitudinales sont nécessaires afin de suivre la variabilité d'utilisation de chaque mode.

Dans cette perspective, cette thèse vise à valoriser de multiples sources de données afin de qualifier et quantifier les interactions entre différents modes de transport urbains. Pour ce faire, la ville de Montréal est étudiée dans six articles. L'utilisation de plusieurs modes, classiques (automobile privée, transport en commun, vélo, taxi) et plus récents (autopartage, vélopartage), est décrite à partir de deux grands types de données : les enquêtes ménages déplacements de la région de Montréal, dites enquêtes Origine-Destination, et différents flux passifs (transactions, comptages, etc.). Les avantages (et inconvénients) de ces deux grands types de données ont souvent été confrontés dans la littérature mais peu de travaux ont tenté de les coupler. Pourtant, les informations multimodales contenues dans l'enquête et la longitudinalité des données passives sont deux caractéristiques intéressantes à rassembler pour l'analyse de l'utilisation de différents modes.

Ainsi, cette recherche propose d'enrichir la journée moyenne d'automne de l'enquête Origine-Destination avec les fluctuations détectées dans des données passives. Dans un premier temps, l'enquête est repondérée et plusieurs bases de données passives sont nettoyées pour pouvoir être comparées sur plusieurs mois. Des indicateurs sont élaborés afin de mesurer des variations quotidiennes systémiques dans l'utilisation de différents modes quels que soient les volumes

transportés et leur unité de mesure. Les résultats montrent que les rythmes comportementaux observés dans les deux types de données sont semblables. Ces similarités sont alors exploitées dans un objectif de fusion de données. Une procédure automatisée basée sur la décomposition de séries temporelles est développée. La tendance et les saisonnalités extraites des données passives sont appliquées à l'enquête afin d'annualiser et de projeter les parts modales moyennes collectées tous les cinq ans. Par conséquent, la méthode proposée permet de suivre l'utilisation de plusieurs modes et d'analyser des interactions en termes de répartition de la demande de transport.

Par ailleurs, différentes sortes de relations entre les modes ont été examinées dans la littérature, et des conclusions parfois contradictoires ont été rapportées quant à la complémentarité ou compétition des modes. Cette recherche tente d'expliquer ces divergences en définissant différents types d'interactions en fonction de l'échelle spatio-temporelle (court terme versus long terme, global versus local), mais aussi en fonction de l'objet d'étude (système complet, personne ou déplacement). Ainsi, un cadre conceptuel original est proposé et une terminologie basée sur les interactions biologiques est adaptée au transport. Plusieurs des interactions définies sont ensuite mises en évidence dans le contexte de Montréal. Leur variabilité est également analysée. De plus, les effets de différents facteurs explicatifs sur ces interactions et les impacts de ces interactions sur l'utilisation des modes sont modélisés.

Au niveau individuel, la multimodalité des comportements de déplacements est mesurée à l'aide d'une variante de l'indice de Dalton. Cet indicateur est appliqué à des séries d'intensités calculées à partir de l'enquête Origine-Destination. De plus, l'influence des attributs socio-démographiques et de la disponibilité des services de mobilité sur les niveaux de multimodalité est évaluée via une procédure d'inférence causale combinant une méthode d'appariement par type de personnes et un modèle de doubles différences. Les résultats prouvent qu'une bonification de l'offre de transport est propice à l'utilisation d'une plus grande diversité de modes.

Au niveau systémique, des corrélations sont calculées entre l'utilisation de différents modes de transport. Des relations intermodales sont estimées dans le voisinage des stations de métro à l'aide d'un modèle multiniveau qui permet de tenir compte de la structure hiérarchique et de l'autocorrélation des données. L'effet de la localisation de la station est également exploré. Plusieurs modes sont potentiellement utilisés en complémentarité du métro pour réaliser le « premier kilomètre » d'un déplacement et atteindre une station, en particulier lorsque la station est

située en périphérie. En revanche, une compétition est plus probable au centre-ville. Les impacts de la pandémie de COVID-19 sur les interactions entre les modes sont également évalués par la détection de points de rupture et la modélisation de séries temporelles interrompues. Des analyses croisées entre plusieurs modes révèlent que des transferts modaux ont potentiellement été déclenchés par la crise sanitaire. Une diminution des corrélations d'usage est également soulignée, suggérant une baisse de complémentarité dans l'utilisation des modes.

À une échelle encore plus désagrégée, au niveau du déplacement, des situations de complémentarité et de compétition sont estimées entre des services de mobilité partagés et le transport en commun. Un algorithme est développé afin de classer des déplacements réalisés en vélopartage et en autopartage en fonction de leur relation potentielle avec le transport en commun. La typologie obtenue indique que les modes partagés peuvent compléter le transport en commun dans le temps et l'espace, en dehors des heures de service et de la zone de desserte, mais aussi en qualité, dans des corridors de déplacements où les services de transport en commun ne sont pas compétitifs (en termes de distance ou de temps de déplacement). Cependant, une partie non négligeable de déplacements auraient pu être faits en transport en commun dans des conditions convenables. L'effet négatif de cette compétition sur l'achalandage du transport en commun est modélisé à l'aide d'un modèle de doubles différences comparant les périodes avant et après la fermeture temporaire du service de vélopartage.

Finalement, cette thèse permet de mieux comprendre certains types d'interactions entre les modes de transport, ainsi que leurs causes et leurs conséquences. Les résultats permettent entre autres d'informer les opérateurs des différents services sur des opportunités de collaborations. De plus, des outils sont fournis pour quantifier les interactions (en niveau de multimodalité, de corrélation ou d'impact sur l'utilisation des modes), et des méthodes permettant d'évaluer l'effet de décisions stratégiques sont mises en application. Néanmoins, des analyses supplémentaires seront nécessaires afin d'approfondir ces premières investigations. En particulier, les comportements individuels devront être précisés, et les interactions à court et à long terme devront être mieux distinguées. Ainsi, cette recherche ouvre des perspectives nouvelles pour l'étude des villes intelligentes et multimodales.

ABSTRACT

Mobility as a Service (or MaaS) and multimodality/intermodality are increasingly common terms in the field of urban mobility. Moreover, diversification of transport modes is generally presented as a way to reduce car dependency, and thus to move towards more sustainable mobility. However, the benefits of this greater availability of modes remain to be clarified. As each mode is usually analyzed independently, its role and interactions with other modes of transport are still poorly understood. In particular, there are questions about the direction of the relationship (complementary or competing) between new shared mobility services and transit.

Similarly, few tools have been developed to model multiple modes simultaneously. The anticipated effect of planning strategies on modal split is thus often based on simplifying assumptions. These shortcomings are due in part to methodological issues. Indeed, several independent data sets must be merged to consider all modes at the same time. In addition, longitudinal analyses are needed to monitor the variability in the use of each mode.

Within this context, this thesis aims to value multiple data sources in order to qualify and quantify the interactions between different urban transportation modes. To this end, the city of Montreal is studied in six papers. The use of several modes, both traditional (private car, transit, bicycle, taxi) and more recent ones (carsharing, bikesharing), is described using two main types of data: household travel surveys of the Montreal region, called Origin-Destination surveys, and different passive data streams (transactions, counts, etc.). The advantages (and disadvantages) of these two main types of data have often been compared in the literature, but few studies have attempted to link them. However, the multimodal information contained in the survey and the longitudinal nature of the passive data are two interesting characteristics to bring together for the analysis of the use of different modes.

Therefore, this research proposes to enrich the average fall day of the Origin-Destination survey with fluctuations detected in passive data. First, the survey is reweighted, and several passive databases are cleaned to be compared over several months. Indicators are developed to measure systemic day-to-day variations in the use of different modes regardless of the volumes and the metric. The results show that the behavioral patterns observed in the two types of data are alike. These similarities are then leveraged for data fusion purposes. An automated procedure based on time series decomposition is developed. The trend and seasonality extracted from the passive data

are applied to the survey to annualize and project the average modal shares collected every five years. Therefore, the proposed method allows monitoring the use of multiple modes and analyzing interactions in terms of transport demand distribution.

Furthermore, different kinds of relationships between modes have been examined in the literature, and sometimes contradictory conclusions have been reported about the complementarity or competition of modes. This research tries to explain these discrepancies by defining different types of interactions according to the spatial-temporal scale (short term versus long term, global versus local), but also according to the object of study (complete system, person or trip). Thus, an original conceptual framework is proposed, and a terminology based on biological interactions is adapted to the transport context. Several of the defined interactions are then highlighted for Montreal. Their variability is also analyzed. In addition, the effects of different explanatory factors on these interactions and the impacts of these interactions on mode use are modeled.

At the individual level, multimodality in travel behavior is measured using a variant of the Dalton index. This indicator is applied to series of intensities calculated from the Origin-Destination survey. In addition, the influence of socio-demographic attributes and the availability of mobility services on the multimodality levels is assessed through a causal inference procedure combining a matching method (by type of persons) and a difference-in-differences model. The results show that an improvement in transport supply is conducive to the use of a greater diversity of modes.

At the systemic level, correlations are calculated between the use of different modes of transport. Intermodal relationships are estimated in the neighborhood of subway stations using a multilevel model that considers the hierarchical structure and autocorrelation of the data. The effect of station location is also explored. Several modes are potentially used to supplement the subway in the “first mile” of a trip and reach a station, especially when the station is located on the outskirts. However, competition is more likely in the city center. The impacts of the COVID-19 pandemic on the interactions between modes are also appraised through the detection of breakpoints and the modeling of interrupted time series. Cross-modal analyses reveal that modal shifts were potentially triggered by the health crisis. A decline in usage correlations is also noted, suggesting a decrease in complementarity in mode use.

At an even more disaggregated scale, at the trip level, situations of complementarity and competition are estimated between shared mobility services and transit. An algorithm is developed

to classify bikesharing and carsharing trips according to their potential relationship with transit. The resulting typology indicates that shared modes can complement transit in time and space, outside of service hours and service area, but also in quality, in travel corridors where transit services are not competitive (in distance or travel time). However, a significant portion of trips could have been made by transit under suitable conditions. The negative effect of this competition on transit ridership is modelled using a difference-in-differences model by comparing the periods before and after the temporary closure of the bikeshare service.

Finally, this thesis allows a better understanding of certain types of interactions between transport modes, as well as their causes and consequences. The outcome can be used to inform operators of different services about opportunities for collaboration. In addition, tools are provided to quantify the interactions (in terms of multimodality, correlation, or impact on mode use), and methods to evaluate the effect of policy and planning decisions are implemented. Nevertheless, further analysis will be required to deepen these first investigations. In particular, individual behaviors need to be clarified, and short- and long-term interactions need to be better distinguished. Therefore, this research opens new perspectives for the study of smart and multimodal cities.

TABLE DES MATIÈRES

DÉDICACE.....	III
REMERCIEMENTS	IV
RÉSUMÉ.....	V
ABSTRACT	VIII
TABLE DES MATIÈRES	XI
LISTE DES TABLEAUX.....	XVIII
LISTE DES FIGURES.....	XXI
LISTE DES SIGLES ET ABRÉVIATIONS	XXV
LISTE DES ANNEXES.....	XXVIII
CHAPITRE 1 INTRODUCTION.....	1
1.1 Mise en contexte.....	1
1.1.1 Évolution vers une mobilité multimodale : concepts associés.....	1
1.1.2 Les bénéfices potentiels de la transition vers une mobilité multimodale.....	3
1.1.3 Le cas de Montréal	4
1.2 Problématique.....	6
1.2.1 Des interactions à préciser.....	6
1.2.2 Enjeux analytiques et méthodologiques.....	8
1.3 Objectifs	9
1.4 Structure du document	11
CHAPITRE 2 REVUE DE LITTÉRATURE	12
2.1 Techniques de fusion de données.....	12
2.1.1 Définition générale de la fusion de données	12
2.1.2 Méthodes et défis généraux de la fusion de données	14

2.1.3	Application au domaine du transport	17
2.1.4	Synthèse spécifique	26
2.2	Analyse des interactions entre les modes de transport.....	27
2.2.1	Évaluation de l'impact des modes émergents sur les modes classiques	27
2.2.2	Étude de la multimodalité	47
2.2.3	Synthèse spécifique	59
2.3	Modélisation de la demande de transport.....	60
2.3.1	Modélisation classique	60
2.3.2	Nouvelles approches de modélisation	68
2.3.3	Facteurs pouvant avoir un effet sur la répartition modale.....	78
2.3.4	Synthèse spécifique	87
CHAPITRE 3	DONNÉES ET MÉTHODOLOGIE GÉNÉRALE	88
3.1	Cadre conceptuel	88
3.1.1	Différents niveaux d'analyse.....	88
3.1.2	Différentes situations de complémentarité/compétition.....	89
3.2	Données utilisées	92
3.2.1	Enquête Origine-Destination de Montréal	92
3.2.2	Données passives.....	93
3.2.3	Comparaison des deux principales sources de données	95
3.2.4	Autres données	96
3.3	Démarche méthodologique	97
3.3.1	Article 1 : Comparaison entre enquête OD et données passives.....	99
3.3.2	Article 2 : Fusion entre enquête OD et données passives	100
3.3.3	Article 3 : Multimodalité des comportements de mobilité.....	101

3.3.4	Article 4 : Interactions systémiques intermodales.....	102
3.3.5	Article 5 : Impacts de la pandémie de COVID-19	102
3.3.6	Article 6 : Complémentarité et compétition au niveau du déplacement	103
CHAPITRE 4 ARTICLE 1 : CROSS-ANALYSIS OF THE VARIABILITY OF TRAVEL BEHAVIORS USING ONE-DAY TRIP DIARIES AND LONGITUDINAL DATA		105
	Abstract	105
4.1	Introduction	106
4.2	Background	107
4.2.1	Variability of mobility behaviors in survey data.....	107
4.2.2	An emerging data landscape: opportunities and challenges.....	109
4.2.3	Data fusion between traditional and emerging transportation data.....	111
4.3	Data and methodology	113
4.3.1	Data description.....	113
4.3.2	Methodological steps.....	116
4.4	Results	123
4.4.1	Expansion of the Montreal origin-destination survey	123
4.4.2	Description of the use of three modes from the two data sources.....	125
4.4.3	Comparison indicators of the two data sources per mode.....	128
4.5	Conclusion.....	134
	Acknowledgments	137
CHAPITRE 5 ARTICLE 2 : COMBINING A REGIONAL HOUSEHOLD SURVEY AND PASSIVE DATA STREAMS FOR LONGITUDINAL MONITORING PURPOSES.....		138
	Abstract	138
5.1	Introduction	139
5.2	Background	140

5.3	Data and methodology	142
5.3.1	Data description and processing.....	142
5.3.2	Method overview.....	144
5.3.3	Seasonal and Trend decomposition using Loess.....	145
5.3.4	Projection and annualization	148
5.4	Results	148
5.4.1	Passive data decompositions	148
5.4.2	Projected and annualized modal shares.....	153
5.5	Conclusion.....	156
	Acknowledgements	158
CHAPITRE 6 ARTICLE 3: MEASURING CHANGES IN MULTIMODAL TRAVEL BEHAVIOR RESULTING FROM TRANSPORT SUPPLY IMPROVEMENT		159
	Abstract	159
6.1	Introduction	160
6.2	Background	160
6.2.1	Analyzing multimodal mobility	161
6.2.2	Quantifying multimodality level	162
6.2.3	Factors affecting multimodal travel behavior	163
6.3	Data and methodology	163
6.3.1	Building a typology of people	165
6.3.2	Defining an indicator of multimodality.....	167
6.3.3	Difference-in-differences approach for causal impact modeling.....	169
6.4	Results	171
6.4.1	Typology of people	171
6.4.2	Multimodality by type of people	173

6.4.3	Multimodality evolution from 2013 to 2018: influence of transport supply.....	178
6.5	Conclusion.....	180
	Acknowledgments.....	182
CHAPITRE 7 ARTICLE 4 : MODELING INTERACTIONS BETWEEN THE MONTREAL SUBWAY AND OTHER URBAN TRANSPORTATION MODES AT THE STATION LEVEL 183		
	Abstract	183
7.1	Introduction	185
7.2	Background	187
7.2.1	Multimodal analysis in transportation research	187
7.2.2	Longitudinal modeling of transport demand	190
7.3	Data and methodology	191
7.3.1	Multiple-source data fusion.....	192
7.3.2	Growth modeling.....	193
7.4	Results	204
7.4.1	Growth model of the daily subway use	204
7.4.2	Growth model by time period and station group.....	208
7.5	Conclusion.....	213
	Acknowledgments.....	216
CHAPITRE 8 ARTICLE 5 : ASSESSING THE IMPACTS OF THE COVID-19 PANDEMIC ON SUBWAY RIDERSHIP AND ON THE INTERACTIONS WITH OTHER TRANSPORTATION MODES 217		
	Abstract	217
8.1	Introduction	218
8.2	Background	219

8.2.1	Literature review	219
8.2.2	COVID-19 timeline in Montreal	222
8.3	Data and methodology	223
8.3.1	Data description.....	225
8.3.2	Analysis strategy	230
8.4	Results	235
8.4.1	Breakpoints in the subway ridership and magnitude of the impacts.....	235
8.4.2	Variations of the impacts by phase	238
8.4.3	Interactions with other transportation modes	242
8.5	Discussion	246
8.6	Conclusion.....	249
	Acknowledgments.....	250
CHAPITRE 9 ARTICLE 6 : ASSESSING COMPLEMENTARY AND COMPETING INTERACTIONS BETWEEN TRANSIT AND SHARED TRANSPORTATION MODES ...		252
	Abstract	252
9.1	Introduction	254
9.2	Background	254
9.3	Data and methodology	258
9.3.1	Case study and datasets	258
9.3.2	Conceptual framework	261
9.3.3	Methodological elements	263
9.4	Results	272
9.4.1	General results of the trip classification.....	272
9.4.2	Temporal analysis	273
9.4.3	Spatial analysis.....	276

9.4.4	Modeling results.....	280
9.5	Conclusion.....	281
	Acknowledgments.....	284
CHAPITRE 10	DISCUSSION GÉNÉRALE	285
10.1	Synthèse des résultats.....	285
10.1.1	Article 1 : Comparaison entre enquête OD et données passives.....	285
10.1.2	Article 2 : Fusion entre enquête OD et données passives	286
10.1.3	Article 3 : Multimodalité des comportements de mobilité.....	286
10.1.4	Article 4 : Interactions systémiques intermodales.....	287
10.1.5	Article 5 : Impacts de la pandémie de COVID-19	288
10.1.6	Article 6 : Complémentarité et compétition au niveau du déplacement	288
10.2	Discussion générale des résultats	289
10.2.1	Intégration de sources diversifiées de données	289
10.2.2	Qualification et quantification de différents types d'interactions entre les modes 291	
10.2.3	Modélisation des interactions entre les modes	294
CHAPITRE 11	CONCLUSION ET RECOMMANDATIONS	296
11.1	Contributions.....	296
11.2	Limites.....	299
11.3	Perspectives	302
	RÉFÉRENCES.....	306
	ANNEXES	338

LISTE DES TABLEAUX

Tableau 2.1 Synthèse des principales classifications de méthodes de fusion de données	15
Tableau 2.2 Indicateurs de multimodalité définis par Diana et Pirra (2016)	58
Tableau 2.3 Classification des modèles d'affectation d'après Bonnel (2004)	66
Tableau 3.1 Interactions biologiques appliquées à deux modes de transport A et B en fonction de l'impact sur leur utilisation respective : positif (+), négatif (-) ou neutre (0)	90
Tableau 3.2 Description sommaire des données passives utilisées pour chaque mode	94
Tableau 3.3 Comparaison de l'enquête Origine-Destination et des données passives	96
Table 4.1 Descriptive statistics on the calibration factors for each Montreal OD survey	124
Table 4.2 Pairwise comparisons of the linear trends in the bicycle databases (p-values of the Tukey tests)	129
Table 4.3 Average ratios (passive data / OD surveys) and coefficient of variations, per mode and day type	131
Table 4.4 Periodicity analysis with spectral density: detected frequencies and periods for the subway from two data sources	133
Table 5.1 Selected values of the smoothing parameters of the seasonal components for each mode and decomposition form	149
Table 5.2 Strength of the trend and seasonal components (F indicator) in the multiplicative data decomposition of each mode	151
Table 5.3 Error rates between the projected and observed number of trips on the 2018 fall weekdays	153
Table 5.4 Variations across the 2018 fall of the daily modal shares (on weekdays only)	155
Table 6.1 Typology of households – Centers: average variables by type	174
Table 6.2 Results of difference-in-differences regression with treatment based on the evolution of transport supply between 2013 and 2018	179
Table 7.1 Description of the tested variables (selected variables in gray)	199

Table 7.2 Definition of the taxonomy of growth models (from A to G).....	204
Table 7.3 Results of the taxonomy of growth models to forecast the daily use of subway	206
Table 7.4 Forecast for the first 15 days of November - comparison of model G with other models	207
Table 7.5 Model results to forecast subway use by station group and time period	211
Table 7.6 Average number of boardings per 5-hour period per station for a working Monday at time 0 (controlling for other variables), according to model 1	212
Table 8.1 COVID-19 Timeline in Montreal.....	224
Table 8.2 Comparison of the average daily use of each mode in 2020 and 2021 vs. 2019	228
Table 8.3 Test results for different number of breakpoints k.....	236
Table 8.4 Magnitude of the impacts of COVID-19-related events based on estimated breakpoint dates.....	237
Table 8.5 Magnitude of the impacts (differences in average daily ridership between pairs of phases) by type of day and period of the day	240
Table 8.6 Interrupted time series models by transportation mode	243
Table 9.1 Descriptive statistics.....	261
Table 9.2 Biological interactions applied to two modes A and B based on the impacts on their ridership: positive (+), negative (-) or neutral (0)	262
Table 9.3 Distribution of the bikesharing and free-floating carsharing trips by interaction class.....	273
Table 9.4 Results of the FE-DID model for four treatment definitions	281
Tableau 10.1 Synthèse des interactions analysées dans chaque article.....	291
Tableau 11.1 Synthèse des contributions de la thèse	297
Tableau 11.2 Synthèse des limites de la thèse	299
Tableau 11.3 Synthèse des perspectives de la thèse.....	304
Tableau A.1 Synthèse des exemples d'application de techniques de fusion de données dans un contexte de transport	338

Tableau B.1 Synthèse de l'évaluation de l'impact des modes émergents sur les modes classiques	
.....	342

LISTE DES FIGURES

Figure 1.1 Évolution mensuelle de l'intérêt pour la recherche du mot « Mobility as a Service » (données Google Trends, de janvier 2014 à décembre 2021, pour tous les pays du monde) ..2	
Figure 1.2 Évolution de l'achalandage annuel (métro et bus) de la STM de 1995 à 2019 7	
Figure 2.1 Schématisation du modèle à 4 étapes, inspirée de McNally (2007) 61	
Figure 3.1 Étude du cas de Montréal : territoire de l'enquête Origine-Destination et des sources de données passives 93	
Figure 3.2 Démarche méthodologie générale de la thèse et positionnement des 6 articles 98	
Figure 4.1 Study area (metropolitan area of Montreal) with the subway network, cycling counters, and highway and bridge traffic sections.....115	
Figure 4.2 Method overview 117	
Figure 4.3 Expansion method applied to the Montreal Origin-Destination (OD) survey 118	
Figure 4.4 Selected cycling and traffic counters in a) 2013 and b) 2018 passive datasets 120	
Figure 4.5 Distribution of the calibration factors over the fall period (2018 Montreal OD survey) 124	
Figure 4.6 Total number of trips per day (all modes of transport) during the entire survey period 125	
Figure 4.7 Proportion of use of three modes (car, subway, and bicycle) over the fall period from 2013 OD survey and passive data 126	
Figure 4.8 Proportion of use of three modes (car, subway, and bicycle) over the fall period from 2018 OD survey and passive data 126	
Figure 4.9 Difference in the average use of three modes (car, subway, and bicycle) by day type compared to the average working day from all databases (2018, 2013, 2008 OD surveys and 2018, 2013 passive data) 127	
Figure 4.10 Linear trends in the bicycle databases 129	

Figure 4.11 Ratios per mode between passive databases and Montreal Origin-Destination (OD) surveys.....	130
Figure 4.12 Periodogram of the hourly number of boardings in the subway (2008 OD survey)	133
Figure 5.1 Data fusion procedure between the Montreal Origin-Destination (OD) survey and passive data	144
Figure 5.2 (a) Additive and (b) multiplicative STL decompositions of the subway tap-in validations	150
Figure 5.3 Multiplicative STL decomposition of the (a) cycling and (b) car counts.....	152
Figure 5.4 (a) Projected and (b) annualized curves of the total number of trips per day in the subway	154
Figure 5.5 Daily modal shares of three modes projected from 2013 to 2018	155
Figure 5.6 (a) Annualized daily modal shares of three modes in 2018 and (b) average modal shares on a 2018 fall week	156
Figure 6.1 Methodological framework.....	165
Figure 6.2 Typology of neighborhoods (proportion of households).....	172
Figure 6.3 Typology of neighborhoods – Centers: normalized average variables by type.....	172
Figure 6.4 Typology of households (by maximum age range and average age)	174
Figure 6.5 Histograms of the multimodality indicator calculated by type of people for different measurement units of the intensities of use (trip segments, trips, travel distance or time) and different mode aggregations (AG1, AG2 or AG3)	175
Figure 6.6 Histograms of the multimodality indicator calculated by type of people for different values of the parameter ε – Example for DAL_m (time) AG1.....	175
Figure 6.7 Regression trees for predicting the multimodality indicator DAL_m (time) AG1 $\varepsilon = 0.6$ calculated by type of people (a) in 2013, (b) in 2018	177
Figure 7.1 Island of Montreal with transit, bikesharing, taxi and free-floating carsharing services	192
Figure 7.2 Hierarchical structure of the growth model developed in this paper	194

Figure 7.3 Trajectories (evolution of the daily number of boardings over time) for 9 randomly selected subway stations.....	196
Figure 7.4 Pearson correlation coefficients between all pairs of variables.....	200
Figure 7.5 Growth model including day type effect (model C): Variations of the station-specific trajectories around the modeled average trajectory.....	203
Figure 7.6 Decomposition of space (a) and time (b).....	210
Figure 8.1 Montreal subway network, selected cycling and car counters, free-floating carsharing zone and carsharing stations.....	226
Figure 8.2 Daily time series for subway, bicycle, car, carsharing and COVID-19 hospitalizations	227
Figure 8.3 Structural change point detection results and associated Bayesian posterior probabilities	236
Figure 8.4 Wavelet power spectrum of the hourly subway time series	241
Figure 8.5 Spectral density function per phase	242
Figure 9.1 Case study city (Montreal) and the transit, bikesharing and free-floating carsharing systems	259
Figure 9.2 Rule-based classification algorithm to assess potential complementary and competing interactions with transit	265
Figure 9.3 Average daily transit ridership for the lines in control and treatment groups (based on treatment #4), with BIXI open and closed	271
Figure 9.4 Trip duration distribution by interaction category of the a) bikesharing and b) carsharing trips.....	274
Figure 9.5 Distribution of all trips by day and distribution of all week trips by hour for bikesharing (a and b) and free-floating carsharing (c and d), with average trip duration by day (e) and hour (f).....	275
Figure 9.6 Spatial representation in a hexagonal grid (100-m diameter hexagons) of the density of bikesharing Origin-Destination desire lines belonging to different interaction classes: a) no	

impact, b) complementarity in space, c) complementarity in service quality, d) competition	
.....	277

Figure 9.7 Spatial representation in a hexagonal grid (100-m diameter hexagons) of the density of free-floating carsharing Origin-Destination desire lines belonging to different interaction classes: a) complementarity in space, b) complementarity in service quality, c) competition	
.....	278

Figure 9.8 Number of a) bikesharing and b) free-floating carsharing trips competing with transit per line based on the transit alternatives calculated by Transition.....	279

LISTE DES SIGLES ET ABRÉVIATIONS

AR	Autoregressive
ARMA	Autoregressive Moving Average
ARIMA	Autoregressive Integrated Moving Average
ARTM	Autorité régionale de transport métropolitain
ATC	Average treatment effect on the controls
ATE	Average treatment effect
ATT	Average treatment effect for the treated
AVL	Automatic vehicle location
CDR	Call detail records
CEO	Chief executive officer
CIT	Conseils intermunicipaux de transport
CMM	Communauté métropolitaine de Montréal
COVID-19	Maladie à coronavirus 2019
DID	Difference-in-differences
DMTI	Desktop Mapping Technologies Inc
ETS	Erreur, Tendances, Saisonnalité
FE	Fixed-effects
GIS	Geographic information system
GPS	Global Positioning System
GTFS	Global Transit Feed Specification
ILUTE	Integrated Land Use, Transportation, Environment
INSPQ	Institut national de santé publique du Québec
LSTM	Long-Short Term Memory

MA	Moving Average
MaaS	Mobility as a Service
MiD	Mobilität in Deutschland (Mobility in Germany)
MOP	Mobilitätspanel (German Mobility Panel)
MOTREM	Modèle de transport de la région de Montréal
MRC	Municipalités régionales de comté
MTQ	Ministère des transports du Québec
NHTS	National Household Travel Survey
OD	Origine-Destination
OMIT	Organisme municipal et intermunicipal de transport
OPTC	Organismes publics de transport en commun
PJCCI	Ponts Jacques Cartier et Champlain Incorporée
RF	Random Forests
RTL	Réseau de transport de Longueuil
SAAQ	Société de l'assurance automobile du Québec
SES	Simple Exponential Smoothing
SEATS	Seasonal Extraction in ARIMA Time Series
STI	Systèmes de transport intelligents
STL	Société de transport de Laval
STL	Seasonal and Trend decomposition using Loess
STM	Société de transport de Montréal
SUITS	Sustainable Urban Integrated Transport Systems
TASHA	Travel Activity Scheduling Model for Household Agents
TC	Transport en commun

TNC	Transportation network companies
VAR	Vector autoregression
VLS	Véhicules en libre-service

LISTE DES ANNEXES

Annexe A Synthèse des exemples d'application de techniques de fusion de données dans un contexte de transport	338
Annexe B Synthèse de l'évaluation de l'impact des modes émergents sur les modes classiques	342

CHAPITRE 1 INTRODUCTION

1.1 Mise en contexte

Au cours des dernières années, le rapide développement des technologies et l'évolution des systèmes d'information et de communication ont permis une diversification des modes de transport urbains. En particulier, des modes partagés tels que le covoiturage, l'autopartage ou le vélopartage, ainsi que des nouvelles formes de services de transport collectif et/ou à la demande (taxi collectif, microtransit, Uber, Lyft, etc.), sont venus compléter l'offre classique (automobile et transport en commun). Les modes actifs (vélo et marche) ont également vu leur popularité augmenter grâce au récent déploiement d'infrastructures et de programmes visant à encourager l'utilisation de ces moyens de transport verts. De plus, de nouveaux acteurs pour la mobilité active sont apparus avec l'introduction de trottinettes et vélos électriques dans les villes. Cette diversification de l'offre de transport semble peu à peu conduire à une modification des comportements de déplacement dans plusieurs pays du monde : l'utilisation d'une plus grande variété de modes de transport différents devient de plus en plus répandue (Buehler & Hamre, 2014), en particulier chez les nouvelles générations (Kuhnimhof et al., 2012).

1.1.1 Évolution vers une mobilité multimodale : concepts associés

Ces tendances ont été accompagnées par l'émergence de nouveaux concepts de mobilité. Dès les années 90, Michel Trudel invente le terme **cocktail modal** dans le cadre d'une réflexion menée au ministère des Transports du Québec (MTQ) afin d'encourager l'intégration du taxi. Repris sous le nom de **cocktail transport** par Vélo Québec et Équiterre, ce concept a été défini par Michel Labrecque comme « l'offre combinée et unifiée de moyens de transport, individuels et collectifs, permettant aux citoyens de se déplacer efficacement et rapidement, tout en économisant et en réduisant au minimum les impacts négatifs sur l'environnement ». Ainsi, le cocktail transport nécessite la combinaison de tous les modes de transport, incluant la voiture, pour contrer « l'usage individuel et quasi exclusif de l'automobile privée pour tous les déplacements ». Toutefois, ce changement de paradigme implique une « unification des services », « une combinaison, une intégration et une bonification tarifaire », ainsi qu'un « maillage des infrastructures et des services » pour faciliter les transferts d'un mode à un autre (Labrecque, 1997).

Plus récemment, le concept de **mobilité intégrée** (ou **Mobility as a service** en anglais, abrégé par **MaaS**) s'est rapidement répandu à l'échelle internationale, en particulier grâce aux différentes interventions de Sampo Hietanen en 2014 (Hietanen, 2014). CEO de ITS Finland à l'époque, Hietanen est généralement considéré comme un des pères fondateurs du MaaS. Les données de Google Trends, utilisées pour produire la Figure 1.1 ci-dessous, confirment le rôle qu'il a joué dans la diffusion de ce concept à travers le monde. Cette figure montre l'évolution croissante de la fréquence à laquelle le mot « Mobility as a service » a été tapé dans le moteur de recherche Google jusqu'en 2019. Une baisse est néanmoins observée après le début de la pandémie de COVID-19. La revue littéraire de Jittrapirom et al. (2017) définit et liste les principales caractéristiques d'un service de mobilité intégrée. Cette notion implique l'intégration de tous les modes de transport dans un seul et même service de mobilité basé sur l'accessibilité (et non plus sur la possession) et répondant à différents types de besoins. De plus, ce service est géré à partir d'une unique plateforme numérique et multifonctions (incluant une fonction de paiement à l'utilisation), centré sur l'utilisateur et personnalisable. Utriainen et Pöllänen (2018) ont quant à eux regroupé les principaux travaux scientifiques portant sur le concept, et Kamargianni et al. (2016) ont recensé différents systèmes MaaS dans le monde en fonction de leur niveau d'intégration.

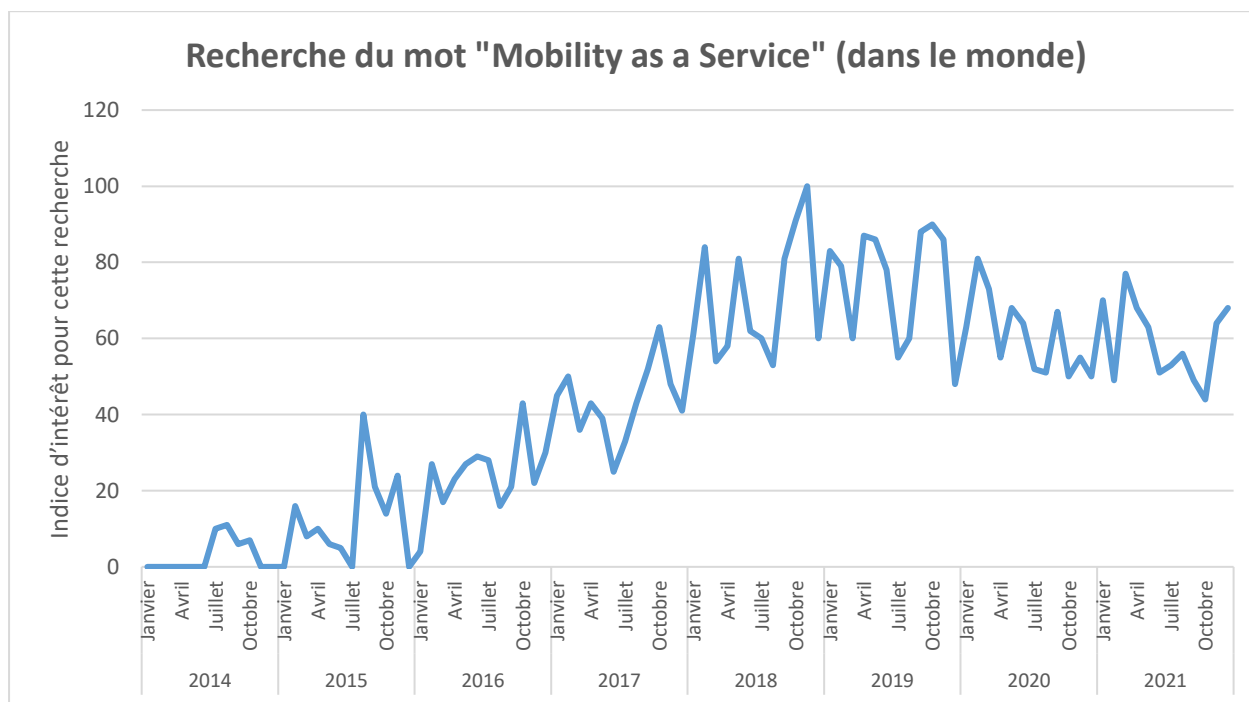


Figure 1.1 Évolution mensuelle de l'intérêt pour la recherche du mot « Mobility as a Service » (données Google Trends, de janvier 2014 à décembre 2021, pour tous les pays du monde)

Dans la même lignée, les termes **multimodalité** et **intermodalité** sont de plus en plus à la mode dans le domaine du transport. Le concept de multimodalité peut s'appliquer à un réseau de transport ou à une personne. Dans le premier cas, le terme multimodal est employé pour désigner la disponibilité de divers modes de transport dans un lieu. Dans le deuxième cas, il implique l'utilisation (par une personne ou un ensemble de personnes) de différents modes de transport pour se déplacer pendant une période donnée (Spickermann et al., 2014). Ce terme est à distinguer de la notion d'intermodalité. Après avoir recensé, comparé et discuté plusieurs définitions du terme « intermodal » dans la littérature, Jones et al. (2000) ont proposé une des définitions les plus connues de l'intermodalité, à savoir: « l'expédition de marchandises et le mouvement de personnes impliquant plus d'un mode de transport au cours d'un seul et même trajet sans interruption » [notre traduction]. Autrement dit, on utilise le terme intermodal pour désigner la combinaison de plusieurs modes de transport dans un même déplacement. L'intermodalité découle intrinsèquement de la multimodalité puisque plusieurs modes de transport doivent être disponibles pour pouvoir être combinés. Par conséquent, on parle parfois de multimodalité des déplacements, par opposition à la multimodalité des personnes (Diana & Pirra, 2016). Dans la suite de ce document, les expressions « déplacements multimodaux » et « déplacements intermodaux » seront ainsi utilisées comme des synonymes. Par ailleurs, le terme **comodalité**, défini par la Commission européenne comme « le recours efficace à différents modes de transport isolément ou en combinaison », regroupe les deux notions (Commission européenne, 2006). Par opposition, la **monomodalité** ou l'**unimodalité** suppose l'utilisation d'un seul mode de transport pour tous les déplacements (Nobis, 2007).

1.1.2 Les bénéfices potentiels de la transition vers une mobilité multimodale

D'après une étude basée sur la théorie de la transition sociotechnique avec application de la méthode Delphi (méthode utilisée en gestion des questions stratégiques pour établir un consensus auprès d'experts), les différentes parties prenantes du transport (autorités publiques, industriels, opérateurs, chercheurs et usagers) s'accordent pour dire que la multimodalité est une caractéristique probable et souhaitable de la mobilité du futur (Spickermann et al., 2014). L'approche interdisciplinaire et multiniveaux (impliquant des acteurs et institutions à différentes échelles) utilisée par les auteurs met en évidence la nécessité de guider les villes dans l'élaboration d'une vision à long terme et d'une stratégie de transition qui s'oriente vers un système de transport

multimodal. Ce consensus sur le fait que « la mobilité de demain sera multimodale ou ne sera pas » est également mis en avant dans la presse (Ferraris, 2018).

Cet engouement pour la multimodalité s'inscrit dans le cadre d'une mobilité qui se veut plus durable. En effet, les concepts exposés précédemment ont été développés en réponse à une motorisation croissante, en nombre d'automobiles par habitant ou en kilométrage parcouru (Labrecque, 1997). La diversification et l'intégration des modes de transport sont présentées comme un moyen de réduire l'utilisation de l'automobile, et donc ses externalités négatives sur l'environnement, la société et l'économie. La multiplication des options de déplacement permet en effet de proposer des alternatives à l'auto solo (automobile privée avec un seul occupant). D'autres avantages peuvent également être tirés d'une mobilité multimodale et intermodale (Vivre en Ville, 2018). D'une part, une plus grande complémentarité des modes de transport améliore l'efficacité et réduit la durée des déplacements en faisant diminuer les temps de correspondance d'un mode à un autre. D'autre part, la création de nouvelles possibilités de trajets (par l'utilisation de modes différents) génère un gain en accessibilité aussi bien au niveau temporel que spatial, les individus pouvant alors atteindre davantage de destinations avec un choix d'horaires plus varié.

Ces trois bénéfices potentiels (proposition d'alternatives à l'auto solo, amélioration de l'efficacité des déplacements et gain en accessibilité) font parfaitement écho aux cibles fixées par la politique de mobilité durable du MTQ pour 2030. En effet, le ministère souhaite entre autres réduire de 20% la part des déplacements effectués en auto solo (en transférant ces déplacements vers des modes de transport plus durables), diminuer de 20% le temps de déplacement moyen entre le domicile et le travail, et augmenter à 70% la part de la population québécoise ayant accès à au moins quatre services de mobilité durable (Ministère des Transports du Québec, 2018). Une mobilité plus multimodale pourrait donc aider le Québec à atteindre ses objectifs en matière de durabilité.

1.1.3 Le cas de Montréal

Montréal est une ville dans laquelle une panoplie de modes de transport sont présents. Outre les modes individuels classiques (automobile privée, vélo, marche), des services de transport en commun (métro, bus, train) sont fournis par deux organismes publics de transport en commun (OPTC) - la Société de transport de Montréal (STM) pour les réseaux de métro et bus, et Exo pour les trains de banlieue - chapeautés par l'Autorité régionale de transport métropolitain (ARTM). Diverses compagnies encadrées par le Bureau du taxi de Montréal offrent également des services

de transport par taxi. De plus, la ville est aussi desservie par Uber depuis 2013 et plusieurs plateformes (Amigo Express, Eva, etc.) permettent le covoiturage. L'entreprise Communauto, arrivée à Montréal en 1995, propose quant à elle deux types de services d'autopartage (le service basé stations avec réservation et retour à la station initiale, et le service FLEX avec véhicules en libre-service), tandis que BIXI exploite un système de vélopartage avec ancrage depuis 2009. À l'été 2019, un projet pilote permettant l'arrivée de trottinettes et vélos en libre-service sans ancrage a également été lancée (Ville de Montréal, 2019b). Non poursuivie les années suivantes, la tentative pourrait toutefois être renouvelée en 2023 (Bergeron, 2022). Ainsi, la ville de Montréal démontre un fort potentiel en termes de multimodalité.

Par ailleurs, des efforts ont été fournis pour favoriser une mobilité intégrée. La STM a établi un partenariat avec l'industrie du taxi afin de proposer des services de taxi collectif (dans des secteurs à faible densité) et de transport adapté (pour les personnes à mobilité réduite). Les vélos sont également acceptés à bord des métros et des trains sous certaines conditions, et la carte à puce OPUS permet le déverrouillage des véhicules FLEX de Communauto. En outre, Communauto et BIXI offrent un rabais tarifaire aux abonnés des deux services. De telles ententes tarifaires existaient également avec la STM, mais elles ont été supprimées après le début de la pandémie de COVID-19 (à l'exception du Passeport Mobilité, conçu par Voyagez Futé à l'intention des promoteurs immobiliers des quartiers centraux de Montréal, qui comprend un an d'abonnement aux services de la STM, de Communauto et de BIXI). Les deux applications mobiles montréalaises Transit et Chrono tendent également à intégrer de plus en plus de modes de transport dans leur calculateur de trajet. En particulier, Transit indique en temps réel l'occupation des stations Communauto et BIXI, et propose différentes alternatives en transport en commun, vélo, BIXI ou transport à la demande, ainsi que des combinaisons de plusieurs de ces modes.

Ces initiatives seront poursuivies dans le cadre du Défi des villes intelligentes du Canada. La Ville de Montréal a en effet remporté le premier prix du concours, soit un montant de 50 M\$ pour pouvoir réaliser les projets proposés. Ceux-ci incluent un volet mobilité intégrée qui prévoit notamment une « nouvelle plateforme numérique ouverte » et « personnalisable » incluant de nombreux modes de transport, un « outil de planification de trajet intermodal » et une « approche de tarification simplifiée » (Ville de Montréal, 2019a). D'abord discuté en interne à la STM avec le projet Céleste (Schnobb, 2018), c'est aujourd'hui l'ARTM qui travaille sur le sujet (ARTM, 2022).

1.2 Problématique

Malgré une diversification de l'offre de transport et une volonté croissante de tendre vers une mobilité plus durable, les politiques et interventions mises en œuvre à Montréal n'ont pas encore eu les effets escomptés. Des recherches supplémentaires sont donc nécessaires afin de mieux comprendre les interactions entre les modes. Cependant, ces analyses doivent faire face à différents enjeux analytiques et méthodologiques.

1.2.1 Des interactions à préciser

En dépit des nouvelles alternatives proposées depuis déjà plusieurs années, les Montréalais ne sont pas moins dépendants de leur automobile (Laviolette et al., 2020). Les chiffres de la SAAQ révèlent que la motorisation continue de croître plus vite que la population (augmentation du taux de motorisation de 7% entre 2006 et 2016 sur l'Île de Montréal). De même, les données du recensement canadien et des enquêtes Origine-Destination de Montréal montrent que le nombre de déplacements en automobile conducteur augmente toujours et que la part modale correspondante est stable. En contrepartie, la proportion des déplacements effectués avec des modes de transport durables n'a que très faiblement progressé. La mobilité des Montréalais demeure donc majoritairement unimodale et centrée sur l'automobile.

En outre, dans le contexte où plusieurs villes américaines ont vu l'achalandage du transport en commun diminuer au cours des dernières années (même avant la pandémie de COVID-19), des questions se posent quant à l'impact des modes émergents sur les modes classiques (Erhardt et al., 2022; Feigon & Murphy, 2018; Graehler Jr et al., 2019): ces modes sont-ils complémentaires au transport au commun ou tendent-ils à le remplacer ? Contrairement aux villes étudiées par les auteurs cités, l'arrivée de nouveaux modes dans l'agglomération de Montréal n'a pas empêché l'achalandage du transport en commun d'augmenter. La Figure 1.2 ci-dessous a été tracée à partir des informations contenues dans les rapports d'entreprise publiés annuellement par la STM (Société de transport de Montréal, 2002). L'achalandage considéré est celui qui est calculé à partir des ventes tarifaires (appelé achalandage « conventionnel » dans les rapports) car, bien qu'il soit plus précis, l'achalandage « électronique » (estimé à partir des données de cartes à puce) n'est rapporté que depuis 2018 par la STM. Cette figure montre une tendance globale croissante du nombre de déplacements réalisés en métro et en bus.

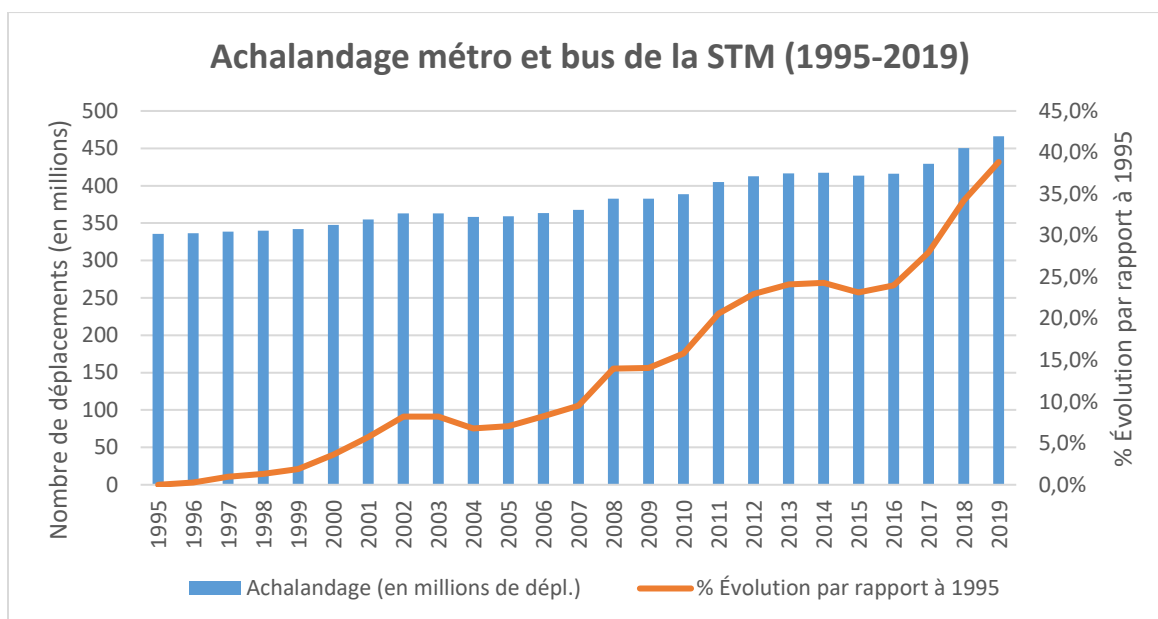


Figure 1.2 Évolution de l’achalandage annuel (métro et bus) de la STM de 1995 à 2019

Toutefois, cette croissance a parfois été ralentie, et l’achalandage a même diminué à deux reprises durant la période 1995-2019 (en 2004 et en 2015). De plus, la pandémie de COVID-19 a ensuite eu un impact négatif important sur l’utilisation du transport en commun, entraînant une chute de plus de 50% de l’achalandage annuel en 2020 et 2021 (par rapport à 2019). Cette baisse peut être en partie expliquée par le développement du télétravail, mais un transfert modal vers d’autres modes (modes actifs, modes individuels partagés ou automobile privée) pourrait également avoir eu lieu.

Ainsi, les effets positifs de la diversification de l’offre de transport sur l’évolution de la mobilité montréalaise sont encore peu perceptibles au niveau systémique. Des travaux plus poussés sont nécessaires afin de valider que les modes émergents sont complémentaires au (et non en concurrence avec le) transport en commun et que les changements de comportement induits sont désirés (c’est-à-dire que les automobilistes se tournent vers des modes de transport plus durables). Pour ce faire, les relations entre les modes doivent être étudiés à différents niveaux (local et global) et différents horizons temporels : plusieurs types d’interactions entre les modes doivent être examinés à court terme, au sein de la mobilité quotidienne (situation quasi-statique), et leurs conséquences sur les comportements à plus long terme doivent également être explorées (situation dynamique). De plus, la variabilité de ces interactions dans l’espace devrait également être considérée. Néanmoins, différents enjeux compliquent ces analyses.

1.2.2 Enjeux analytiques et méthodologiques

Les travaux sur la mobilité intégrée sont généralement d'ordre stratégique : ils mettent l'accent sur des solutions technologiques et communicationnelles (application mobile ou web, calculateur de chemin intermodal, informations voyageurs personnalisées et en temps réel, etc.), matérielles (infrastructures intermodales) ou organisationnelles (collaboration entre les différents intervenants, intégration des tarifs), mais peu de recherches ont été menées afin de dépasser les enjeux analytiques et méthodologiques liés à l'étude de plusieurs modes de transport.

Au niveau analytique, les connaissances que l'on a sur la demande multimodale sont relativement limitées ; quels sont les comportements associés et quelles sont les motivations qui incitent ces comportements (Gebhardt et al., 2016) ? Chaque moyen de transport étant le plus souvent analysé de manière indépendante, ses interactions avec les autres modes restent également mal comprises. Dans quelle(s) situation(s) deux modes sont-ils utilisés en complémentarité ou en concurrence ? En outre, quel effet, à court et à long terme, aura l'introduction d'un nouveau mode ou d'une nouvelle infrastructure de transport sur l'utilisation des autres modes ? Cet impact est-il uniforme dans la ville ? Plus généralement, quel est le rôle de chaque mode dans la mobilité urbaine ?

Au niveau méthodologique, de nouveaux outils devront également être développés pour permettre une planification intégrée de l'ensemble des modes. Dans le cadre du projet européen SUITS visant à renforcer les capacités des villes dans l'élaboration de plans de mobilité urbaine durable, plusieurs villes ont exprimé des besoins de quantification en soutien à la mise en œuvre de mesures stratégiques (Diana et al., 2018). En particulier, des indicateurs permettant d'évaluer des niveaux de multimodalité ou d'intermodalité ainsi que la complémentarité (ou la compétition) entre les modes devraient être disponibles pour suivre l'évolution de la mobilité et informer les différents intervenants du transport. De même, l'impact d'une modification de l'offre ou de toute autre intervention stratégique sur la demande de déplacement et/ou sur la répartition modale devrait pouvoir être mesuré afin de guider les actions des planificateurs (Fuller et al., 2013).

En outre, la majorité des modèles de prévision de la demande de transport n'incluent qu'un nombre limité de modes. Généralement, deux modes de transport seulement sont considérés. Dans le cas de Montréal, le MOTREM (Modèle de transport de la région de Montréal), utilisé par le Service de la modélisation des systèmes de transport du MTQ, comprend un modèle de transfert modal permettant de traiter les déplacements bimodaux, mais seuls les modes automobile et transport en

commun sont pris en compte (Tremblay, 2002). Par ailleurs, le MOTREM est calibré à l'aide de l'enquête Origine-Destination de Montréal, qui recueille des déplacements pour une journée de semaine moyenne d'automne tous les cinq ans. Cette enquête régionale transversale ne permet donc pas de mesurer la variabilité des comportements de mobilité (entre les différents jours de la semaine mais aussi à plus long terme), ni d'évaluer les effets des grands changements survenus entre deux enquêtes quinquennales. De même, peu d'informations sont disponibles pour analyser la variabilité d'utilisation des modes dans le temps et dans l'espace.

Ces limitations sont essentiellement dues à un manque de données qui soient à la fois multimodales et longitudinales. En effet, il n'existe pas de base de données unique décrivant l'utilisation continue 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 actifs et alternatifs sont généralement sous-représentés (voire non-représentés) et les informations sont transversales. D'autre part, des flux longitudinaux de données passives peuvent être exploités, mais ceux-ci sont collectés de manière indépendante et hétérogène pour chaque mode ; elles rendent donc difficile l'analyse des interactions entre les modes. En particulier, l'intermodalité n'est pas capturée dans les flux passifs car les segments des déplacements réalisés avec plusieurs modes sont recueillis dans des ensembles de données différents. Par conséquent, de nouvelles méthodes doivent donc être développées afin de permettre l'intégration de diverses sources de données.

1.3 Objectifs

En réponse aux enjeux discutés à la section précédente, cette recherche s'inscrit dans la perspective d'effacer les frontières analytiques et méthodologiques entre l'étude des différents modes de transport et des différentes unités spatio-temporelles pour tendre vers une analyse multimodale et longitudinale de la demande de transport. Plus particulièrement, cette thèse vise à mieux comprendre, à évaluer et à modéliser les interactions entre différents modes de transport d'un système multimodal urbain. Pour ce faire, des sources diversifiées de données devront d'abord être exploitées et intégrées. Les contributions proposées seront découpées en six articles et répondront aux trois objectifs suivants :

- **Élaborer des procédures automatisées de valorisation et d'enrichissement de données provenant d'enquêtes et de flux passifs afin de permettre un suivi longitudinal de l'utilisation de différents modes de transport**

Des méthodes de pondération et de projection seront développées en vue de comparer (**article 1**) puis de fusionner (**article 2**) ces deux types de données. Dans un premier temps, les deux sources de données seront valorisées pour analyser la variabilité temporelle de l'utilisation de différents modes de transport sur plusieurs mois d'automne. Dans un deuxième temps, les indicateurs de suivi de l'utilisation des différents modes, calculés à partir de l'enquête Origine-Destination de Montréal pour une journée moyenne d'automne, seront annualisés et projetés grâce aux données passives.

- **Proposer un cadre conceptuel et une boîte à outils afin de qualifier et quantifier différents types d'interactions entre les modes de transport à plusieurs niveaux d'analyse**

Plusieurs types d'interactions seront définis et mis en évidence à différents niveaux d'analyse (individu, système, déplacement) et pour différentes échelles spatio-temporelles (du global au local, et du long terme au court terme). Au niveau individuel, la multimodalité des comportements de déplacements sera évaluée à l'aide d'un indicateur et son évolution sur cinq ans sera étudiée (**article 3**). Au niveau systémique, des relations intermodales dans le voisinage des stations de métro seront examinées (**article 4**) et des corrélations d'usage entre plusieurs modes seront estimées (**article 5**) à l'aide de plusieurs mois de données. Au niveau du déplacement, un algorithme sera développé afin de détecter différentes situations de complémentarité et de compétition entre deux modes pendant plusieurs journées (**article 6**).

- **Développer un ou plusieurs modèles pour expliquer/contextualiser les interactions entre les modes et évaluer leurs impacts sur la demande de transport**

L'effet de différents facteurs contextuels en lien avec les caractéristiques socio-démographiques, l'offre de transport (**article 3**), le calendrier, la météo, l'environnement bâti et l'utilisation du sol (**article 4**), ou encore les mesures et événements de la pandémie de COVID-19 (**article 5**), sera analysé et modélisé. Finalement, l'impact causal de la concurrence entre les modes sur l'utilisation du transport en commun sera évalué (**article 6**).

1.4 Structure du document

Cette thèse est découpée en plusieurs sections. Le présent chapitre (Chapitre 1) introduit le contexte dans lequel s'inscrit cette recherche et précise le cas d'étude (Montréal). Trois objectifs sont également formulés afin de répondre à la problématique énoncée et aux enjeux associés. Le Chapitre 2 résume la littérature existante pour chacun des trois objectifs fixés. Le Chapitre 3 propose un cadre conceptuel pour définir différents types d'interactions entre les modes et décrit les données utilisées ainsi que la méthodologie générale appliquée pour qualifier, quantifier et modéliser ces interactions. Les Chapitres 4 à 9 présentent les différentes contributions réalisées regroupées en six articles. Le Chapitre 10 fournit une synthèse des principaux résultats obtenus et discute de leur portée. Finalement, le Chapitre 11 conclut cette thèse en soulignant les contributions et limites de chaque article et en recommandant des perspectives d'analyses futures.

CHAPITRE 2 REVUE DE LITTÉRATURE

La revue de littérature proposée dans ce chapitre est composée de trois sections qui correspondent chacune à un des trois objectifs formulés précédemment. Tout d’abord, différentes techniques de fusion de données sont recensées afin de mettre en contexte la combinaison entre des données passives et les données d’une enquête de déplacements traditionnelle. Ensuite, plusieurs travaux ayant trait à l’analyse des interactions entre les modes de transport sont rapportés. Finalement, différentes approches de modélisation de la demande de transport sont présentées. Cet état de l’art permet ainsi de dresser un portrait détaillé des résultats analytiques déjà obtenus sur le sujet de la mobilité multimodale et intégrée. Il permet également de discuter les méthodes et outils de fusion et de modélisation existants afin d’en pointer les lacunes à dépasser dans cette recherche.

2.1 Techniques de fusion de données

Le concept de fusion (ou d’intégration) de données n’est pas récent. Cependant, il est aujourd’hui un sujet de recherche en pleine expansion en raison de la disponibilité croissante de données (développement du Big Data), conjuguée avec des besoins informationnels de plus en plus exigeants (on veut des données complètes, désagrégées, en temps réel, etc.). Cette popularité a conduit à un grand nombre de publications visant à théoriser et définir le concept de manière générale (section 2.1.1), ainsi qu’à classer les méthodes et défis associés (section 2.1.2).

Historiquement, les techniques de fusion ont d’abord été utilisées dans les domaines civil et militaire pour la reconnaissance et le suivi d’objets, la surveillance de l’habitat d’espèces sauvages et la détection des risques pour l’environnement (El Faouzi et al., 2011). Elles ont ensuite été diffusées dans différentes spécialités de la recherche scientifique, incluant l’étude des systèmes multisensoriels, l’analyse des réseaux de capteurs, le traitement d’images et de vidéos et la conception de systèmes intelligents. La robotique, le biomédical, la santé, l’environnement et bien sûr le transport, spécifiquement couvert dans cette recherche (section 2.1.3), sont ainsi des exemples de domaines d’application (Khaleghi et al., 2013; Lahat et al., 2015).

2.1.1 Définition générale de la fusion de données

Du fait de cette multidisciplinarité, il existe une multitude de synonymes au mot « fusion de données ». En anglais, les termes suivants sont employés à tour de rôle : *data fusion*, *information*

fusion, decision fusion, data combination, data aggregation, data integration, multisensor data fusion, sensor fusion, statistical matching, file grafting, ou encore *microdata merging*. De même, plusieurs définitions ont été proposées dans la littérature. La définition la plus ancienne et sûrement aussi la plus citée est celle du département de défense américain (White, 1991), qui désigne la fusion de données comme :

“A process dealing with the automatic detection, association, correlation, estimation, and combination of data and information from single and multiple sources to achieve refined position and identity estimates, and complete and timely assessments of situations and threats and their significance.”

Cette définition issue du domaine militaire est orientée sur l'évaluation des risques et de situations dangereuses. Toutefois, des efforts de conceptualisation ont ensuite été réalisés afin de généraliser la théorie de la fusion de données à tous les domaines d'application. Après avoir discuté différentes définitions recensées dans la littérature, Boström et al. (2007) arrivent à la conclusion suivante :

“Information fusion is the study of efficient methods for automatically or semi-automatically transforming information from different sources and different points in time into a representation that provides effective support for human or automated decision making.”

Cette définition plus standard implique la transformation et la combinaison d'informations qui peuvent provenir de différentes sources, ou d'une seule source pour différentes périodes temporelles. Les auteurs précisent cette définition en indiquant que les sources peuvent être de types variés (capteurs, bases de données, simulations, etc.) et que l'information traitée peut également se présenter sous différentes formes (texte, nombre, graphique, etc.). De plus, les auteurs supposent l'application de méthodes automatisées ou semi-automatisées. Leur définition intègre également une réflexion sur la représentation des informations issues de la transformation. Par ailleurs, la procédure de fusion de données doit être efficiente et efficace : les contributions au domaine seront donc évaluées à partir de l'effet de cette procédure sur les processus décisionnels mais aussi sur le coût dépensé pour arriver à cet effet (en termes de temps et de ressources).

De manière encore plus générale, en croisant différents travaux de la littérature, il est possible d'énoncer qu'une fusion de données se traduit par la combinaison de différentes données ou informations (provenant d'une seule ou de plusieurs sources) dans l'objectif d'augmenter ou d'améliorer les connaissances initiales qui auraient été obtenues avec une seule base de données. Le principe derrière cette définition est de compenser les limitations respectives de chaque base de

données et de tirer parti d'un tout (Cherchi & Bhat, 2018). Une information améliorée peut être une information moins coûteuse, de meilleure qualité, plus fiable, plus complète ou plus pertinente (Castanedo, 2013). Les techniques de fusion de données peuvent notamment être appliquées pour produire un meilleur modèle ou une meilleure estimation d'un système ou d'un phénomène (El Faouzi et al., 2011). D'autres avantages peuvent être tirés de cette manipulation de données, tels que la suppression des doublons dans les informations collectées, l'extension de la couverture spatiale ou temporelle de l'analyse, ou l'accroissement de la disponibilité des données, entre autres par l'augmentation de la taille de l'échantillon initial (Khaleghi et al., 2013). Les résultats de la fusion pourront prendre différentes formes, par exemple via les paramètres d'un modèle ou la construction d'une seule base de données à partir de plusieurs (Venigalla, 2004).

2.1.2 Méthodes et défis généraux de la fusion de données

De nombreuses méthodes peuvent être utilisées pour la fusion de données, depuis les méthodes classiques de classification (K-moyennes, K plus proches voisins) jusqu'à des algorithmes d'apprentissage automatique plus complexes (exemple : réseaux de neurones), en passant par des méthodes de filtrage (exemple : filtre de Kalman), des distributions de probabilités et des techniques statistiques d'inférence bayésienne ou basée sur la théorie de Dempster-Shafer. La revue de littérature de Castanedo (2013) peut être consultée pour des exemples d'application de ces méthodes. Néanmoins, le choix de la méthode la plus appropriée étant une des étapes les plus difficiles dans un processus d'intégration de données, des typologies de méthodes ont été formalisées afin de guider les spécialistes.

Les principales classifications sont résumées dans le Tableau 2.1. Elles sont basées sur différents critères, notamment la nature et les propriétés des données (Dasarathy, 1997; Durrant-Whyte, 1988; Khaleghi et al., 2013; Luo, R. C. et al., 2002), l'objectif visé et son niveau d'abstraction (Castanedo, 2013; White, 1991), l'architecture du processus de fusion (Esteban et al., 2005) ou les propriétés mathématiques des méthodes (Alam et al., 2017; El Faouzi et al., 2011). Plus récemment, des classifications multi-perspectives combinant plusieurs de ces critères ont été proposées (Lau et al., 2019; Ounoughi & Ben Yahia, 2022; Wang, Meisong et al., 2016; Zheng, 2015). Le tableau montre également une évolution dans les exemples d'application des méthodes, d'abord centrés sur des réseaux de capteurs puis élargis à des sujets plus larges tels que l'internet des objets et les systèmes intelligents, par exemple les systèmes de transport intelligents (STI).

Tableau 2.1 Synthèse des principales classifications de méthodes de fusion de données

Référence	Application	Critère(s) de classification
(Durrant-Whyte, 1988)	Capteurs	Relations entre les sources de données à fusionner : <i>complémentaires, redondantes</i> ou <i>coopératives</i>
(White, 1991)	Militaire	5 niveaux de traitement : <i>niveau 0</i> = prétraitement des données, <i>niveau 1</i> = détection de l'objet étudié, <i>niveau 2</i> = évaluation de la situation, <i>niveau 3</i> = évaluation de l'impact, <i>niveau 4</i> = optimisation du processus complet
(Dasarathy, 1997)	Capteurs	Nature des données (données brutes, attributs ou décisions) en entrée et en sortie du processus de fusion
(Luo, R. C. et al., 2002)	Capteurs	Niveau d'abstraction des données : <i>faible</i> (mesures brutes : signaux, pixels, etc.), <i>moyen</i> (caractéristiques ou attributs tels que la forme, la texture ou la position) ou <i>élevé</i> (décisions)
(Esteban et al., 2005)	Capteurs	Architecture/organisation du processus de fusion (nombre de nœuds de calcul, rôle et position de chaque nœud, etc.) : <i>centralisée, décentralisée, distribuée</i> ou <i>hiérarchique (hybride)</i>
(El Faouzi et al., 2011)	STI	Propriétés mathématiques : approches <i>statistiques</i> , approches <i>probabilistes</i> (bayésiennes) ou approches basées sur l' <i>intelligence artificielle</i>
(Castanedo, 2013)	Capteurs	Objectif de la procédure de fusion : <i>associer des données, estimer un état</i> ou <i>prendre une décision</i>
(Khaleghi et al., 2013)	Capteurs	Enjeux rencontrés au niveau des données à fusionner : <i>imperfection</i> (problème d'incertitude, d'imprécision ou de granularité), <i>corrélation, inconsistance</i> (valeurs conflictuelles, aberrantes ou en désordre) ou <i>disparité</i>
(Zheng, 2015)	Fusion inter-domaines	Approche multicritère : type de méthodes (par étapes, basées sur un attribut ou sur la signification sémantique), volume (petit ou gros) et propriétés (fixes ou flexibles) des données, objectif (5 types), type d'apprentissage (supervisé, semi-supervisé ou non-supervisé) et ampleur de la tâche
(Wang, Meisong et al., 2016)	Internet des objets	Approche multicritère : architecture informatique, prise en compte du contexte, niveaux d'interaction sémantique, de configuration dynamique et de complexité, existence d'un actionneur dans le système automatisé, type de procédure, nombre de domaines d'application, existence d'une méthode d'implémentation et d'une procédure d'évaluation de la performance
(Alam et al., 2017)	Internet des objets	Propriétés mathématiques : approches <i>probabilistes</i> , approches basées sur l' <i>intelligence artificielle</i> et approches basées sur la théorie de Dempster-Shafer
(Lau et al., 2019)	Ville intelligente	Approche multicritère : objectif (4 types), famille de techniques (9 types), nature des données en entrée et sortie (5 types), source des données (4 types), échelle (5 types), architecture de la plateforme (4 types)
(Ounoughi & Ben Yahia, 2022)	STI	Approche multicritère : approche mathématique (4 types), propriété des données (7 types), technologie de collecte des données (30 types), évaluation (cas réel ou simulé), domaine d'application (7 types)

Outre l'identification d'une méthode, de nombreux autres défis sont liés à la fusion de données. Plusieurs chercheurs (Bayart et al., 2009; Cherchi & Bhat, 2018; El Faouzi et al., 2011; Lahat et al., 2015; Lau et al., 2019) ont tenté de les lister de manière exhaustive. Communs à tous les domaines, ces défis peuvent être rapportés en deux grandes catégories.

➤ **Enjeux liés aux méthodes**

Le coût associé à chaque méthode, notamment en termes de temps de calcul, est un des principaux enjeux à considérer dans un processus de fusion de données. La plupart des méthodes répertoriées dans les typologies précédentes sont en effet très lourdes. De plus, elles sont généralement complexes et difficiles à interpréter. Ainsi, malgré les efforts de formalisation et de classification des méthodes de fusion existantes, leur applicabilité dans des cadres analytiques parfois très différents reste souvent assez difficile à mettre en œuvre. Trouver et comprendre les interactions entre les bases de données à fusionner est également une étape délicate. Enfin, tous les auteurs s'accordent pour dire que des méthodes de validation (ou des indicateurs de performance) doivent être développées afin de quantifier la fiabilité des processus de fusion.

➤ **Enjeux liés aux données**

Au-delà des enjeux se rapportant à l'acquisition des données et à la protection de la vie privée lorsque les données en question sont des données désagrégées, un défi de standardisation des protocoles et des formats (souvent hétérogènes) est souvent relevé. L'incompatibilité du contenu des diverses bases de données à fusionner peut provenir de différences dans les unités (problème de commensurabilité), dans la résolution ou la couverture spatio-temporelle, ou dans la taille des échantillons. Bayart et al. (2009) illustrent trois grands types d'incompatibilités (spatiale, temporelle ou sémantique) avec des données du domaine du transport. Des problèmes d'alignement (ou de recalage en traitement de l'image) dus à des distorsions spatiales peuvent également survenir dans le cas de l'intégration de données géospatiales. De plus, il est nécessaire de développer des méthodes pour valider la qualité des données fournies en entrée du processus de fusion (bruit, incertitudes, données manquantes, etc.) et contrôler leurs interdépendances. La communauté scientifique souligne aussi un besoin d'adaptation des systèmes d'exploitation commerciaux et de la gestion des bases de données (stockage des données massives, mise à jour rapide, continue et en temps réel, etc.). Enfin, des interfaces flexibles et conviviales doivent être créées pour faciliter la représentation des données et des résultats de la fusion.

2.1.3 Application au domaine du transport

2.1.3.1 Contexte

Le concept d'intégration ou de fusion de données n'est pas non plus nouveau en transport. On a très vite remarqué que les multiples approches employées pour réaliser des enquêtes allaient poser problème une fois le moment venu de les assembler (Goulias, 1997). Les premières mentions du terme « fusion de données » ont été rapportées pour la combinaison d'enquêtes de préférences déclarées et d'enquêtes de préférences révélées (Cherchi & Bhat, 2018). De même, bien qu'on ne dénomme pas souvent ce processus comme une fusion de données, on utilise depuis longtemps des données de recensement pour redresser les enquêtes de déplacements : il s'agit là d'une forme d'intégration de deux types d'enquêtes. En particulier, la synthèse de population soulève des enjeux similaires à la fusion de données (Bayart et al., 2009).

Au cours des dernières années, le recours au concept de fusion de données dans le domaine du transport s'est intensifié. D'après Bayart et al. (2009), ce regain d'intérêt peut être expliqué par les trois raisons principales détaillées ci-après.

2.1.3.1.1 *La baisse de la représentativité des enquêtes de transport*

Une baisse des taux de réponse aux enquêtes traditionnelles de transport est constatée depuis quelques années déjà (Stopher & Greaves, 2007; Zmud, 2003). Cette tendance est due à différents facteurs, notamment à la lassitude des répondants (du fait du nombre croissant d'enquêtes, en particulier des sondages réalisés à des fins commerciales) et à une anxiété grandissante de divulguer ses informations personnelles. Ces phénomènes occasionnent des refus de plus en plus nombreux et des questionnaires de moins en moins complets. Aussi, des difficultés sont rencontrées lorsqu'il s'agit de joindre toutes les catégories de la population avec un seul mode d'enquête (Behrens et al., 2009; Verreault & Morency, 2018) : les plans d'échantillonnage sont souvent basés sur des listes de numéros de téléphone résidentiel mais, du fait de la popularité des téléphones portables, un nombre croissant de ménages ne disposent plus de ligne fixe.

Ces problématiques de sous-représentativité conduisent généralement à l'utilisation de méthodes de pondération. Cependant, ces méthodes impliquent des hypothèses (basées sur les caractéristiques socio-démographiques des personnes) qui sont souvent critiquées (Forsman et al., 2007). L'intégration de différentes enquêtes apparaît comme une meilleure solution, d'autant plus

qu'elle permet également de réduire le recours à des processus lourds et coûteux de collecte (Venigalla, 2004). Différents types ou méthodes d'enquêtes (exemples : recensement et enquête de déplacements, enquête nationale et enquête locale) peuvent être combinés pour générer des informations plus complètes. Différents modes d'enquêtes (exemple : enquête téléphone et enquête web) peuvent également être exploités pour obtenir un échantillon plus gros et plus représentatif (taille et couverture plus grandes).

2.1.3.1.2 La prise de conscience de la complexité des phénomènes urbains

On observe également un intérêt croissant pour la recherche multidisciplinaire, motivant l'utilisation de différents ensembles de données afin de considérer toutes les dimensions d'une problématique. Les données des systèmes de transport peuvent ainsi être croisées avec des sources de données provenant d'autres domaines comme la géographie (Rodrigue, 2020), l'urbanisme (Handy, 2005), la santé (Sallis et al., 2004; Schauder & Foley, 2015) ou le commerce (Kressner, 2017). L'étude des interactions entre la mobilité et le développement durable implique également l'intégration d'informations économiques, environnementales et sociales aux données traditionnelles. La multiplication des questions portant sur tous ces domaines dans un même questionnaire conduirait à un processus beaucoup trop lourd pour les répondants. Par conséquent, il est nécessaire de combiner des bases de données existantes.

2.1.3.1.3 La disponibilité croissante de systèmes d'observation

Comme dans de nombreux domaines, de nouvelles données opérationnelles et collectives ont été rendues disponibles en transport grâce au récent développement des technologies. Ces données sont collectées via des systèmes GPS, des compteurs, des téléphones cellulaires, des médias sociaux, des capteurs, des cartes à puce ou d'autres systèmes automatisés de collecte transactionnelle (Cherchi & Bhat, 2018). Elles sont dites passives car elles recueillent des informations sans la participation active du sujet. Les informations peuvent être enregistrées de manière continue sur de longues périodes; elles sont donc longitudinales. Cela conduit à de grandes quantités de données, c'est pourquoi elles sont aussi appelées données massives ou Big Data, qui sont souvent caractérisées par « 4 V »: volume, vélocité, variété et véracité (Callegaro & Yang, 2018). Par ailleurs, cette prolifération de données a été accompagnée par le développement d'outils de calcul modernes et puissants.

De nombreux chercheurs voient dans ces données émergentes une opportunité pour compléter les données traditionnelles, en particulier les enquêtes (Bonnell & Munizaga, 2018; Miller et al., 2018). En revanche, ils partagent le point de vue selon lequel une substitution n'est pas envisageable pour le moment; les deux types de données doivent être combinés pour maximiser leur valeur respective (Callegaro & Yang, 2018). En transport en particulier, chaque base de données prise séparément ne contient généralement pas toutes les informations nécessaires pour répondre aux besoins que l'on a (Polak & Cornelis, 2009). Ainsi, les avantages et les inconvénients de chaque type de données sont souvent listés afin de comprendre comment les différentes informations disponibles peuvent se compléter (Bayart et al., 2009; Callegaro & Yang, 2018; Cherchi & Bhat, 2018) dans le cadre d'un changement de paradigme du « Big Data » vers le « Rich Data » (Callegaro & Yang, 2018).

Les données passives, initialement prévues pour des fonctions autres que l'analyse et la modélisation des comportements de mobilité (par exemple, les systèmes de cartes à puce ont été conçus pour la collection des recettes tarifaires), ont le potentiel de fournir des informations spatio-temporelles très précises. En revanche, elles ne contiennent aucune donnée contextuelle sur les individus, principalement en raison des enjeux sur la confidentialité et la protection de la vie privée. Au contraire, les enquêtes traditionnelles apportent des renseignements complets sur les caractéristiques socio-démographiques des répondants. Cependant, les enquêtes traditionnelles collectent généralement des carnets de déplacement d'une seule journée, sans prendre en compte les fluctuations observées d'un jour à l'autre de la semaine (Verreault & Morency, 2011), et elles sont transversales, c'est-à-dire collectées à des moments précis séparés de plusieurs années. Par conséquent, les informations recueillies deviennent rapidement datées et moins représentatives des conditions actuelles, alors que les données émergentes sont longitudinales et peuvent être extraites à tout moment (parfois même en temps réel).

Bien qu'elles soient généralement menées sur plusieurs semaines ou plusieurs mois, les grandes enquêtes ménages déplacements sont aussi considérées comme transversales car les échantillons d'une journée sont regroupés pour représenter une « journée type » (El-Assi et al., 2020), concept qui a été beaucoup remis en question dans la littérature (Schönfelder & Axhausen, 2010). Il existe des alternatives aux enquêtes transversales, notamment les enquêtes multi-jours, les enquêtes par panel et les enquêtes continues - voir les travaux de Ampt (2013) et Ortúzar et al. (2011) pour des revues complètes sur le sujet, mais elles sont beaucoup plus coûteuses et laborieuses que les

données émergentes, collectées passivement et donc sans fardeau pour le répondant. De plus, les bénéfices de ces enquêtes ont encore peu été démontrés (El-Assi et al., 2020; Miller et al., 2018).

Du fait de ces différences de contenu et de nature, les données traditionnelles et émergentes sont généralement exploitées pour des objectifs distincts. Les premières permettent de modéliser et d'expliquer les comportements de mobilité, alors que les secondes sont particulièrement adaptées pour étudier la variabilité (inter et intrapersonnelle) de ces comportements - voir par exemple le mémoire de maîtrise de Deschaintres (2018) pour la variabilité d'utilisation du transport en commun évaluée à partir de données de cartes à puce. Par conséquent, la complémentarité entre les sources de données traditionnelles et les sources de données émergentes tient notamment dans le fait que l'une donne le « pourquoi » et l'autre le « comment » (Callegaro & Yang, 2018).

D'autres arguments ont été avancés pour expliquer en quoi les données émergentes ne peuvent pas remplacer les enquêtes traditionnelles. Les préoccupations concernant la qualité, l'exhaustivité et la représentativité des données passives en font partie (Bonnel & Munizaga, 2018; Miller et al., 2018). En effet, même si elles sont disponibles en grande quantité, ces données peuvent également être incomplètes et imparfaites : le quatrième V (véracité) du Big Data devrait toujours être vérifié (Lovelace et al., 2016). En outre, ces données massives sont difficiles à traiter, et les méthodes de collecte de données sont encore en évolution, entraînant des changements rapides dans l'état, la forme et la quantité des données recueillies (Bayart et al., 2009; Miller et al., 2018).

2.1.3.2 Méthodes et défis spécifiques au transport

En appliquant le cadre conceptuel de D'Orazio et al. (2006) et D'Ambrosio et al. (2007) dans un contexte de transport, Bayart et al. (2009) définissent le problème de fusion de données comme la combinaison d'un ensemble de données *donneur* et d'un ensemble de données *récepteur* afin de transmettre des informations supplémentaires (*variables d'intérêts* observées dans l'ensemble de données donneur mais pas dans l'ensemble de données récepteur) grâce à une *fonction de liaison* se basant sur des *variables communes* aux deux ensembles de données. Les variables communes peuvent être spatiales, temporelles ou sémantiques et dépendent de la granularité des données. Ainsi, un processus de fusion s'articule en quatre étapes principales (Venigalla, 2004) : 1) Identifier les bases de données appropriées et les éléments (variables ou attributs) en jeu; 2) Examiner les caractéristiques des données de chacune des sources de données impliquées; 3) Identifier des éléments communs (ou similaires, c'est-à-dire statistiquement compatibles) pour faciliter la fusion

de données; 4) Analyser et intégrer les ensembles de données en choisissant une méthode, une plateforme d'analyse (un logiciel statistique, un format de base de données, etc.) et en trouvant un moyen de présenter les résultats de la fusion.

Les méthodes appliquées en transport sont généralement classées en deux catégories (Bayart et al., 2009; Zhu et al., 2018) : les méthodes de correspondance exacte (*exact matching*) lorsqu'il existe un identifiant unique permettant de relier directement les deux jeux de données ou, plus couramment, les méthodes basées sur des modèles statistiques (*statistical matching*). Ces derniers peuvent être *explicites* ou *implicites*. La méthode explicite consiste à entraîner un modèle classique (régression multiple, régression logistique ou modèle non paramétrique tel qu'un arbre de décision) sur le jeu de données donneur, puis à l'appliquer sur le jeu de données récepteur. La méthode implicite est utilisée lorsque chaque enregistrement de l'ensemble de données donneur est associé à un (appariement *contraint*) ou plusieurs (appariement *non contraint*) enregistrements similaire(s) de l'ensemble de données récepteur, par exemple via l'algorithme des K plus proches voisins.

Toutefois, bien que l'intérêt de fusionner des données, notamment pour augmenter les connaissances sur les comportements de déplacements et améliorer les modèles existants (de plus en plus exigeants en termes de résolution et de données), soit reconnu par la communauté scientifique (Bayart et al., 2009), et malgré l'existence de nombreuses techniques dans d'autres domaines qui ont plus d'expérience sur le sujet, très peu de méthodes ont déjà été appliquées dans un contexte de transport. Des recherches supplémentaires et interdisciplinaires sont nécessaires afin de mieux définir le rôle de chaque source de données et développer de nouveaux outils permettant de les intégrer (Cherchi & Bhat, 2018).

L'intégration de différents modes d'enquêtes implique l'élimination d'un biais de sélection du mode d'enquête. Pour dépasser ce problème, des méthodes permettant de distinguer les différences de réponse dues à la méthodologie et les différences réelles de comportement doivent être élaborées. Il faut donc modéliser à la fois le choix du mode d'enquête et le phénomène de transport étudié (Bayart et al., 2009). De même, dans le cas de l'intégration entre des données massives et des enquêtes traditionnelles, plusieurs défis spécifiques peuvent être soulignés. Le premier enjeu rencontré est lié à la disponibilité des données émergentes. En effet, les différentes bases de données (généralement une par mode ou système de transport) peuvent appartenir à différents opérateurs privés (Miller et al., 2018). D'autres difficultés sont liées à l'hétérogénéité des formats

et des unités de mesure de l'utilisation des modes (ou indicateurs de mobilité) entre les différents ensembles de données. Les enquêtes contiennent des déplacements alors que des transactions, des comptages, des traces GPS ou d'autres types d'information sont enregistrés dans les données passives. En outre, des individus sont interrogés dans les enquêtes alors que les données passives peuvent seulement être regroupés par carte à puce, par compteur ou par système de collecte, sans information sur les individus observés. Ces enjeux de comparabilité des données posent notamment problème pour trouver des variables communes (Bonnell & Munizaga, 2018).

Au-delà de cette dissimilitude sémantique, il existe également des incompatibilités temporelles et spatiales entre les données d'enquêtes et les données passives, qui peuvent avoir été collectées avec des niveaux de résolution différents. En particulier, les zones de couverture (région entière versus zone de desserte d'un mode donné) et la population de référence correspondante diffèrent d'un type de données à l'autre. Les déplacements des non-résidents (visiteurs) et les déplacements commerciaux ne sont pas couverts par les enquêtes traditionnelles, alors que tous les déplacements réalisés via le système sont saisis automatiquement dans les données passives (Miller et al., 2018). De même, des échantillons plus ou moins représentatifs versus des inventaires presque complets (excluant les données perdues par fraude ou dysfonctionnement du système de collecte) sont disponibles respectivement dans les enquêtes et les données passives. Toutes ces disparités s'expliquent par les rôles initiaux différents des deux sources de données. Pour les compenser et permettre la valorisation des deux types de données, des prétraitements et une agrégation des données est généralement nécessaire, entraînant alors une perte de précision.

2.1.3.3 Exemples d'application

Ces multiples enjeux expliquent le nombre relativement limité d'exemples d'application de la fusion de données en transport. Ces derniers sont regroupés en plusieurs sous-sections ci-dessous. L'Annexe A fournit une synthèse des différents travaux recensés (dans leur ordre d'apparition) en précisant les données utilisées, l'objectif visé, la méthode employée et son type dans la classification présentée précédemment.

2.1.3.3.1 Fusion inter-domaines

Les exemples les plus courants sont des cas de combinaison de données inter-domaines, c'est-à-dire lorsque les données du transport sont croisées avec des données provenant d'autres secteurs.

Cette combinaison est le plus souvent réalisée à partir de variables temporelles ou spatiales communes, donc via une correspondance exacte (*exact matching*) établie à un niveau plus ou moins désagré. Dans le cas spatial, des systèmes d'information géographique (SIG) sont utilisés afin de combiner une base de données d'objets géolocalisés avec une autre base de données géospatiales contenant des informations contextuelles, décrivant par exemple l'utilisation du sol (An et al., 2019; Jun et al., 2015). La fusion se fait généralement à partir des coordonnées, mais il existe d'autres méthodes pour intégrer des jeux de données spatiales à partir du nom ou du type de lieu, d'une mesure de similarité combinée (Sehgal et al., 2006), ou via des algorithmes de fusion plus évolués basés sur la géométrie (Stankutė & Asche, 2009). Le résultat de l'intégration des deux bases de données permet alors de caractériser l'environnement autour des objets étudiés. Par ailleurs, différentes méthodes de cartospondance (ou appariement cartographique, *map matching* en anglais) ont été appliquées en transport afin de faire correspondre des points GPS à un élément du réseau, tel qu'un lien routier ou un arrêt d'autobus (Quddus et al., 2007).

Pour des applications liées aux secteurs de la santé et du développement durable, notamment le croisement de données du transport avec des informations sur l'activité physique et la mixité sociale, les travaux de maîtrise de Mageau-Béland (2019) peuvent être consultés. Aussi, des informations autodéclarées (données d'enquête) et des données médicales (examen médical et données de laboratoire) ont été combinées afin de modéliser l'impact du transport actif sur différentes variables reflétant le niveau de santé (IMC, obésité, pression artérielle, taux de cholestérol, etc.) (Schauder & Foley, 2015).

2.1.3.3.2 Fusion intra-domaine entre plusieurs enquêtes

En matière de fusion de données intra-domaine (données du transport combinées avec d'autres données du transport), l'intégration de plusieurs enquêtes est le cas le plus fréquent. Différents chercheurs (Bhat & Castelar, 2002; Bradley & Daly, 1997; Cherchi & Ortúzar, 2002) ont ainsi proposé des modèles logit de choix discret incorporant des variables provenant d'enquêtes de préférences déclarées et d'enquêtes de préférences révélées (modèles mixtes). Dans la même lignée, toute une série de modèles ont été développés dans l'objectif d'intégrer des informations comportementales aux données des enquêtes traditionnelles, via des modèles hybrides (modèles de choix discret incluant un modèle de variables ou de classes latentes) (Ben-Akiva et al., 2002; Kim et al., 2014) ou, plus récemment, via des modèles généralisés de données hétérogènes qui peuvent

provenir de différentes sources (Bhat, 2015). En plus d'être utilisées pour redresser les enquêtes de déplacements, les données de recensement peuvent également être exploitées pour les enrichir à un niveau totalement désagrégé. Certains attributs sociaux, économiques et démographiques du recensement peuvent ainsi être intégrés à l'étude des comportements de mobilité (Morency, 2004; Venigalla, 2004). En réponse aux enjeux de baisse de la représentativité discutés précédemment, les travaux les plus récents se sont concentrés sur la combinaison de plusieurs modes d'enquête (Bonnell et al., 2015; Hilgert et al., 2017; Verreault & Morency, 2018) et sur l'intégration d'une enquête satellite spéciale (par exemple, ciblant les étudiants) à une enquête ménages déplacements centrale (Verreault & Morency, 2018; Wang, Kaili et al., 2021). Ces fusions d'enquêtes peuvent être réalisées grâce à une méthode de repondération (Verreault & Morency, 2018), une méthode d'appariement implicite (Hilgert et al., 2017) ou un modèle explicite (Wang, Kaili et al., 2021).

2.1.3.3 Fusion intra-domaine entre plusieurs sources de données passives

La plupart des exemples de fusion entre plusieurs sources de données passives sont recensés dans les domaines de l'ingénierie du trafic routier, de la sécurité routière et des systèmes de transport intelligents, secteurs dans lesquels l'utilisation de différentes sortes de capteurs (boucles de détection, véhicules équipés de dispositifs Bluetooth ou GPS, système automatique de reconnaissance des plaques d'immatriculation, caméras, etc.) est répandue (El Faouzi et al., 2011). Des travaux ont été réalisés pour l'estimation de variables d'état du trafic (débit, temps de parcours, vitesse, etc.), la prévision et le suivi des volumes de trafic routier, l'estimation de la demande sous forme de matrices origine-destination, la détection automatique de différents phénomènes ou objets (incidents, collisions, types de véhicules, etc.), l'estimation précise de la position, ou bien pour différentes nécessités dans le cadre de systèmes avancés d'informations pour les voyageurs, d'assistance au conducteur, de gestion des services de transport ou de communication entre véhicules (véhicules autonomes) – voir les revues de littérature de El Faouzi et Klein (2016) et Ounoughi et Ben Yahia (2022).

Plusieurs bases de données passives indépendantes ont également été utilisées en complémentarité pour valider de manière croisée (Lovelace et al., 2016) ou compléter (Giraud et al., 2016) certaines informations, comme les origines et/ou les destinations des déplacements. Les enregistrements des différentes sources de données peuvent être reliés grâce à l'horodatage et la localisation sur le réseau, mais aussi en se basant sur les habitudes des usagers et le service programmé.

2.1.3.3.4 *Fusion intra-domaine entre enquêtes et données passives*

Du fait des enjeux discutés précédemment, les exemples d'intégration entre des enquêtes et des données passives se font encore plus rares. La plupart du temps, de simples comparaisons sont réalisées entre les deux types de données afin de valider les informations recueillies, voire parfois de les corriger. Par exemple, des données de cartes à puce ont été comparées de manière agrégée à des données d'enquête, via des distributions spatiales (matrices Origine-Destination) et temporelles (Egu & Bonnel, 2020; Spurr et al., 2014) ou par le calcul d'indicateurs (Trépanier et al., 2009). Des comparaisons désagrégées ont aussi été réalisées lorsqu'il était possible de relier directement les deux sources de données (Munizaga et al., 2014). En outre, une méthode de repondération de l'enquête Origine-Destination de Montréal a été proposée afin de corriger la surestimation des déplacements réguliers en heures de pointe dans l'achalandage du métro à partir des données de cartes à puce (Spurr et al., 2014).

Au-delà de simples comparaisons, des procédures d'enrichissement ont aussi été développées afin d'ajouter des informations supplémentaires dans des données passives grâce à des enquêtes, par exemple les lieux d'origine et de destination (Long & Thill, 2015) ou le motif de déplacement (Alsger et al., 2018; Kusakabe & Asakura, 2014; Zhong et al., 2014) dans des données de cartes à puce, ou bien l'heure de départ dans des données de téléphones cellulaires (Alexander et al., 2015). Ce transfert d'attributs est produit via des algorithmes basés sur des règles et entraînés sur les données d'une enquête, ou grâce à des modèles probabilistes basés sur la distribution d'une ou de plusieurs variables de déplacement (temporelles, spatiales ou fréquentielles) communes supposée constante entre les deux bases de données. L'inférence bayésienne permet également de modéliser la relation entre un attribut pouvant uniquement être dérivé à partir des données de cartes à puce (exemple : la fréquence de déplacement) et un attribut observé uniquement dans les données de l'enquête (exemple : le motif de déplacement), c'est-à-dire une relation qui ne peut pas être étudiée à partir de l'un ou l'autre jeu de données pris séparément (Kusakabe & Asakura, 2014).

Des méthodes implicites contraintes ont également été proposées afin de faire correspondre chaque numéro de carte à puce à un répondant de l'enquête Origine-Destination de Montréal (Spurr et al., 2015) ou à un agent d'une population synthétique (Grapperon et al., 2016). Ces appariements permettent alors de transférer les informations socio-démographiques d'une enquête vers des données passives. De même, dans le contexte du transport commercial, des traces GPS converties

en déplacements ont été reliées aux déplacements d'une enquête par interception de camions afin d'imputer les variables manquantes dans les données GPS (Zhu et al., 2018).

2.1.4 Synthèse spécifique

Ainsi, malgré une littérature très abondante sur la fusion de données dans de nombreux domaines, la complexité des méthodes et les enjeux associés à ce type de problème ont conduit à peu d'exemples d'application dans un contexte de transport, et encore moins dans le cas d'une fusion intra-domaine entre des données d'enquêtes et des données passives. La plupart des auteurs cités précédemment se sont servis des enquêtes traditionnelles (ensemble de données donneur) pour compléter les informations sur les déplacements ou les individus dans les données passives (ensemble de données récepteur), ou bien les deux types de données (les deux en tant que donneur) ont été exploités à des fins de validation/comparaison, mais très peu de travaux ont utilisé les deux types de données dans l'autre direction, c'est-à-dire en enrichissant une enquête à partir de données passives. Considérer l'enquête comme jeu de données récepteur est une idée relativement nouvelle et fait écho à l'approche de collecte noyau-satellite proposée par Goulias et al. (2013) et repris dans plusieurs articles, où l'enquête ménages déplacements est le noyau et les ensembles de données passives sont auxiliaires (Miller et al., 2018; Wang, Kaili et al., 2021). Des données de téléphones cellulaires ont récemment été manipulées pour compléter ou corriger les données d'enquêtes ménages-déplacements dans la synthèse de population ou la construction de modèles de génération de déplacements (Bwambale et al., 2021; Janzen, 2017). Toutefois, aucune recherche n'a été menée pour permettre le deuxième exemple d'application proposé par Bayart et al. (2009), visant à estimer la variabilité des comportements à partir des données passives pour ensuite intégrer cette information aux données transversales d'une enquête. C'est justement là le premier objectif de cette thèse : projeter et annualiser les grands indicateurs de suivi de la mobilité calculés à partir de l'enquête Origine-Destination de Montréal en appliquant les variations longitudinales capturées dans des données passives pour plusieurs modes de transport. Par ailleurs, différentes sources de données devront aussi être intégrées pour pouvoir analyser les interactions entre plusieurs modes de transport.

2.2 Analyse des interactions entre les modes de transport

Le développement des technologies a entraîné une multiplication des sources de données mais aussi une diversification de l'offre de transport dans les villes, ce qui a fait naître des questions sur la complémentarité et la compétition entre les modes. Différentes recherches ont ainsi été menées afin d'évaluer l'impact des modes émergents sur les services de transport plus traditionnels, tels que l'automobile privée ou le transport en commun (section 2.2.1). Bien que l'analyse des interactions entre plus de deux modes de transport soit rare dans la littérature, quelques auteurs se sont penchés sur les concepts de multimodalité et d'intermodalité, ainsi que sur l'étude et la quantification des comportements correspondants (section 2.2.2).

2.2.1 Évaluation de l'impact des modes émergents sur les modes classiques

L'impact des modes de transport qui ont été récemment introduits dans les villes est un sujet qui fait l'objet de plus en plus de recherches. Même si le nombre de publications relatives à cette question reste encore assez limité pour certains modes, quelques résultats peuvent être rapportés quant à l'effet du vélopartage, de l'autopartage, des TNC (transportation network companies) et des autres modes émergents (trotinettes électriques, véhicules autonomes, etc.) sur l'offre classique de transport. L'Annexe B fournit une synthèse des articles recensés (dans leur ordre d'apparition). Les données et méthodes utilisées pour évaluer ces impacts seront aussi discutées.

2.2.1.1 Impact du vélopartage

2.2.1.1.1 Impact du vélopartage sur l'automobile

Dans un contexte où les modes de transport émergents sont généralement présentés comme des nouveaux moyens de déplacement alternatifs à l'automobile privée, plusieurs auteurs ont examiné l'impact du vélopartage sur l'utilisation de la voiture et/ou la possession automobile. À partir d'une enquête réalisée auprès des membres de différents services de vélopartage, Shaheen et al. (2013) ont observé une diminution de l'utilisation des modes motorisés (voiture privée mais surtout taxi) dans plusieurs villes nord-américaines. Cette réduction est notamment expliquée par des déplacements qui combinent le vélopartage et le transport en commun pour remplacer des déplacements qui auraient été faits auparavant en automobile. En particulier, 37% des Montréalais interrogés ont répondu utiliser moins souvent leur automobile, 44% ont déclaré moins emprunté le

taxi, et 41% ont dit avoir déjà combiné le vélopartage avec un mode de transport en commun pour réaliser un déplacement qui aurait été fait en voiture. Par ailleurs, environ 12% des répondants de cette même enquête (56% d'entre eux provenant de Montréal) ont vendu/donné un véhicule ou considèrent le vendre depuis qu'ils sont devenus membres du service de vélopartage, l'abonnement au vélopartage ayant joué un rôle très important ou relativement important dans la décision pour 55% de ceux s'étant départis d'un véhicule (Shaheen et al., 2012).

En termes de transfert modal, c'est entre 0.3% et 1.3% (selon le scénario) des déplacements automobiles qui pourraient avoir été remplacés par jour en 2010 grâce à l'implantation du système de vélopartage à Montréal en 2009 (Fuller et al., 2013). Malgré les kilomètres ajoutés pour l'opération du service (c'est-à-dire la redistribution et la maintenance de la flotte de vélos), le vélopartage a également permis de réduire le kilométrage annuel (véhicules-kilomètres) parcouru en automobile dans la majorité des villes étudiées par Fishman et al. (2014) : pour chaque kilomètre parcouru par les véhicules motorisés associés à l'opération du service, 2 à 4 kilomètres effectués par des voitures particulières ont été évités. Par conséquent, l'introduction d'un système de vélopartage permet également une diminution de la congestion routière (Hamilton & Wichman, 2018), en particulier dans les villes les plus grandes et les moins riches, où le réseau de transport en commun est plus développé et les taux de possession automobile moins élevés (Wang, Mingshu & Zhou, 2017). La combinaison du vélopartage et du transport en commun a alors plus de chances de se substituer à la voiture privée. Toutefois, la diminution de la congestion pourrait aussi potentiellement être expliquée par le remplacement de certains déplacements en bus par le vélopartage.

2.2.1.1.2 Impact du vélopartage sur les modes durables

Les bénéfices du vélopartage pour tendre vers une mobilité moins polluante et moins dépendante de l'automobile sont limités du fait de la substitution d'autres modes de transport durables (transport en commun et transport actif). En effet, bien qu'un transfert modal soit aussi observé depuis l'automobile privée, la majorité des déplacements auraient été faits en transport en commun en l'absence du service de vélopartage (Fishman et al., 2014; Fuller et al., 2013). De même, une diminution de l'usage du bus et du métro a été rapportée par les utilisateurs du service dans plusieurs villes nord-américaines (Shaheen et al., 2013). Le vélopartage peut remplacer le transport en commun comme une option de transport à part entière lorsqu'il offre une alternative plus rapide

(sans temps d'attente), moins chère et plus directe (sans correspondance) que le transport en commun, notamment pour des courts déplacements (Jin et al., 2019). Cette rivalité est potentiellement encore plus visible lorsque le système de vélopartage est sans ancrage (vélos en libre-service), car le service est alors encore plus flexible (Ma, Xiaolei et al., 2019), et lorsque les vélos partagés sont électriques, car ils permettent alors de couvrir des distances similaires à celles du transport en commun (Bieliński et al., 2021; Guidon et al., 2019).

Toutefois, les résultats de plusieurs enquêtes montrent que les individus ne font généralement pas un transfert modal « strict » d'un mode à l'autre. En réalité, ils combinent les deux modes de manière complémentaire dans leurs déplacements, augmentant ainsi leur multimodalité (Fuller et al., 2013; Shaheen et al., 2013; Tingting et al., 2011). Le vélopartage est alors utilisé pour connecter les origines et destinations des déplacements aux stations de métro, arrêts de bus ou gares de train, c'est-à-dire pour améliorer l'accès au transport en commun (enjeu du premier/dernier kilomètre). Cette intégration modale est encore plus probable lorsque les utilisateurs du vélopartage sont abonnés au service et qu'ils peuvent donc l'emprunter de manière illimitée (Kong et al., 2020).

Le sens de l'impact (positif ou négatif) du vélopartage sur l'achalandage du transport en commun mesuré au niveau systémique à partir de données passives est aussi parfois contradictoire. Entre autres, l'augmentation de l'utilisation du système de vélopartage de Washington a été corrélée avec une augmentation de l'achalandage du métro (Ma, T. et al., 2015) alors qu'une diminution de l'achalandage des lignes de bus localisées dans des zones qui disposent d'infrastructures pour le vélopartage (par rapport aux lignes de bus où le service n'est pas disponible) a été estimée à New York (Campbell & Brakewood, 2017). En exploitant les données opérationnelles de 22 grandes villes américaines, Graehler Jr et al. (2019) ont mis en évidence un effet différent selon le mode de transport en commun : le vélopartage complète le rail en jouant un rôle dans le premier ou le dernier kilomètre (Ma, T. & Knaap, 2019; Ma, Xinwei et al., 2018) voire les deux (Tarpin-Pitre & Morency, 2020), alors qu'il concurrence le bus sur des déplacements qui sont en moyenne plus courts et donc faisables entièrement en vélopartage (Campbell & Brakewood, 2017).

Les données passives permettent également de mesurer des variations spatio-temporelles dans les interactions entre les deux modes. Leur complémentarité est plus souvent observée en semaine, pendant les heures de pointe (et donc probablement pour des déplacements pendulaires domicile-travail), alors que les deux modes sont en compétition pendant la fin de semaine, pour des activités

non contraintes (Kong et al., 2020; Ma, Xiaolei et al., 2019). Les effets devraient également être évalués à différents horizons temporels : comme pour le vélo personnel, le vélopartage est potentiellement un substitut du transport en commun à court terme mais il peut aussi le compléter de manière synergique à long terme (Singleton & Clifton, 2014).

Au niveau spatial, les taux de substitution observés tendent à être plus élevés dans les grandes villes à forte densité de population que dans les plus petites villes à plus faible densité (Shaheen et al., 2013). Des impacts différents ont également été révélés au sein d'une même ville. En particulier, les personnes ayant diminué leur utilisation du transport en commun depuis leur abonnement au service de vélopartage vivent majoritairement dans le centre urbain à forte densité de population, alors que les répondants ayant augmenté leur utilisation du transport en commun sont plus fréquents dans les régions à faible densité de la périphérie urbaine (Martin & Shaheen, 2014). De même, des corrélations négatives entre le vélopartage et l'achalandage du transport en commun ont été évaluées dans le centre-ville et des corrélations positives ont été mesurées en banlieue (Ma, T. & Knaap, 2019; Tang et al., 2021). Ainsi, le vélopartage encourage l'usage du transport en commun dans les zones moins bien desservies (où le réseau est moins développé) en comblant les lacunes du réseau et en servant de moyen d'accès, tandis qu'il peut être plus compétitif (plus rapide) que le transport en commun pour les courtes distances dans les zones denses, surtout lorsque le réseau est saturé. Toutefois, à long terme, le vélopartage peut permettre de décharger le réseau et ce désengorgement peut être bénéfique pour augmenter le confort et la fiabilité du transport en commun, et donc son attractivité pour de nouveaux utilisateurs (Singleton & Clifton, 2014).

Par ailleurs, l'implantation d'un système de vélopartage est souvent accompagnée d'interventions sur l'environnement bâti (par exemple, la création de pistes cyclables) qui peuvent potentiellement inciter l'utilisation d'autres modes actifs (Fuller et al., 2013). De même, la découverte du vélopartage peut entraîner un changement vers des comportements de déplacement plus actifs. Dans les enquêtes, on constate effectivement une augmentation de l'utilisation du vélo personnel suite à l'implantation d'un système de vélopartage (Shaheen et al., 2013), même parmi les non utilisateurs du service (Fuller et al., 2013). En revanche, le vélopartage remplace aussi des déplacements qui étaient auparavant faits à pied (Fishman et al., 2014; Shaheen et al., 2013). Cette substitution peut être due à la plus grande vitesse du vélo, qui lui confère un avantage concurrentiel sur la marche.

2.2.1.2 Impact des autres modes de micromobilité

Outre le vélopartage, d'autres formes de micromobilité (incluant différentes sortes de véhicules légers à faible vitesse) sont apparues ces dernières années (Abduljabbar et al., 2021). Des vélos électriques sont parfois disponibles dans les systèmes de vélopartage (utilisation partagée), mais de plus en plus de personnes possèdent leur propre vélo électrique (utilisation privée). D'après une métaanalyse internationale basée sur 24 études, les propriétaires de vélos électriques ont principalement remplacé des déplacements en transport en commun (33%), en vélo traditionnel (27%), en automobile privée (24%) et à la marche (10%). Toutefois, les travaux les plus récents rapportent une plus grande substitution de l'automobile (Bigazzi & Wong, 2020). En outre, l'achat d'un vélo électrique conduit à une diminution de l'utilisation du vélo traditionnel (sans assistance électrique) mais à une forte hausse de l'utilisation totale du vélo, faisant alors augmenter la part modale du vélo (en kilomètres) par rapport aux autres modes empruntés par la personne (Fyhri & Beate Sundfør, 2020). Ainsi, le kilométrage parcouru en automobile par les propriétaires de vélos électriques, initialement plus dépendants que les cyclistes traditionnels, a diminué significativement dans plusieurs études (Cairns et al., 2017; Sun et al., 2020).

Par ailleurs, des flottes de trottinettes électriques partagées ont été ajoutées dans certaines villes. La revue de littérature de Wang, Kailai et al. (2022) montre que ce nouveau mode a permis de supprimer de nombreux déplacements en automobile (voiture particulière, taxi et TNC), avec des taux de substitution compris entre 25 et 40% dans la plupart des cas. C'est particulièrement vrai dans les villes américaines, où presque un tiers des déplacements automobiles font moins de deux miles d'après l'enquête nationale ménages déplacements (NHTS, pour National Household Travel Survey). Toutefois, la marche est le mode le plus substitué dans presque tous les travaux recensés (taux de substitution de 30 à 60%). En France, les utilisateurs de trottinettes ont ainsi été qualifiés de « piétons augmentés » (Krier et al., 2019). Du fait des similarités entre les deux services, les trottinettes électriques et les systèmes de vélopartage tendent également à se concurrencer, notamment durant les heures hors-pointe (Yang, H. et al., 2021). En effet, les trottinettes sont moins susceptibles d'être utilisées pour les déplacements domicile-travail que les vélos partagés, en particulier parce qu'il n'existe pas encore d'abonnement régulier (McKenzie, 2019).

Les interactions entre les trottinettes électriques et le transport en commun sont plus incertaines. Elles varient en fonction du motif de déplacement : l'utilisation des trottinettes pour des activités

sociales peut avoir un impact positif sur l'achalandage du transport en commun alors que les déplacements utilitaires ont un impact négatif puisque les services de transport en commun sont centrés sur cette demande (Ziedan et al., 2021). De plus, des situations de compétition ont été observées au centre-ville alors que des déplacements complémentaires ont été détectés en dehors du centre-ville, là où le service est plus limité (Luo, H. et al., 2021). Les trottinettes sont aussi parfois utilisées pour faire le premier/dernier kilomètre permettant de compléter un déplacement en transport en commun (Yan et al., 2021; Zuniga-Garcia et al., 2022), ou bien pour faire un trajet aller uniquement alors que le transport en commun est emprunté pour le retour (Krier et al., 2019).

Ces nouveaux modes étant très récents, peu de travaux sur le sujet ont été publiés et il est donc difficile de tirer des conclusions générales. Cependant, la direction des impacts observés est importante à déterminer car des analyses de cycle de vie ont prouvé que certaines formes de micromobilité, en particulier les services partagés de vélos (électriques ou non) et de trottinettes électriques, étaient plus polluantes que le transport en commun, notamment le bus (Hollingsworth et al., 2019; Luo, H. et al., 2019). La redistribution des vélos et des trottinettes est la principale source d'émissions de CO₂ de ces systèmes. Ainsi, Reck et al. (2022) sont arrivés au résultat inquiétant que les émissions totales des vélos et des trottinettes (dans leur forme partagée et électrique) sont supérieures à celles des modes qu'ils remplacent à Zurich.

2.2.1.3 Impact de l'autopartage

2.2.1.3.1 Impact de l'autopartage sur l'automobile

Tous les travaux recensés s'accordent pour dire que l'autopartage, autant le service basé stations que les véhicules en libre-service, entraîne une diminution de la possession automobile et de l'utilisation de la voiture. La réduction du nombre de voitures par ménage ou du kilométrage parcouru est généralement évaluée en comparant les indicateurs avant et après avoir joint un service d'autopartage (Lane, 2005; Le Vine & Polak, 2019; Martin et al., 2010), ou entre les membres et les non-membres du service (Becker et al., 2018; Cervero et al., 2007), par exemple des utilisateurs enquêtés versus des répondants à une enquête ménages déplacements ayant accès au service (Clewlow, 2016; Sioui et al., 2013).

Globalement, les ménages avec au moins un abonné au service d'autopartage possèdent moins d'automobiles que les ménages de la population de référence (Clewlow, 2016; Sioui et al., 2013).

En outre, une part plus importante des véhicules appartenant aux ménages membres d'un service d'autopartage sont des véhicules hybrides ou électriques (Clewlow, 2016). L'inscription à un service d'autopartage aurait deux types d'impacts sur la possession automobile : dans la plupart des cas (pour environ 80% des Londoniens ayant déclaré que l'autopartage a eu une incidence sur leur possession de voitures privées), il dissuade l'achat d'une nouvelle automobile, mais il encourage aussi parfois à vendre ou céder une voiture déjà possédée (Le Vine & Polak, 2019). En effet, même si environ 60% des ménages ayant souscrit au service étaient initialement des ménages non motorisés (0 automobile), le nombre moyen de véhicules par ménage a diminué de moitié après l'inscription à un service d'autopartage dans plusieurs villes d'Amérique du Nord (Martin et al., 2010). Une grande partie de cette diminution est due à des ménages étant passés de 1 à 0 voiture. Toutefois, la plupart des véhicules supprimés étaient des vieux véhicules (âge médian = 11 ans) utilisés pour faire peu de kilomètres.

La revue de littérature de Becker et al. (2018) résume les résultats de différentes études portant sur l'effet de l'autopartage sur la possession automobile (voir Table A.4. de l'article). La réduction des taux de possession automobile mesurée dans ces différentes études varie de 0.14 à 0.20 véhicule par ménage pour l'autopartage basé stations, et de 0.02 à 0.19 véhicule par ménage pour l'autopartage en libre-service. Ces pourcentages suggèrent un impact moins important dans le cas où les véhicules partagés sont en libre-service, ce qui pourrait être expliqué par la plus grande flexibilité du service, le rendant ainsi moins prévisible et donc moins fiable que le système basé stations. Tous types de services confondus, les calculs agrégés réalisés dans les différents travaux recensés par Becker et al. (2018) ont permis d'énoncer que pour chaque véhicule d'autopartage ajouté, le nombre de véhicules particuliers retirés sur les routes était compris entre 3,5 et 23.

Néanmoins, l'ampleur des effets mesurés sur le taux de possession semble dépendre de plusieurs paramètres, tels que l'intensité d'utilisation du service d'autopartage. Les usagers les plus fréquents du service sont ceux qui ont eu les impacts les plus importants sur leur possession automobile. En revanche, être hautement qualifié et avoir un revenu plus élevé est associé au maintien du niveau de possession de la voiture. Ainsi, les ménages dont les ressources économiques sont plus limitées sont plus susceptibles d'utiliser le système pour remplacer (et non augmenter) l'utilisation d'une voiture privée (Le Vine & Polak, 2019). L'impact varie également en fonction du lieu de domicile et de l'environnement bâti : les plus grandes diminutions du taux de possession automobile sont

observées dans les quartiers urbains denses (Clewlow, 2016), là où d'autres alternatives sont également présentes.

En termes d'intensité d'utilisation de l'automobile, plusieurs auteurs ont relevé un nombre de véhicules-kilomètres parcourus et/ou une consommation d'essence plus faible chez les membres que chez les non-membres d'un service d'autopartage (Cervero et al., 2007; Clewlow, 2016; Lane, 2005; Trépanier et al., 2013). Toutefois, cette différence serait significative seulement dans les quartiers à faible densité, c'est-à-dire là où le potentiel de réduction de l'utilisation de l'automobile est le plus grand (Clewlow, 2016). En outre, les plus grandes diminutions ont lieu au cours des 1 à 2 premières années du programme. De 3 à 4 ans après l'inauguration de service, les baisses précédentes semblent se stabiliser (Cervero et al., 2007). De plus, Sioui et al. (2013) ont constaté que quelle qu'était l'intensité d'utilisation du service d'autopartage, la part modale automobile observée (en déplacements) chez les membres de Communauto n'atteignait jamais celle des ménages similaires (avec la même composition) de la population possédant une ou plusieurs voitures. Ces derniers résultats laissent donc présager un usage plus efficace de l'automobile et l'utilisation de modes alternatifs chez les abonnés à l'autopartage.

2.2.1.3.2 Impact de l'autopartage sur les modes durables

Les conclusions tirées au sujet de l'impact de l'autopartage sur les modes durables (transport en commun et transport actifs) ne sont pas aussi unanimes que celles rapportées précédemment dans le cas de l'automobile. Des plus grandes parts modales (en déplacements) pour le transport en commun, le vélo et la marche ont été enregistrées chez les membres que chez les non-membres d'un service d'autopartage (Clewlow, 2016). Toutefois, pour le même service (City CarShare, San Francisco), une partie non négligeable des répondants à l'enquête de Cervero et al. (2007) ont déclaré que leurs déplacements auraient été faits en transport collectif si le service d'autopartage n'avait pas été disponible. Ce taux de substitution est néanmoins souvent plus faible (5 fois inférieur à Turin, Italie) que celui des automobiles privées (Ceccato et al., 2021). Globalement, une diminution de l'utilisation du transport en commun (tant le métro que le bus) a été mise en évidence dans plusieurs enquêtes, mais les deux directions (augmentation et diminution) sont observées dans les données désagrégées (Martin & Shaheen, 2011; Trépanier et al., 2013). En outre, la possession d'un titre de transport en commun ferait augmenter les chances d'être membre du service d'autopartage, mais il ferait diminuer la probabilité de l'utiliser (Becker et al., 2017).

La direction des impacts mesurés au niveau individuel semble être liée aux changements dans la possession automobile. D'un côté, l'abonnement au service d'autopartage a donné accès à un véhicule et à un mode supplémentaire à de (nombreux) ménages non motorisés. De l'autre côté, l'autopartage a permis à des ménages qui possédaient déjà une (ou plusieurs) automobiles de s'en départir ou de moins l'utiliser. Ainsi, les membres qui ont réduit leur possession automobile tendent à emprunter davantage le transport en commun, alors que les membres qui ont gagné un accès à la voiture grâce à l'autopartage ont diminué leur utilisation du transport en commun et ont augmenté leur kilométrage parcouru en automobile (Lane, 2005; Martin & Shaheen, 2011).

L'autopartage peut concurrencer le transport en commun en offrant une alternative souvent plus rapide (Martin & Shaheen, 2011) ou en fournissant le confort et la flexibilité qu'un conducteur aurait avec une voiture privée (Silvestri et al., 2021). Cependant, il peut aussi le compléter à des périodes ou endroits où le transport collectif est moins fréquent ou moins développé, pour des déplacements qui sont moins appropriés en transport en commun (motif magasinage, mauvaises conditions météo, etc.), ou comme un moyen de relier les nœuds du réseau de transport en commun à l'origine et/ou à la destination finale des utilisateurs (Ceccato & Diana, 2021). La majorité des déplacements en autopartage sont en effet réalisés pendant les heures creuses, pour des trajets relativement longs à l'intérieur de la banlieue (Ye et al., 2022).

L'interaction entre l'autopartage et le transport en commun dépend également du type de service d'autopartage. L'autopartage basé stations n'étant généralement pas utilisé pour faire des déplacements quotidiens, il a plus de chances d'être complémentaire au transport collectif et d'être utilisé pour des motifs spécifiques. À l'inverse, l'autopartage en libre-service est plus susceptible de remplacer le transport en commun dans des déplacements réguliers (motifs travail/études) ou pour se rendre à l'aéroport car il permet des déplacements aller seulement (*one-way trip*) (Becker et al., 2017; Le Vine et al., 2014; Papu Carrone et al., 2020). Néanmoins, cette plus grande flexibilité en fait aussi une meilleure option pour servir le premier et/ou le dernier kilomètre d'un déplacement en transport en commun.

Concernant les transports actifs, leur usage tendrait à augmenter après l'inscription à un service d'autopartage (Martin & Shaheen, 2011; Trépanier et al., 2013). C'est notamment le cas pour la marche, qui est le mode le plus utilisé pour accéder aux stations d'autopartage (Cervero et al., 2007; Ye et al., 2022). Une forte complémentarité entre le vélopartage et l'autopartage a également été

mesurée (Ceccato & Diana, 2021). Finalement, en combinant tous les modes de transport durables ensemble, il y aurait globalement plus de personnes qui ont augmenté (plutôt que diminué) leur utilisation du transport en commun et des modes actifs après s'être abonnées à l'autopartage (Martin & Shaheen, 2011).

2.2.1.4 Impact des TNC (transportation network companies)

Malgré un manque de données souvent signalé (Graehler Jr et al., 2019; Hall et al., 2018; Henao & Marshall, 2019), de nombreuses recherches ont été menées très récemment pour analyser l'impact des entreprises de transport technologiques (*transportation network companies* ou TNC en anglais, aussi connues sous le nom de services de *ridesourcing* ou *ridehailing*) telles que Uber ou Lyft sur l'usage des autres modes de transport.

2.2.1.4.1 Impact des TNC sur le taxi traditionnel et l'automobile

La littérature existante s'est d'abord penchée sur les interactions entre les TNC et le taxi du fait de la similarité des deux services. Ce nouveau type de transport rémunéré de personnes par automobile a sans aucun doute eu un impact négatif sur la demande des taxis traditionnels. En effet, l'augmentation de la popularité et de l'utilisation des TNC coïncide généralement avec une forte diminution des courses en taxi (Contreras & Paz, 2018; Nie, 2017; Roy et al., 2020; Schaller, 2017; Wallsten, 2015). De plus, dans les enquêtes, le taxi fait partie des premiers modes substitués par les TNC - voir le tableau 4 de la revue de littérature faite par Tirachini (2020). Cette forte concurrence entre les deux modes peut être expliquée par plusieurs avantages qu'ont les TNC sur le taxi : les prix sont généralement moins élevés, les temps d'attente sont plus courts, et ils offrent une meilleure expérience client grâce à l'utilisation d'une application mobile (Hall et al., 2018; Rayle et al., 2016; Roy et al., 2020).

Les répercussions sur l'automobile privée sont plus divergentes. Dans plusieurs enquêtes, la possession automobile des utilisateurs de TNC n'a pas été impactée significativement, sauf lorsque les TNC sont empruntés très régulièrement, ce qui est rarement le cas (Clewlow & Mishra, 2017; Rayle et al., 2016). À l'échelle d'une ville, à la fois des diminutions (Feigon & Murphy, 2018) et des augmentations (Ward et al., 2021) du nombre d'automobiles ont été observées, mais une absence d'effets a aussi été signalée (Diao et al., 2021). Ces changements systémiques sont les résultats d'un équilibre entre les nouveaux achats réalisés par les conducteurs de TNC et les achats

reportés/annulés ou les ventes d'anciens véhicules par les passagers des TNC. Les impacts devront être examinés à plus long terme pour que la situation se stabilise (Rayle et al., 2016).

En général, les propriétaires de voitures ont tendance à moins conduire leur propre voiture quand ils adoptent les services de TNC (Clewlow & Mishra, 2017; Rayle et al., 2016). Les TNC peuvent parfois remplacer l'utilisation de véhicules particuliers, notamment lorsque le stationnement est un enjeu ou pour éviter de conduire après avoir bu de l'alcool (Clewlow & Mishra, 2017). Cependant, étant donné que les TNC ajoutent des déplacements sur les routes (nouvelles destinations accessibles, conducteurs en attente de clients, etc.), le bilan est plutôt une hausse de l'ensemble des kilomètres parcourus en véhicule motorisé (Erhardt et al., 2019; Henao & Marshall, 2019; Schaller, 2017; Tirachini & Gomez-Lobo, 2019). En effet, les kilomètres supplémentaires dus aux chauffeurs circulant à vide entre deux clients (ou entre leur lieu de domicile et le premier/dernier client) pourraient augmenter de plus de 40% les kilomètres parcourus (Henao & Marshall, 2019). Des kilomètres sont également ajoutés si les TNC entraînent des transferts modaux depuis des modes plus durables (voir section suivante). De plus, la grande disponibilité du service crée de nouveaux déplacements (demande induite) : environ 10% des personnes enquêtées ont déclaré qu'elles ne se seraient pas déplacées en l'absence des TNC (Henao & Marshall, 2019; Rayle et al., 2016). Un moyen de faire diminuer ce kilométrage est le partage des courses via des services spécifiques comme UberPool, mais les courses partagées représentent environ 20% des déplacements seulement dans les endroits où l'option est disponible (Gehrke et al., 2019; Hou et al., 2020), ce qui ne suffit pas à compenser les véhicules-kilomètres ajoutés (Schaller, 2021).

Cette augmentation du kilométrage engendre inévitablement une hausse des émissions polluantes et de la congestion automobile. Une augmentation des temps de déplacement et une diminution des vitesses ont en effet été mesurées depuis l'arrivée des TNC dans plusieurs villes (Diao et al., 2021; Erhardt et al., 2019; Hall et al., 2018; Qian et al., 2020).

2.2.1.4.2 Impact des TNC sur les modes durables

Les résultats de plusieurs enquêtes indiquent que le transport en commun ou des modes actifs auraient été empruntés si les TNC n'avaient pas été disponibles (Clewlow & Mishra, 2017; Gehrke et al., 2019; Henao & Marshall, 2019; Rayle et al., 2016; Tirachini & Gomez-Lobo, 2019). Néanmoins, l'ampleur de la substitution varie selon le contexte et la ville étudiée (de 15 à 42% pour le transport en commun dans les travaux cités). Encore une fois, les transferts modaux

observés diffèrent selon la possession automobile : les propriétaires de voiture(s) ont plus de chances de substituer les TNC à leur voiture privée ou au taxi, alors que les individus non-propriétaires d'automobile sont plus susceptibles de substituer un mode de transport collectif (Rayle et al., 2016). Une corrélation positive entre l'utilisation des deux modes a toutefois été constatée parmi les ménages moins motorisés (Zhang, Y. & Zhang, 2018).

Dans plusieurs analyses longitudinales long-terme réalisées pour identifier les déterminants permettant d'expliquer la diminution ou la stagnation de l'utilisation du transport en commun observée depuis quelques années, l'arrivée des TNC fait partie des principaux facteurs ayant conduit à une baisse d'achalandage (Diab, Kasraian, et al., 2020; Erhardt et al., 2022; Graehler Jr et al., 2019). Toutefois, seule la date d'entrée d'Uber ou de Lyft dans les villes est connue, ce qui ne permet pas d'évaluer l'ampleur de l'impact. En outre, l'impact moyen cache en réalité une grande hétérogénéité : l'effet estimé dépend de la taille de la ville et de l'agence, mais aussi du mode de transport. Ainsi, l'impact des TNC sur l'achalandage total du transport en commun est positif dans les grandes villes (densément peuplées) et les petites agences de transport, mais négatif dans les villes plus petites (moins denses) et les grandes agences de transport (Diab, Kasraian, et al., 2020; Erhardt et al., 2022; Hall et al., 2018). Les TNC tendent à compléter les petites agences car elles ont un réseau de transport collectif moins développé. À l'inverse, l'achalandage du transport en commun est potentiellement plus résilient dans les grandes villes car ces dernières disposent généralement d'un réseau de métro. En effet, la modélisation des impacts a permis de mettre en évidence que l'achalandage des autobus était généralement plus négativement influencé que celui des trains ou des métros urbains (Erhardt et al., 2022; Graehler Jr et al., 2019). Une relation de complémentarité entre les TNC et le rail a même été rapportée dans certaines études (Babar & Burtch, 2020; Li, W. et al., 2022).

Les TNC peuvent être utilisés en remplacement du transport en commun dans certaines situations, en particulier lorsque le service est privilégié pour sa plus grande rapidité et ses temps d'attente moins longs que ceux du transport en commun (Dong et al., 2021; Feigon & Murphy, 2018; Graehler Jr et al., 2019; Young et al., 2020). Ces modes pourraient par exemple être préférés en cas d'urgence. Cependant, les TNC sont le plus souvent utilisés de façon occasionnelle (Clewlow & Mishra, 2017; Feigon & Murphy, 2018; Gehrke et al., 2019), généralement pour des motifs récréatifs (Tirachini, 2020); ils ne constituent donc pas un mode central pour la plupart des utilisateurs enquêtés. Par ailleurs, les TNC peuvent servir de complément à l'extérieur de la

desserte du transport en commun, pour des déplacements dans les régions à faible densité (Hall et al., 2018; Young et al., 2020), ou en dehors des horaires planifiés du service, notamment la nuit (Feigon & Murphy, 2018; Gehrke et al., 2019), période durant laquelle un sentiment d'insécurité peut aussi vouloir faire minimiser l'attente d'une personne. De plus, les TNC peuvent servir dans la connexion avec le transport en commun (Rayle et al., 2016; Yan et al., 2019), notamment grâce à l'application Transit (Davidson et al., 2017). Finalement, les TNC sont parfois présentés comme une solution de désengorgement des réseaux de transport collectif pendant les heures de pointe (Rayle et al., 2016).

2.2.1.5 Impact des véhicules autonomes

La prochaine étape dans le développement du transport automobile est probablement l'arrivée des véhicules autonomes. Quelques auteurs ont déjà proposé des modèles et simulations (souvent basés agents) afin d'appréhender les changements qui seront causés par cette nouvelle forme de mobilité. Avec leur revue de littérature sur le sujet, Soteropoulos et al. (2019) résument que les études recensées associent généralement les véhicules autonomes privés (utilisation individuelle) à un impact moindre sur la possession automobile, à une augmentation du kilométrage automobile parcouru (pour les mêmes raisons que celles évoquées précédemment pour les TNC : déplacements à vide, transfert modal depuis d'autres modes, demande induite) et à une réduction des parts modales du transport en commun et des modes lents. En particulier, ces véhicules attireront de nouvelles clientèles vers l'automobile, notamment les personnes trop jeunes pour conduire, les personnes âgées et les personnes handicapées (Truong et al., 2017).

Cependant, le sens de cet impact est moins clair pour les services de véhicules autonomes partagés, par exemple des taxis autonomes. Ils ont plus de chances de faire baisser l'usage des automobiles privées, mais l'ampleur de cet impact dépendra de leur taux de pénétration dans le marché qui est encore très incertain (Narayanan et al., 2020). L'introduction d'un service de taxis autonomes avec des tarifs potentiellement bas pourrait réduire significativement l'utilisation du transport en commun, notamment dans les zones suburbaines (Bischoff et al., 2019). Toutefois, une analyse de coût a permis de démontrer que le transport en commun devrait rester compétitif dans les zones urbaines denses (Bösch et al., 2018). Par ailleurs, la coordination entre des systèmes de mobilité autonome à la demande et le transport en commun (utilisation de véhicules autonomes partagés pour faire le premier/dernier kilomètre) pourrait générer des avantages significatifs en termes de

temps de déplacement, d'émissions de GES (véhicules-kilomètres) et de coût monétaire par rapport à un système autonome fonctionnant de manière isolée (Salazar et al., 2018; Shen et al., 2018). Toutefois, le nombre de déplacements multimodaux (combinant véhicules autonomes et transport en commun) ne serait peut-être pas suffisant pour compenser le nombre de déplacements en véhicules autonomes qui auraient pu être faits en transport en commun (Gurumurthy et al., 2020). Néanmoins, les auteurs cités reconnaissent que la représentation des changements comportementaux par les simulations proposées reste assez limitée (ces modèles ne prennent pas en compte toute la complexité des phénomènes modélisés et sont basés sur de nombreuses hypothèses). Par conséquent, des recherches supplémentaires sont nécessaires pour préciser le sens des transferts modaux qui seront produits en réponse à la mobilité autonome.

2.2.1.6 Données et méthodes d'évaluation des impacts

2.2.1.6.1 Des enquêtes transversales jusqu'aux données longitudinales

L'Annexe B montre que la grande majorité des recherches réalisées pour comprendre l'impact des modes de transport émergents sur les services classiques sont basées sur des enquêtes transversales. En particulier, presque aucune étude sur les interactions avec l'autopartage ou avec les TNC ne s'appuie sur des données d'achalandage, probablement en raison de la faible accessibilité de ces données (possédées par des entreprises privées). Les enquêtes qui ont uniquement ciblé les membres d'un service, sans groupe contrôle pour comparaison, ont été vivement critiquées (Becker et al., 2018). Toutefois, lorsque des non-membres sont enquêtés, les réponses à certaines questions peuvent parfois être difficiles à interpréter car les personnes interrogées au hasard ne sont pas nécessairement familières au service (Ceccato & Diana, 2021). Dans le cas de l'autopartage, un problème de sous-estimation des membres inactifs peut également être rencontré, ces derniers répondant moins aux enquêtes proposées (Sioui et al., 2013). Néanmoins, lorsque ces membres peu actifs ou inactifs sont inclus dans l'analyse, il peut être compliqué de les écarter et les effets moyens ou agrégés calculés pour les membres sont alors réduits (Martin et al., 2010). Plus globalement, la représentativité des échantillons collectés est un enjeu majeur.

Par ailleurs, la plupart des enquêtes utilisées étant rétrospectives, des biais de réponse sont introduits, en particulier lorsqu'il est demandé de rapporter un kilométrage parcouru dans le passé. Reconnaisant ces limites, certains auteurs se justifient en déclarant que même si l'information

recueillie est sujette à des imprécisions et qu'elle ne permet pas de mesurer avec certitude l'ampleur des changements, elle capture au moins la direction du transfert modal (Martin & Shaheen, 2011). De même, on invite parfois les personnes enquêtées à se projeter (par exemple, concernant l'achat futur ou la vente d'un véhicule) et les changements observés peuvent être bien différents des changements déclarés au départ. Par conséquent, des enquêtes de type panel en plusieurs vagues devraient être privilégiées (Becker et al., 2018).

En outre, Mishra et al. (2019) ont mis en évidence deux importants types de biais dans l'estimation des impacts de l'autopartage par des enquêtes transversales. Premièrement, les résultats peuvent être erronés par un biais d'auto-sélection lié au caractère non aléatoire de l'appartenance au groupe contrôle. Autrement dit, les impacts mesurés peuvent être dus à des différences intrinsèques (observées ou non observées) entre les membres et les non-membres du service (par exemple, le choix de leur lieu de résidence), et non à l'abonnement en soi. Deuxièmement, le vrai effet de causalité peut être confondu avec un biais de simultanéité ou de causalité inversée entre une variable indépendante et la variable dépendante. Par exemple, la possession automobile a une influence sur la décision de s'inscrire à un service d'autopartage, qui à son tour influe sur la possession automobile à une période ultérieure. De même, si les personnes qui utilisent davantage les transports en commun sont plus susceptibles d'adhérer à un service d'autopartage, alors la variable d'intérêt (utilisation des transports en commun) peut également influencer la variable causale (adhésion à l'autopartage). La solution est encore une fois d'utiliser des données longitudinales, qui permettent de confirmer l'ordre des événements.

En plus d'être biaisées, les enquêtes ne permettent généralement que des analyses qualitatives. Les questions posées invitent les répondants à indiquer quel mode a été remplacé par un autre dans la plupart des cas ou pour un déplacement spécifique (par exemple, le dernier), ce qui ne permet qu'un seul type d'interaction (compétition ou complémentarité) par répondant alors que le même répondant peut à la fois combiner et substituer les deux modes dans des conditions différentes. Une personne peut par exemple déclarer qu'elle utilise un certain mode en remplacement du transport en commun pour un déplacement donné, mais si ce mode a permis à la personne de se départir de sa voiture et que ce nouveau mode est utilisé en alternance avec le transport en commun, l'impact réel sera plutôt une augmentation de l'usage du transport en commun (Tirachini, 2020). De plus, aucune information quantitative sur la fréquence de l'interaction n'est recueillie. Au mieux, des intensités d'utilisation déclarées rétrospectivement sont compilées. Par conséquent, la tendance

générale estimée au niveau du répondant peut ne pas refléter l'impact sur l'achalandage total (Luo, H. et al., 2021). Des tentatives de généralisation des résultats d'une enquête (échantillon) à toute la population ont été faites, mais elles reposent généralement sur des hypothèses très simplistes, impliquant par exemple des distances de déplacement constantes entre les modes de transport échangés par les individus ayant transféré d'un mode à un autre (Fishman et al., 2014; Martin et al., 2010). De surcroît, les enquêtes ne contiennent pas d'informations assez précises pour étudier la variabilité des interactions entre les modes dans le temps et dans l'espace.

Toutes ces raisons expliquent pourquoi les études les plus récentes se basent sur des données opérationnelles longitudinales (voir tous les exemples de l'Annexe B). Celles-ci sont notamment exploitées pour évaluer des effets au niveau systémique sur l'achalandage. Toutefois, cet impact global est la somme de différents effets individuels (des relations de concurrence et de complémentarité peuvent coexister au même endroit), et il ne tient pas compte des interactions qui n'influencent pas nécessairement l'achalandage à court terme (par exemple, lorsque d'autres modes complètent le transport en commun en dehors des heures planifiées ou de la zone de service, il n'y a pas d'impact direct sur l'utilisation du transport en commun) (Luo, H. et al., 2021). De plus, des corrélations positives ou négatives sont mesurées sans vérifier si la substitution ou la complémentarité des déplacements est vraiment possible (faisabilité spatio-temporelle). Seule la proximité spatiale entre les modes est considérée, sans tenir compte des horaires planifiés des services de transport en commun (Song, Ying & Huang, 2020). Par conséquent, les interactions ont été récemment explorées au niveau du déplacement à partir de données passives désagrégées (Kong et al., 2020; Luo, H. et al., 2021; Song, Ying & Huang, 2020). Pour chaque déplacement en vélopartage ou en trottinette, les auteurs ont cherché à évaluer s'il s'agissait d'une potentielle situation de complémentarité ou de concurrence avec le transport en commun en fonction de la faisabilité spatio-temporelle et/ou de la qualité de l'alternative en transport en commun.

2.2.1.6.2 Des statistiques descriptives jusqu'aux modèles d'inférence causale

Ce changement de type de données exploitées a également été accompagné d'une évolution des méthodes utilisées pour évaluer les impacts. La plupart des résultats d'enquêtes sont rapportés sous la forme d'une analyse descriptive des réponses, visualisés à l'aide de graphiques et parfois validés avec des tests statistiques (Clewlow, 2016; Martin et al., 2010; Shaheen et al., 2013). Au mieux, un indicateur de transfert modal a été proposé (Fuller et al., 2013), et des modèles de régression

logistique binaire ou ordinaire ont été développés pour modéliser, au niveau individuel, l'existence d'un impact [modification ou non du taux de possession automobile] (Le Vine & Polak, 2019), la direction du transfert modal observé [augmentation ou diminution de l'utilisation du transport en commun] (Martin & Shaheen, 2014), le type d'impact causé [dissuasion à l'achat ou vente/cession d'un véhicule privé] (Le Vine & Polak, 2019) ou encore l'ampleur de cet impact [réduction de 2 autos ou plus, réduction de 1 auto, pas de changement, augmentation de 1 auto, augmentation de 2 autos ou plus] (Cervero et al., 2007). Toutefois, ces modèles désagrégés restent basés sur des réponses déclarées.

Avec les données passives, des méthodes statistiques permettant de quantifier l'ampleur des impacts systémiques et de contrôler l'influence d'autres facteurs contextuels indépendants ont été appliquées. En particulier, plusieurs modèles proposés dans la littérature s'appuient sur des séries temporelles. Les plus basiques sont des régressions linéaires où la variable dépendante représente l'achalandage du mode de transport impacté par l'arrivée d'un nouveau mode (Ma, T. & Knaap, 2019; Ma, T. et al., 2015). Des modèles à effets aléatoires et/ou fixes (ou modèles multiniveaux) ont également été développés (Erhardt et al., 2022; Graehler Jr et al., 2019; Ziedan et al., 2021). Comparés aux régressions traditionnelles, ces modèles ont l'avantage de pouvoir considérer simultanément les données de plusieurs villes (ou autres zones d'agrégation spatiale) en tenant compte de l'autocorrélation des données et en réduisant les biais liés à l'omission de variables. Des précisions seront apportées sur ces modèles et d'autres méthodes de modélisation longitudinale basées sur des séries temporelles seront répertoriées plus tard dans cette revue (section 2.3.2.1).

Dans ces approches longitudinales, l'effet des modes émergents est souvent modélisé par l'intermédiaire d'une variable liée à l'offre, indiquant l'existence ou non du service, par exemple une variable binaire rapportant la présence ou non d'un système ou d'une station de vélopartage (Graehler Jr et al., 2019; Ma, T. & Knaap, 2019). Cette variable représente un effet immédiat et permanent à partir de son apparition (c'est-à-dire à partir de l'instant où la variable binaire devient égale à 1) et suppose un effet constant dans le temps. Dans le cas des TNC, le nombre d'années depuis l'implantation du service est parfois considéré (Erhardt et al., 2022; Graehler Jr et al., 2019), mais cette variable présume un effet linéaire dans le temps. Étant donnée la forte croissance de l'achalandage des modes émergents, le nombre de déplacements devrait être modélisé via une variable continue. De même, les changements dans l'offre devraient être inclus du fait de leur rapide évolution (Graehler Jr et al., 2019). L'indisponibilité des données sur l'offre et la demande

des nouveaux modes est la principale cause de la non prise en compte de ces phénomènes. Par ailleurs, une limite majeure des modèles longitudinaux est qu'ils permettent seulement de modéliser des corrélations et non des effets de causalité (Hyndman & Athanasopoulos, 2018).

Néanmoins, dans les dernières publications sur le sujet, des techniques statistiques spécifiquement conçues pour faire de l'inférence causale sont souvent utilisées. Ces méthodes permettent d'évaluer l'impact causal d'un « traitement », d'une politique ou d'une intervention sur une variable d'intérêt en comparant un groupe « traité » à un groupe témoin. Elles sont employées dans le cas où on dispose d'observations ou de données quasi-expérimentales (c'est-à-dire lorsque le traitement ou l'intervention n'est pas attribué au hasard). Ces méthodes servent ainsi à réduire les biais dus au caractère non aléatoire de l'expérience dans l'estimation de l'effet du traitement. Trois types d'impact peuvent être mesurés : l'effet moyen du traitement sur un individu choisi de manière aléatoire (ATE - *average treatment effect*), l'effet moyen du traitement sur un individu traité (ATT - *average treatment effect for the treated*), ou l'effet moyen du traitement sur un individu non traité (ATC - *average treatment effect on the controls*), l'effet ATT étant le plus couramment étudié dans la littérature (Mishra et al., 2019). Dans un contexte de transport, la relation de causalité mesurée peut par exemple être l'effet de l'abonnement à l'autopartage sur la possession automobile (Becker et al., 2018) ou l'effet de la présence de stations de vélopartage sur la congestion routière (Hamilton & Wichman, 2018). Dans le premier cas, la méthode utilisée permet de corriger partiellement le biais d'auto-sélection des membres tel que discuté précédemment. Dans le deuxième cas, elle permet de tenir compte de la localisation non aléatoire des stations de vélopartage. Dans la suite, le terme « individu » est utilisé (en parallèle avec la notion de « traitement »), mais d'autres objets peuvent être étudiés, par exemple des unités d'analyse spatiale.

Les méthodes les plus anciennes de la recherche sur l'inférence causale sont les méthodes d'appariement, employées depuis longtemps dans diverses disciplines (Stuart, 2010). Elles consistent à faire correspondre les individus traités à des individus non traités ayant les mêmes caractéristiques observées (la distribution des covariables doit être semblable dans les deux groupes), de sorte que la différence entre les deux groupes de la variable dépendante étudiée soit seulement due au fait d'être traité ou non. Cependant, quand le nombre de variables déterminant la sélection augmente, il devient de plus en plus difficile de trouver des individus comparables. C'est pourquoi le score de propension a été inventé, donnant naissance à la méthode d'appariement par score de propension (*propensity score matching*), utilisé entre autres par Hamilton et Wichman

(2018) et Mishra et al. (2019) dans un contexte de transport. Cette méthode permet d'apparier les individus à partir de leur similarité sur un seul indice (le score de propension), traduisant la probabilité d'un individu d'avoir été traité à partir des informations contenues dans toutes les covariables qui caractérisent la population étudiée. Ce score de propension est donc exprimé comme un modèle de régression logistique binaire, conditionnel aux caractéristiques des individus (Centre for Research on Impact Evaluation, 2015). Sur la base de cet indice, l'appariement peut alors se faire avec différentes méthodes, la plus connue étant l'algorithme des plus proches voisins. Différents types de distances peuvent être appliqués et chaque individu traité peut être apparié à un ou plusieurs individus non-traité(s) selon les paramètres choisis. Finalement, l'effet moyen ATT est calculé comme la différence moyenne entre les individus traités et non-traités appariés.

Une alternative à la méthode d'appariement est la méthode de pondération par score de propension inverse (*inverse propensity score weighting*) qui consiste, dans chaque groupe, à donner des poids plus élevés aux individus sous-représentés (par rapport aux covariables considérées) et, inversement, des poids plus faibles aux individus surreprésentés. Ces poids sont en fait l'inverse du score de propension, et l'effet moyen pondéré obtenu est l'ATE (Austin & Stuart, 2015). L'avantage de la méthode de pondération par rapport à la méthode d'appariement est qu'elle permet d'inclure tous les individus (aucun individu n'est exclu parce qu'il ne peut pas être apparié à un individu de l'autre groupe de traitement), ce qui est particulièrement intéressant pour des échantillons de plus petite taille.

Ces méthodes possèdent des avantages indéniables sur la régression linéaire (en particulier, elles assurent la comparabilité des individus comparés et elles permettent de s'intéresser plus directement au phénomène causal mesuré avec moins de paramètres à estimer). Cependant, leur plus gros défaut est lié à l'hypothèse très forte d'indépendance conditionnelle, selon laquelle toutes les variables influençant à la fois le fait d'être traité et la variable dépendante (variables confondantes, source d'endogénéité) sont connues. Si certaines caractéristiques ne sont pas observées, il est possible d'utiliser des variables dites instrumentales (corrélées à la variable non observée sans être corrélées au terme d'erreur non expliqué), mais celles-ci peuvent être difficiles à trouver. C'est pourquoi la méthode des doubles différences (DID : *difference-in-differences*) a été développée (Centre for Research on Impact Evaluation, 2015; Lechner, 2011). Comme son nom l'indique, cette méthode consiste à modéliser des doubles différences : entre les deux groupes (individus traités versus non traités) et entre deux périodes temporelles (avant et après traitement).

Plus précisément, la différence entre les deux groupes avant traitement (ou avant la mise en place d'une intervention ou l'entrée en vigueur d'une politique) est soustraite de la différence entre les deux groupes après traitement. La différence entre ces deux différences est l'effet estimé du traitement (ATT).

Ainsi, contrairement aux modèles de séries temporelles expliqués plus haut, qui mesurent uniquement des différences dans le temps, et les méthodes transversales ou d'appariement, qui prennent seulement en compte les différences entre le groupe traité et le groupe contrôle, la méthode des doubles différences considère ces deux types de différences dans le même modèle. Elle supprime donc les biais causés par les différences permanentes entre les deux groupes (contrôle et traitement), ainsi que les biais causés par les tendances temporelles affectant l'ensemble de l'échantillon. Des données avant/après sont nécessaires pour l'appliquer (par exemple, avant et après l'implantation d'un nouveau service de mobilité), d'où l'utilisation de données de panel ou de données longitudinales. Cette méthode a été utilisée dans un contexte de transport dans quelques-uns des travaux cités précédemment, soit avec un modèle de base mesurant des différences entre deux groupes [membres versus non-membres] et deux périodes [un point avant et un point après] (Becker et al., 2018), soit avec un modèle généralisé à plusieurs groupes et plusieurs périodes (Campbell & Brakewood, 2017; Hall et al., 2018; Luo, H. et al., 2021).

Toutefois, la méthode suppose que les attributs qui différencient le groupe contrôle du groupe traité soient constants dans le temps. De plus, elle contrôle seulement pour les changements survenus au cours du temps qui ont eu le même impact sur les deux groupes. Cela signifie qu'en l'absence de traitement, la variable d'intérêt (variable dépendante) devrait suivre la même tendance pour les deux groupes (courbes parallèles), ce qui est impossible lorsque de trop grandes différences existent entre les individus. Pour limiter cet enjeu, la méthode d'appariement par score de propension et la méthode DID sont parfois combinées (Ma, Xiaolei et al., 2019; Wang, Mingshu & Zhou, 2017), c'est-à-dire que le modèle DID est appliqué sur les individus appariés uniquement. Il existe également une méthode de pondération des quatre groupes impliqués dans la méthode DID [groupe traité avant traitement, groupe traité après traitement, groupe contrôle avant traitement, groupe contrôle après traitement] (Stuart et al., 2014). Par ailleurs, ces méthodes permettent seulement d'évaluer l'interaction entre deux phénomènes (ou entre deux modes dans le cas spécifique du transport). Dans un contexte de multimodalité, les effets entre plusieurs modes de transport devraient pouvoir être inclus dans un même modèle.

2.2.2 Étude de la multimodalité

La multimodalité, concept impliquant la prise en compte simultanée de l'ensemble des modes de transport, est un thème qui reçoit de plus en plus d'attention dans le domaine de la mobilité urbaine (voir le contexte et la définition de toutes les notions associées dans le Chapitre 1 Introduction). Toutefois, la littérature scientifique sur ce sujet est limitée en raison d'un manque de données qui soient à la fois multimodales et longitudinales (ou au moins disponibles sur plusieurs jours).

Historiquement, dans les enquêtes de déplacements réalisées, peu d'informations étaient collectées sur la multimodalité. Jusqu'en 1980-1990, ces enquêtes étaient centrées sur l'automobile et le transport en commun, les déplacements effectués avec des modes non motorisés n'étant souvent même pas retranscrits. De même, elles ne comprenaient généralement pas de questions sur la séquence de modes empruntés, et encore moins sur les segments de déplacements en direction (*access travel*) ou en provenance (*egress travel*) du ou des modes principaux utilisés (Clifton & Muhs, 2012). Même lorsque cette information est connue, beaucoup d'enquêtes restent encore basées sur le concept de « mode de transport principal », où chaque déplacement est représenté par un seul mode central (Kagerbauer et al., 2015). De plus, des questions se posent toujours sur la façon de collecter et de définir les déplacements multimodaux : comme chaque déplacement commence et se termine à pied, des seuils (en temps ou en distance) doivent généralement être fixés afin de déterminer si la marche peut être considérée comme un mode à part entière dans une séquence de déplacements (Gebhardt et al., 2016).

Par ailleurs, les comportements de mobilité doivent être suivis sur plusieurs jours de données pour pouvoir étudier la multimodalité des individus, c'est-à-dire l'utilisation de différents modes de transport dans le temps. Kuhnimhof et al. (2006) justifient par exemple la nécessité de disposer d'informations sur au moins 7 jours. Dans le cas où ces données ne seraient pas disponibles, il faut au moins détenir les réponses à des questions rétrospectives sur les fréquences d'utilisation de plusieurs modes pendant une période donnée (Diana & Pirra, 2016). Toutefois, les enquêtes de déplacements traditionnelles ne collectent pas ces informations. Au contraire, les données passives sont longitudinales et disponibles sur de longues périodes. Cependant, ces données sont la plupart du temps unimodales, c'est-à-dire qu'elles ne contiennent des informations que sur un seul mode, et il n'existe pas de moyen direct pour les relier.

D'après Spickermann et al. (2014), l'expansion de « la mobilité multimodale suppose trois conditions : la présence de plusieurs modes de transport adaptés à chaque situation spécifique ; des informations suffisantes pour pouvoir choisir facilement entre plusieurs options en temps réel (notamment avec le support de la technologie mobile) ; et un changement dans les habitudes et les comportements sociaux » [notre traduction]. Sur ce dernier point, quelques recherches ont été réalisées afin de mieux comprendre la mobilité multimodale (section 2.2.2.1). De plus, un besoin d'intégration et de coopération entre les modes a été souligné, que ce soit sous la forme de tarifs intégrés, d'informations voyageurs ou d'infrastructures intermodales (Gebhardt et al., 2016). Pour y parvenir, Spickermann et al. (2014) insistent sur deux grandes nécessités : celle de développer plus de collaborations entre les différents opérateurs du transport, et celle de guider les villes dans l'élaboration d'une vision à long terme par le développement d'une stratégie de planification intégrée qui s'oriente vers un système multimodal et durable. Cependant, les méthodes actuelles d'analyse des déplacements ne semblent pas tout à fait aptes à informer de telles politiques et des indicateurs de mesure de la performance des villes en matière de multimodalité doivent être développés (Diana & Pirra, 2016). Quelques enjeux et outils liés à ce besoin de quantification de la multimodalité sont donc rapportés dans la section 2.2.2.2.

2.2.2.1 Analyse de la mobilité multimodale

Au niveau individuel, Diana et Pirra (2016) rappellent que le concept de mobilité multimodale peut être appliqué aux déplacements, lorsque plusieurs modes sont utilisés pour faire un même trajet (intermodalité), mais également aux personnes, définissant ainsi la multimodalité comme l'utilisation de plusieurs moyens de transport par un individu sur une période donnée. C'est sur cette deuxième définition que la plupart des recherches ont porté, en vue de mieux comprendre les comportements multimodaux (étape préliminaire nécessaire pour modéliser la demande multimodale, discutée plus tard dans cette revue de littérature). Au niveau systémique, très peu de travaux ont été réalisés sur la demande multimodale à cause d'un manque de données adéquates.

2.2.2.1.1 Au niveau individuel : analyse de la multimodalité des personnes

Dans un contexte où l'accès à l'automobile augmente et où les options de mobilité se diversifient, comment les individus se comportent-ils et quels sont leurs choix de mobilité ? Quelques auteurs ont tenté de répondre à cette question d'actualité. En particulier, plusieurs travaux se sont basés sur des enquêtes nationales allemandes : la MOP (Mobilitätspanel), enquête par panel sur 7 jours, et/ou

la MiD (Mobilität in Deutschland), enquête transversale sur une seule journée mais comprenant des questions sur la fréquence d'utilisation de chaque mode de transport. Dans ces études, la multimodalité des individus est définie comme l'utilisation de plus de deux modes de transport en une semaine, les modes de transport considérés étant l'automobile, le transport en commun et le vélo. Ainsi, parmi les 49% d'utilisateurs multimodaux allemands recensés par Nobis (2007), environ 38% ont utilisé la voiture et le transport collectif en complémentarité, 35% la voiture et le vélo, 5% le transport collectif et le vélo, et 21% des personnes ont emprunté les trois modes (voiture, transport en commun, vélo) pendant une semaine moyenne. Plus spécifiquement, les comportements multimodaux des automobilistes (Kuhnimhof et al., 2006) et des cyclistes (Kuhnimhof et al., 2010) ont été examinés. Ces travaux ont révélé que les automobilistes multimodaux avaient une utilisation occasionnelle du transport collectif, pour des motifs spécifiques et lorsque le transport en commun est une meilleure option que la voiture, alors que les cyclistes empruntaient d'autres modes de manière beaucoup plus fréquente (pour 70% de leurs déplacements durant une semaine moyenne). Le vélo est préféré pour des courtes distances mais l'automobile est le mode privilégié pour des déplacements motif magasinage ou pour pouvoir se déplacer à plusieurs. Au-delà d'un certain périmètre autour du lieu de domicile, les cyclistes utilisent davantage le transport en commun que les autres voyageurs. Plus récemment, une enquête distincte a été réalisée à Munich et a également permis de montrer que les usagers des systèmes de vélopartage et d'autopartage empruntaient une grande variété de modes différents pendant un mois, le transport en commun jouant un rôle central dans la plupart des cas (Miramontes et al., 2017).

Outre ces études désagrégées servant à mieux comprendre les comportements individuels d'utilisation de différents modes de transport, des recherches ont également porté sur l'évolution du niveau de multimodalité (mesurée par le pourcentage de personnes multimodales). Des analyses longitudinales et générationnelles (Kuhnimhof et al., 2011; Kuhnimhof et al., 2012) ont notamment permis de détecter une réduction de la dépendance à l'automobile chez les jeunes en Allemagne et en Grande Bretagne, en partie due à des comportements de plus en plus multimodaux qui reposent sur la combinaison de modes alternatifs à la voiture. Le potentiel croissant de la multimodalité face à l'usage exclusif de l'automobile privée dans la mobilité quotidienne allemande (au moins en zones urbaines denses) a également été confirmé par Schuppan et al. (2014). Dans le contexte américain, l'enquête nationale NHTS révèle des tendances similaires (Buehler & Hamre, 2014, 2015) : la part d'utilisateurs multimodaux a significativement augmenté entre 2001 et 2009.

Finalement, ces études s'accordent pour énoncer un changement de paradigme dans la mobilité quotidienne : la voiture n'est plus considérée comme une nécessité absolue pour se déplacer mais plutôt comme une option parmi d'autres modes de transport urbains. Toutefois, bien que l'acceptation des modes alternatifs grandisse et que les mentalités changent, l'automobile reste encore souvent perçue comme plus confortable ou plus appropriée dans certaines situations (Schuppan et al., 2014).

Au cours des dernières années, des méthodes plus évoluées que de simples analyses descriptives ont été développées pour créer des typologies de comportements multimodaux en fonction des modes utilisés et de leur fréquence d'utilisation pendant une période donnée. Une technique exploratoire, l'analyse d'associations (*association analysis*), a été appliquée à l'enquête NHTS 2009 afin de détecter des combinaisons fréquentes de modes qui sont généralement utilisés ensemble (dans le temps) par une même personne (Diana, 2012). Un modèle de classes latentes a également été construit à partir de l'enquête nationale des Pays-Bas (Molin et al., 2016), et l'algorithme de classification des K-moyennes a été utilisé dans les contextes danois (Olafsson et al., 2016) et français (Yin & Leurent, 2022) pour segmenter les usagers enquêtés et analyser leurs caractéristiques. La majorité des types de comportements multimodaux identifiés sont centrés sur un mode (la voiture, le transport en commun ou le vélo) et d'autres modes sont utilisés en complémentarité de manière plus occasionnelle, les cyclistes ayant le plus fort potentiel de multimodalité (il y a peu de cyclistes unimodaux). Les comportements actuels mais aussi parfois les comportements projetés sont inclus dans la classification obtenue afin de distinguer les personnes ouvertes à de nouveaux modes de transports de celles qui sont plus conservatrices dans leurs habitudes de déplacement (Rodriguez Cote & Diana, 2017).

Les résultats de ces dernières recherches (majoritairement européennes) révèlent à chaque fois un nombre important d'usagers multimodaux et une évolution croissante de leur contribution dans la mobilité quotidienne. C'est pourquoi l'adoption d'une approche multimodale dans l'étude des comportements de déplacement devient de plus en plus pertinente, en particulier dans les régions urbaines où les parts modales du transport en commun et des modes non motorisés sont importantes. Cependant, il est parfois nécessaire de se ramener à l'échelle du déplacement (plutôt qu'à celle de l'individu) pour pouvoir se rendre compte de la complémentarité entre les modes : par exemple, le groupe d'usagers du transport en commun identifié par Olafsson et al. (2016) est associé à la fréquence d'utilisation la plus faible du vélo en tant que mode seul, ce qui peut être

maladroitement interprété comme une forte compétition entre les modes alors que le vélo est en fait fréquemment utilisé pour aller ou revenir du réseau de transport en commun.

2.2.2.1.2 *Au niveau individuel : analyse de la multimodalité des déplacements*

Les études sur le sujet de la multimodalité des déplacements (ou de l'intermodalité) sont encore plus anecdotiques et la plupart d'entre elles ont porté sur l'examen des trajets de longue distance (Oostendorp & Gebhardt, 2018). Quelques recherches se sont néanmoins focalisées sur l'analyse de combinaisons de deux modes spécifiques, un mode (*feeder*) ayant alors la fonction d'en alimenter un autre (*feedee*). Généralement, le mode alimenté est le transport en commun, qui joue un rôle central dans la plupart des déplacements multimodaux du fait de sa zone de desserte fixe et limitée qui peut être complétée par d'autres modes. Par exemple, les travaux sur l'association du vélo et du transport en commun dans un même déplacement (Kager et al., 2016; Kuhnimhof et al., 2010; Martens, 2004; Olafsson et al., 2016; Papon et al., 2017) permettent de conclure que le vélo est utilisé en complémentarité du transport en commun dès que la distance s'allonge, tandis que les deux modes sont plus compétitifs pour de courtes distances. La combinaison du transport en commun et de l'automobile a également été analysée, notamment pour modéliser le choix des stations et les besoins en stationnement dans les déplacements *park and ride* [automobile conducteur + transport en commun] ou *kiss and ride* [automobile passager + transport en commun] (Olaru et al., 2014; Weiss & Habib, 2017). Plus récemment, la complémentarité des modes émergents avec le transport en commun dans le premier/dernier kilomètre a été explorée (voir les exemples présentés dans la section 2.2.1), mais souvent de manière très basique à l'aide d'un pourcentage de déplacements intégrant les deux modes.

Toutefois, très peu d'études traitent de l'intermodalité dans la mobilité quotidienne avec plus de deux modes de transport. De même, peu d'informations sont connues sur les raisons, le contexte et les préférences permettant d'expliquer l'utilisation de certaines combinaisons intermodales par rapport à d'autres ou par rapport à des déplacements unimodaux. Pourtant, il a été démontré que l'intermodalité pouvait permettre d'augmenter l'accessibilité à de multiples destinations dans les villes (Oostendorp et al., 2019). L'Allemagne a encore un train d'avance sur le sujet avec quelques travaux qui s'inscrivent dans le cadre de projets visant à développer l'intermodalité entre les modes (par exemple, le projet IeMM - Intermodal eMobility Management). Ces travaux sont basés sur des enquêtes nationales (Gebhardt et al., 2016) ou des enquêtes spécifiquement conçues pour recueillir

des informations sur la mobilité intermodale et les séquences de modes empruntés, collectant des déplacements sur une journée par personne (Kagerbauer et al., 2015) ou des fréquences d'utilisation de différentes combinaisons de modes sur un mois (Oostendorp & Gebhardt, 2018).

Alors qu'environ 83.5% des répondants ont déclaré combiner plusieurs modes pour faire certains déplacements à une fréquence plus ou moins régulière (Oostendorp & Gebhardt, 2018), seulement 17.2% des personnes enquêtées sur une journée typique à Berlin ont réalisé des déplacements multimodaux (représentant 9.4% des déplacements totaux) (Gebhardt et al., 2016). Ce résultat prouve encore une fois l'importance de s'appuyer sur plusieurs jours d'enquêtes. En outre, ces pourcentages varient fortement selon si la marche permettant d'accéder à un mode est considérée comme un mode à part entière et selon si les différents modes de transport en commun (métro, bus, train) sont séparés ou non (Kagerbauer et al., 2015). Les déplacements multimodaux recueillis dans les études allemandes citées sont majoritairement composés de plusieurs modes de transport en commun et ont lieu dans les quartiers connectés bien desservis. Les deux autres principales combinaisons observées, « transport en commun + automobile » et « transport en commun + vélo », sont respectivement plus fréquentes dans les zones décentralisées et les zones urbaines (Oostendorp & Gebhardt, 2018). La majeure partie de ces associations de modes servent pour des déplacements motif travail ou études (Gebhardt et al., 2016; Oostendorp & Gebhardt, 2018).

Finalement, les résultats des travaux sur l'intermodalité au niveau individuel restent très qualitatifs et basés sur des réponses déclarées. La manière de prendre en compte les modes divergent également d'une étude à l'autre. En outre, seuls les modes de transport classiques (automobile, transport en commun, marche et vélo) sont considérés, l'utilisation des modes de transport émergents et partagés n'étant pas rapportée (ou non distinguée de celle des modes classiques) dans les enquêtes réalisées. Alors que l'analyse individuelle de la multimodalité (basée sur des enquêtes) entraîne la considération d'un nombre limité de modes de transport, une analyse systémique (basée sur des données opérationnelles) permettrait d'avoir une vision plus globale de l'ensemble des modes (incluant les nouveaux services de mobilité).

2.2.2.1.3 Au niveau systémique : analyse de la demande multimodale

Tous les travaux cités précédemment (analyse de la mobilité multimodale au niveau individuel) sont basés sur des données d'enquêtes. En effet, les données passives étant souvent indépendantes pour chaque mode de transport (car chaque mode est géré par un opérateur différent), ce sont

seulement des segments de déplacements multimodaux qui sont collectés dans chaque base de données (et non pas les déplacements complets avec tous les modes combinés). Les données de téléphones cellulaires sont une exception, mais les modes doivent être estimés via des méthodes encore en cours de développement (Huang, H. et al., 2019). En outre, comme certaines bases de données passives ne contiennent pas d'informations sur les individus, soit parce que les systèmes de collecte ne le permettent pas (par exemple, des données de comptages), soit parce que les informations sont anonymisées ou cachées dans les fichiers distribués (par souci de confidentialité), il est parfois impossible de relier une personne à ses déplacements et de la suivre sur plusieurs jours. Ainsi, ces données ne permettent pas d'étudier ni l'intermodalité ni la multimodalité des comportements individuels. Au mieux, elles peuvent être exploitées pour analyser la demande multimodale au niveau systémique en croisant l'achalandage de différents modes de transport.

Très peu de chercheurs se sont penchés sur ce genre d'analyse car rares sont ceux qui bénéficient d'assez de données pour examiner simultanément différents systèmes modaux. En général, seulement deux modes (un mode classique et un mode émergent) sont considérés (section 2.2.1). Toutefois, on trouve dans la littérature récente quelques exceptions pour lesquelles des flux de données passives étaient disponibles pour un grand nombre de modes de transport (Morency et al., 2020; Morency et al., 2018; Tang et al., 2021). Les données sont alors combinées (fusion de différentes bases de données passives) au moyen d'une unité spatio-temporelle commune dans laquelle les déplacements (ou segments de déplacements) sont agrégés. Par exemple, les déplacements (ou segments de déplacements) sont sommés par jour et par TAZ - *traffic analysis zone* (Tang et al., 2021) ou par MIAZ - *mobility interaction analysis zone* (Morency et al., 2020).

Ensuite, des corrélations sont mesurées soit entre l'utilisation de plusieurs modes dans le temps (Ma, T. et al., 2015; Morency et al., 2020; Morency et al., 2018; Tang et al., 2021), soit entre les origines (ou destinations) d'un mode et les destinations (ou origines) d'un autre mode (Ma, Xiaolei et al., 2019; Ziedan et al., 2021; Zuniga-Garcia et al., 2022). Dans le premier cas, la corrélation indique si l'intensité d'usage d'un mode tend à augmenter avec celle d'un autre, suggérant qu'au niveau désagrégé les modes sont combinés par les personnes (multimodalité). Selon la longueur de la période étudiée, des tendances à long terme ou des variations saisonnières sont capturées. Dans le deuxième cas, la corrélation est un proxy pour estimer des relations d'intermodalité. Par ailleurs, ces corrélations peuvent être calculées à différents endroits. Ainsi, les interactions sont analysées en fonction de trois dimensions : le temps, l'espace et le mode de transport (Morency et al., 2020).

2.2.2.2 Quantification de la multimodalité

Bien que la multimodalité soit un sujet de recherche présentement « à la mode », la littérature sur la question de sa quantification est extrêmement limitée. De nouveaux outils doivent donc être développés pour pouvoir mesurer et ensuite modéliser la mobilité multimodale (Diana & Pirra, 2016). Cette modélisation passant d'abord par une meilleure compréhension des comportements individuels de mobilité, quelques indicateurs ont été proposés pour quantifier le niveau de multimodalité des personnes (section 2.2.2.2.2). Par ailleurs, des indicateurs doivent aussi être disponibles pour pouvoir évaluer la performance des villes en matière d'offre multimodale et d'intégration des modes de transport (section 2.2.2.2.1).

2.2.2.2.1 Quantification de l'offre multimodale

Des indicateurs doivent être développés afin d'aider les villes à mesurer la quantité et la qualité de leur offre de service multimodale. La quantité d'offre associée à chacun des modes de transport est facilement chiffrable et divers indicateurs sont déjà largement répandus. Après avoir interrogé neuf villes européennes, Diana et al. (2018) ont listé un grand nombre de ces indicateurs (voir le tableau 2 de l'article). De plus, Litman (2019) a fait correspondre plusieurs de ces indicateurs d'offre (*service indicators*) à des indicateurs de sortie (*outcome indicators*) liés à la demande afin de pouvoir évaluer dans quelle mesure les infrastructures mises en place permettent de répondre aux objectifs fixés en matière de multimodalité. Cependant, il n'existe pas encore d'indicateur de performance global permettant de synthétiser toutes ces informations. De plus, le niveau d'intégration de ces différents modes de transport est assez difficile à quantifier et les indicateurs traditionnels d'interconnectivité du réseau ne sont pas suffisants dans un contexte de comodalité où d'autres formes d'intégration doivent être mesurées (de Stasio et al., 2011). Par conséquent, quelques rares études se sont penchées sur la question.

Au-delà de l'intégration spatiale des modes, analysée via la superposition des infrastructures de transport et le calcul de distances ou de temps d'accès entre les modes (Li, X. et al., 2022), plusieurs formes d'intégration ont été définies dans le cadre de recherches sur les systèmes de mobilité intégrée (MaaS). Kamargianni et al. (2016) ont ainsi proposé un indice pour mesurer le niveau d'intégration d'un système MaaS à partir de quatre types de coopération : l'intégration du titre de transport (un seul titre pour accéder à tous les modes de transport), l'intégration du paiement (une seule facture pour l'ensemble des modes utilisés), l'intégration de forfaits de mobilité sur mesure

(un forfait prépayé d'un certain montant, en durée ou en distance, pour une combinaison de services de mobilité) et l'intégration technologique (une seule application ou interface web pour accéder aux informations sur tous les modes). Chaque type d'intégration est alors évaluée séparément puis un score total est calculé en additionnant les scores individuels pour permettre la comparaison de plusieurs villes. De même, s'inspirant des 5 niveaux d'automatisation des véhicules autonomes, des auteurs ont proposé des typologies en plusieurs niveaux d'intégration (Lyons et al., 2019; Sochor et al., 2018). La taxonomie de Lyons et al. (2019), nivelée en fonction des efforts cognitifs à fournir par les utilisateurs pour se déplacer, est basée sur trois types d'intégration : l'intégration opérationnelle (diminution des pénalités de correspondance et amélioration de l'expérience porte-à-porte), l'intégration informationnelle (disponibilité des informations pour la planification et l'exécution du déplacement dans une seule interface) et l'intégration transactionnelle (paiement, réservation et billetterie via une seule interface). Sochor et al. (2018) classent ces trois types d'intégration (niveau 0 – pas d'intégration, niveau 1 – intégration informationnelle, niveau 2 – intégration transactionnelle, niveau 3 – intégration opérationnelle) et ajoutent un quatrième niveau 4 – intégration des enjeux sociétaux, incluant notamment les objectifs des politiques de transport en matière d'écologie, d'équité, d'accessibilité, etc.

2.2.2.2.2 *Quantification de la multimodalité des comportements individuels*

En ce qui concerne la quantification de la multimodalité des comportements individuels, la revue de littérature de Heinen et Chatterjee (2015) pointe les limites des indicateurs actuels. La plupart du temps, la multimodalité des personnes est mesurée à l'aide de simples statistiques telles que, comme on l'a vu précédemment, le pourcentage de voyageurs qui utilisent un nombre donné de moyens de transport. La fréquence d'utilisation de chaque mode n'est souvent pas considérée, ou bien seulement par l'application d'un seuil maximal. Par exemple, dans les résultats de Nobis (2007), chaque mode utilisé par un individu défini comme multimodal ne peut pas représenter plus de 70% de tous ses déplacements, de sorte à éviter un usage quasi exclusif d'un seul des modes.

En outre, l'article de Diana et Pirra (2016) sur la mesure de la multimodalité au niveau individuel est particulièrement intéressant et fait ressortir les principaux défis associés à cette question. Par conséquent, la suite de cette section résume la discussion très instructive de l'article. Les deux chercheurs proposent de développer un indicateur de multimodalité à partir d'une série d'intensités d'utilisation de différents modes de transport. Pour y parvenir, les auteurs commencent par

déterminer les propriétés que cet indicateur devrait avoir, ce qui implique différents enjeux et des choix à faire au niveau de la définition même de la multimodalité.

Tout d'abord, l'indicateur doit considérer la diversité de l'ensemble des modes utilisés par l'individu (nombre de modes différents), mais aussi la distribution des intensités d'utilisation de chaque mode. Pour illustrer cette condition, les auteurs prennent l'exemple d'un individu qui conduit une automobile et prend le bus 5 fois par semaine (2 modes avec une intensité de 5), versus un individu qui prend sa voiture chaque jour, mais utilise le bus une fois par semaine et le vélo une fois par semaine également (3 modes avec des intensités respectives de 7, 1 et 1). Le comportement le plus multimodal dépendra alors de la sensibilité que l'on veut accorder à chacun de deux aspects suivants : le nombre de modes différents (facteur qui fait augmenter la multimodalité) et la prédominance d'un mode (facteur qui fait diminuer la multimodalité).

Toutefois, la fréquence d'utilisation des modes est associée à un enjeu majeur. En effet, l'intensité d'usage d'un mode peut être exprimée dans différentes unités (en nombre de déplacements, en distance, en durée, etc.) qui conduiront nécessairement à des résultats différents puisque les modes de transport ont des vitesses différentes et sont habituellement empruntés pour des types de déplacements différents (contraints ou non) réalisés à des fréquences différentes. D'après les deux auteurs, le choix de l'unité doit dépendre du contexte. Par exemple, pour des problèmes impliquant un point de vue subjectif, on se basera plutôt sur les durées de déplacement car la perte de temps est ce qui est le plus mal perçue par les individus. Si on veut en revanche regarder les impacts du transport sur l'environnement, on prendra les distances.

Un autre enjeu abordé par les auteurs est lié aux similarités entre les types de modes de transport utilisés. Doit-on considérer chaque mode séparément, ou doit-on prendre en compte les points communs qu'il y a entre certains modes ? Le cas échéant, quelle est l'agrégation ou la classification de modes la plus appropriée ? Par exemple, faut-il considérer tous les modes de transport en commun (bus, métro, train, etc.) ensemble ou bien les distinguer ? Ces questions réfèrent au même enjeu d'indépendance des alternatives que dans les modèles de choix modal. Par ailleurs, les chercheurs évoquent la possibilité de considérer la fréquence de combinaisons de modes plutôt que de modes séparés, voire de donner un poids différent selon la combinaison : un déplacement automobile + TC pourrait alors être considéré comme plus « multimodal » qu'un déplacement

marche + TC. Aussi, bien que cette idée ne soit pas proposée dans l'article, il serait possible de donner un poids différent à chaque mode selon son niveau de durabilité.

En résumé, d'après la réflexion de Diana et Pirra (2016), la définition d'un indice de multimodalité individuelle devrait tenir compte des aspects suivants :

- le nombre de modes utilisés;
- les différences dans les intensités d'utilisation de différents modes;
- les valeurs absolues des intensités d'utilisation d'un ou plusieurs modes (par exemple, l'intensité d'utilisation du mode le plus utilisé);
- les types de modes utilisés.

Les trois premiers aspects sont analogues à trois concepts issus de la diversité des espèces en biologie (richesse, distribution et dominance), tandis que le quatrième implique une classification des modes. Les enjeux associés n'ont pas encore été traités dans la littérature et devront faire l'objet d'analyses de sensibilité.

Cette question étant nouvelle dans le domaine du transport, les auteurs font une revue de différents indices utilisés dans d'autres disciplines (par exemple, en économie et en biologie) et servant à mesurer l'inégalité, la diversité ou, plus globalement, l'hétérogénéité de différentes valeurs (par exemple, l'inégalité des richesses ou la diversité des espèces). D'un point de vue mathématique, c'est la même notion que l'on cherche à mesurer quand on s'intéresse à la multimodalité, c'est-à-dire la variabilité de l'utilisation de différents modes de transport. Les indicateurs candidats retenus par les chercheurs sont les suivants : l'indice de Gini (GI), l'indice de Herfindahl-Hirschman (HH), une formulation revisitée par les auteurs du même indice (HH_m), deux indices basés sur l'entropie de Shannon (OM_{PI} et OM_{MI} , le deuxième étant plus sensible au niveau de mobilité, c'est-à-dire plus sensible au nombre de modes de transport qu'à l'uniformité des intensités d'utilisation), l'indice de Theil (TH), l'indice de Dalton (DAL), une formulation revisitée par les auteurs du même indice (DAL_m) et l'indice d'Atkinson (ATK).

Les définitions de tous ces indicateurs, rapportées dans le Tableau 2.2 et adaptées dans un contexte de transport, impliquent le nombre de modes et les fréquences d'utilisation associées à chaque mode. Ces indices varient entre une valeur minimale et une valeur maximale, où l'une désigne le cas de l'utilisation exclusive d'un seul mode de transport (parfaite monomodalité) et l'autre le cas

où les intensités d'utilisation sont égales pour tous les modes (parfaite multimodalité). Le sens de l'interprétation varie selon l'indicateur, c'est-à-dire que soit la valeur minimale soit la valeur maximale peut correspondre au niveau de multimodalité le plus élevé.

Tableau 2.2 Indicateurs de multimodalité définis par Diana et Pirra (2016)

Indicateur	Formule	Minimum	Maximum	Quand l'indicateur ↑, la multimodalité ...
GI	$\frac{2 \sum_{i=1}^n i * f_i}{n \sum_{i=1}^n f_i} - \frac{n+1}{n}$	0	1	↘
HH	$\frac{1}{n} \left[\frac{n \sum_{i=1}^n (f_i - \bar{f})^2}{(\sum_{i=1}^n f_i)^2} + 1 \right]$	0	1	↘
HH_m	$\frac{1}{m} \left[\frac{n \sum_{i=1}^n (f_i - \bar{f})^2}{(\sum_{i=1}^n f_i)^2} + 1 \right]$	0	1	↘
OM_{PI}	$\sum_{i=1}^n \left[\frac{f_i}{\sum_{k=1}^n f_k} \ln \left(\frac{\sum_{k=1}^n f_k}{f_i} \right) \frac{1}{\ln n} \right]$	0	1	↗
OM_{MI}	$\sum_{i=1}^n \left[\frac{f_i}{nM} \left(1 + \ln \frac{M}{f_i} \right) \right]$	0	1	↗
TH	$\frac{1}{n} \sum_{i=1}^n \frac{f_i}{\bar{f}} \ln \left(\frac{f_i}{\bar{f}} \right)$	0	$\ln n$	↘
DAL	$1 - \frac{\frac{1}{n} \sum_{i=1}^n (f_i^{1-\varepsilon} - 1)}{\bar{f}^{1-\varepsilon} - 1}$	0	$\frac{1 - n^{-\varepsilon}}{1 - (f/n)^{\varepsilon-1}}$	↘
DAL_m	$1 - \frac{\frac{1}{n} \sum_{i=1}^m (f_i^{1-\varepsilon} - 1)}{\left(\frac{1}{m} \sum_{i=1}^m f_i \right)^{1-\varepsilon} - 1}$	0	$\frac{n-1}{n}$	↘
ATK	$1 - \left[\frac{1}{n} \sum_{i=1}^n (f_i/\bar{f})^{1-\varepsilon} \right]^{\frac{1}{1-\varepsilon}}$	0	$1 - n^{-\varepsilon(1-\varepsilon)}$	↘

où n est le nombre de modes étudiés, $f_1, \dots, f_n$ les fréquences d'utilisation correspondantes, $\bar{f}$ la valeur moyenne, M la fréquence maximale, m le nombre de modes utilisés, ε un paramètre représentant l'influence décroissante des modes les plus utilisés.

Quelques-uns de ces indicateurs, notamment l'entropie de Shannon (Diana & Mokhtarian, 2009), l'indice de Gini (Xiao et al., 2020) ou encore l'indice de Herfindahl-Hirschman (Kuhnimhof, 2009; Susilo, Yusak O & Axhausen, 2014) ont d'ailleurs déjà été appliqués dans la littérature afin d'analyser longitudinalement des variations dans les comportements d'usage de différentes modes. Heinen et Chatterjee (2015) ont également testé l'indice de Herfindahl-Hirschman pour deux classifications différentes de l'ensemble des modes considérés et l'ont comparé à deux autres

mesures plus simples : le nombre total de modes utilisés et la différence (en proportion) d'utilisation entre le mode le plus utilisé et le second mode le plus utilisé.

Cependant, l'application numérique réalisée par Diana et Pirra (2016) pour comparer deux individus A et B fictifs montre que les différents indicateurs donnent des indications contradictoires sur lequel de ces deux individus est le plus multimodal. Les auteurs ont alors recensé différentes propriétés mathématiques des indices de variabilité et ont déterminé si ces propriétés étaient désirables ou non dans un contexte d'évaluation de la multimodalité (en fonction des enjeux soulignés précédemment). L'analyse multicritère effectuée aboutit à la conclusion qu'il n'existe pas de mesure de la multimodalité qui se démarque systématiquement des autres dans toutes les circonstances. Néanmoins, certains indicateurs peuvent être préférables dans des situations spécifiques. En particulier, l'indice de Dalton revisité DAL_m est plus sensible aux écarts d'intensités d'utilisation et peut être ajusté pour mieux considérer les modes les moins utilisés. Par conséquent, cet indicateur est particulièrement adapté aux situations où certains modes sont moins utilisés que d'autres, ce qui est le cas à Montréal où la voiture joue un rôle prédominant.

2.2.3 Synthèse spécifique

En résumé, différents types d'interactions modales ont été mis en évidence dans la littérature, que ce soit entre des modes émergents et des modes classiques ou dans les comportements multimodaux des individus. Ces interactions ont été analysées à l'aide de diverses sources de données (enquêtes versus données opérationnelles), à différents niveaux d'analyse (individuel versus systémique) et pour différents horizons temporels (court, moyen ou long terme). Ce doctorat cherchera à les formaliser et à définir un cadre conceptuel permettant de toutes les englober en fonction de ces différentes dimensions. En outre, le sens des relations (complémentarité ou compétition) entre les modes n'est pas toujours très clair et semble dépendre de plusieurs paramètres. Bénéficiant d'un large éventail de données, cette recherche permettra de préciser ces interactions dans le contexte montréalais. L'enquête Origine-Destination de Montréal sera utilisée pour examiner la multimodalité des comportements au niveau individuel, alors que plusieurs bases de données passives permettront d'explorer l'utilisation de différents modes de transport au niveau systémique. Pour aller au-delà d'une simple description de ces interactions, elles seront également quantifiées en utilisant notamment des méthodes d'inférence causale et des indicateurs de multimodalité. Par ailleurs, les impacts de différents facteurs sur ces interactions seront modélisés.

2.3 Modélisation de la demande de transport

Une meilleure compréhension des interactions entre les modes permettra d'améliorer la modélisation de la demande de transport. Pour voir le lien entre ces deux importants défis, il est nécessaire de revenir à quelques éléments de base sur la modélisation du transport des personnes en milieu urbain. Cette modélisation est importante pour prévoir la demande future à plus ou moins long terme, mais aussi pour identifier les différents paramètres explicatifs qui peuvent jouer sur cet exercice de prévision. Bien que de nouvelles approches soient en cours de développement, « la connaissance des apports et limites des modèles à quatre étapes permet le plus souvent de mieux apprécier la portée des nouveaux modèles » (Bonnell, 2004). À la suite d'une brève présentation de ces méthodes classiques (section 2.3.1), deux types d'approches émergentes et cohérentes avec le sujet de cette thèse seront discutés : les approches longitudinales (section 2.3.2.1) et les approches multimodales (section 2.3.2.2). Cette recherche doctorale étant surtout en lien avec la troisième étape du modèle classique, différents facteurs exerçant une influence sur la répartition modale et donc, plus globalement, sur les interactions entre les modes, seront recensés (section 2.3.3).

2.3.1 Modélisation classique

Les premiers modèles de prévision de la demande de transport sont apparus aux États-Unis au début des années 1950. Ces modèles, initialement appliqués pour prévoir le trafic sur un axe isolé, ont ensuite été généralisés pour modéliser l'ensemble du réseau. De même, un changement de paradigme s'est peu à peu produit parallèlement à des besoins en planification qui ont également évolué : la modélisation, d'abord centrée sur la circulation des voitures, s'est finalement dirigée vers la prévision des déplacements des personnes, quel que soit le mode utilisé. En outre, bien que la modélisation classique ait d'abord été développée avec une approche agrégée, des modèles désagrégés l'ont peu à peu enrichie afin de mieux comprendre les facteurs influençant les comportements individuels (Bonnell, 2004).

La procédure séquentielle classique, ou procédure classique à quatre étapes, est composée, comme son nom l'indique, de quatre étapes : génération, distribution, répartition modale et affectation des déplacements. Dans la séquence des choix que doit faire un individu pour la réalisation d'un déplacement, ces différentes étapes correspondent respectivement au choix de se déplacer ou non, au choix de la destination (et de l'origine), au choix du mode de transport, puis au choix de

l'itinéraire (Bonnel, 2004). Cette procédure est illustrée par le schéma de la Figure 2.1. McNally (2007) regroupe les données d'entrée (exogènes) de cette procédure en deux grands systèmes : le système de transport **T** et le système d'activités **A**. Le premier correspond aux réseaux de transport, généralement représentés sous la forme de graphes définis par des liens et des nœuds. Le deuxième se réfère aux caractéristiques socio-démographiques et économiques de la population et du territoire étudié. Ce dernier est découpé en zones, chaque zone étant représentée par un centroïde relié aux réseaux de transport et d'activités par un système de connecteurs. En sortie, la procédure séquentielle classique permet d'obtenir des volumes à l'équilibre sur des liens.

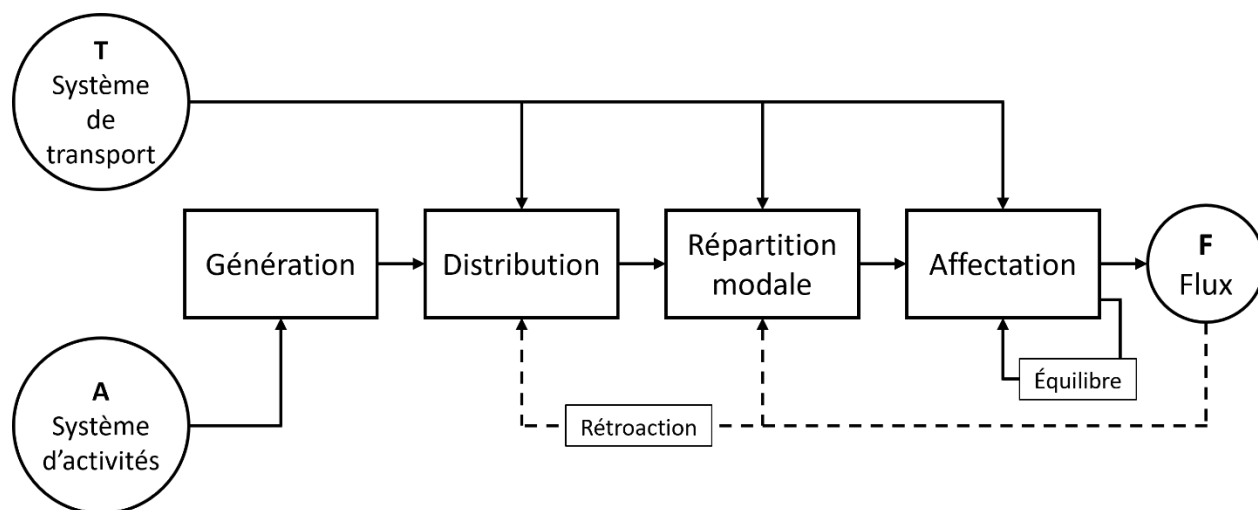


Figure 2.1 Schématisation du modèle à 4 étapes, inspirée de McNally (2007)

Les sous-sections qui suivent décrivent le rôle de chacune des quatre étapes, les principales méthodes utilisées et les limites correspondantes. Il s'agit d'un résumé de plusieurs livres fondamentaux qui peuvent être consultés pour plus de détails (Bonnel, 2004; McNally, 2007; Ortúzar & Willumsen, 2011).

2.3.1.1 Génération

La première étape de la procédure vise à déterminer le nombre total de déplacements effectués par les résidents de chaque zone en fonction des caractéristiques de cette zone, généralement pendant un jour moyen de semaine. Les déplacements sont différenciés en déplacements produits (ou émis ou sortant) et attirés (ou reçus ou entrant), c'est-à-dire en déplacements ayant leur origine ou leur destination dans la zone. Ces productions et attractions sont calculées séparément par deux modèles distincts, les unes à partir des attributs de la population résidente, les autres en se basant sur les

caractéristiques des activités disponibles. La génération des déplacements est souvent réalisée par motif, ou au moins en distinguant les trois catégories suivantes : les déplacements travail basés domicile, les autres déplacements basés domicile et les déplacements non basés domicile.

Plusieurs types de méthodes peuvent être appliqués pour exécuter cette première étape. En l'absence de données, des **modèles normatifs** (taux de génération fixes en fonction de certains paramètres) sont utilisés. Toutefois, le plus souvent, des **modèles de régression linéaire** sont entraînés à partir des données d'enquêtes de déplacements. Pour tenir compte de la variabilité des comportements, des **modèles catégoriels** (aussi appelés classifications croisées) ont été proposés : ils consistent à définir des classes de comportements supposés homogènes puis à construire un modèle de régression pour chaque classe. Des **modèles désagrégés** de choix discrets sont également parfois appliqués afin de modéliser au niveau individuel la probabilité de réaliser 0, 1, 2, etc. déplacements par jour ou par semaine.

Les limites associées à ces méthodes sont notamment liées au non-respect des hypothèses de la régression linéaire (non-linéarité de la relation avec la variable dépendante, hétéroscédasticité des résidus et multicollinéarité des variables explicatives), à l'inégalité des productions et des attractions totales, et à la non prise en compte de l'offre de transport (pouvant par exemple créer une demande induite). De plus, l'étape de génération est non-tendancielle, c'est-à-dire qu'elle ne permet pas de prendre en compte les tendances et suppose que les comportements sont stables dans le temps. Cette étape assume également un taux de mobilité homogène dans chaque zone (ce taux est le même partout dans la zone). En outre, les résultats de ces méthodes dépendent du choix et de la disponibilité des variables explicatives, ainsi que de la qualité des données.

2.3.1.2 Distribution

La deuxième étape de la procédure permet de spatialiser les déplacements en les distribuant entre les différentes zones sous la forme d'une matrice origine-destination. Au niveau désagrégé, cela équivaut à associer à chaque déplacement une zone d'origine et une zone de destination. Cette matrice origine-destination est calculée à partir de ses marges (obtenues à l'étape de génération), correspondant aux émissions et attractions de chaque zone.

Pour ce faire, différentes méthodes peuvent être employées. Tout d'abord, des **modèles à facteurs de croissance** peuvent être appliqués. Ils se traduisent par des déformations successives d'une matrice de référence pour arriver à une matrice horizon dont on ne connaît que les marges, fournies

par l'étape de génération. Différents algorithmes de déformation existent (facteurs uniformes, de croissance, Furness, Fratar). Le modèle de distribution le plus connu reste néanmoins le **modèle gravitaire**, inspiré de la loi gravitationnelle de Newton. Ce modèle prend les déplacements produits par une zone et les distribue dans les autres zones selon leur pouvoir attractif, exprimé par le nombre de déplacements attirés mais réduit par une résistance impliquant le temps, la distance à parcourir ou le coût généralisé du déplacement entre les deux zones. Plusieurs formulations de cette fonction de résistance sont possibles. Des **méthodes basées sur la maximisation d'une entropie** peuvent également être utilisées : le principe de ces méthodes consiste à trouver le méso-état (c'est-à-dire la matrice OD somme des micro-états) le plus probable à partir des connaissances que l'on a sur l'état du système au niveau macroscopique (c'est-à-dire les émissions et attractions par zone). Enfin, des **modèles d'opportunité**, basés sur une approche comportementale et probabiliste, ont été développés. L'idée derrière cette méthode est que les individus choisissent la meilleure opportunité parmi celles qui peuvent satisfaire à leur besoin, c'est-à-dire la destination la plus proche ou la moins coûteuse. Cette notion implique une comparaison relative, et non plus absolue comme dans le modèle gravitaire, des zones entre elles. Toutefois, elle suppose que chaque individu connaît toutes les alternatives qui s'offrent à lui.

L'étape de distribution en elle-même possède d'autres limites. Tout d'abord, les déplacements intra-zones et les déplacements sortant du périmètre d'étude, c'est-à-dire avec une destination dans une zone externe, ne sont pas modélisés. Des problèmes peuvent également survenir à la suite de la segmentation des données de référence en zones : cela peut amener à des cellules nulles qui, même multipliées par un facteur de croissance, resteront nulles. De plus, la pertinence de la fonction de coût généralisé est souvent questionnée : le choix ne dépend pas que du coût et toutes les opportunités ne sont pas équiprobables. Par ailleurs, la position de l'étape de distribution dans la séquence de la procédure est parfois remise en doute : est-ce que le choix de la destination précède ou succède au choix du mode ?

2.3.1.3 Répartition modale

La troisième étape de la procédure consiste à répartir les déplacements par mode, c'est-à-dire à décomposer la matrice origine-destination produite à l'étape de distribution en autant de sous-matrices qu'il y a de modes considérés. Cette étape est basée sur une hypothèse de rationalité selon laquelle les individus choisissent le mode présentant la plus grande utilité (exprimée par une

fonction de différentes variables). Par ailleurs, l'offre de transport étant inégale tout au long de la journée, les matrices de demande sont parfois segmentées en différentes périodes avant l'étape de choix modal (distribution temporelle des déplacements). Le plus souvent, on distingue les périodes de pointe (AM et PM) des périodes creuses.

Au niveau agrégé, l'étape de répartition permet de prévoir la proportion d'individus qui choisiront chaque mode. Les comportements individuels sont alors supposés aléatoires. En l'absence de données, des **modèles normatifs** sont disponibles (répartitions normées qui prennent parfois la forme d'abaques). Sinon, un **modèle logit** (ou **régression logistique**) est souvent calculé. Pour deux modes de transports, ce modèle est basé sur une loi binomiale, où la probabilité de choisir un mode est représentée par une courbe en S bornée en 0 et 1 en fonction de la différence d'utilité entre les deux modes. Cette méthode est généralisable à une loi multinomiale pour plusieurs modes. Lorsque l'hypothèse d'indépendance des alternatives disponibles n'est pas respectée, des modèles hiérarchiques avec une structure emboîtée doivent être appliqués. Ces modèles permettent par exemple de représenter le choix d'un mode individuel (la voiture) versus un mode collectif (transport en commun), puis un deuxième choix est impliqué entre les différents modes de transport en commun existants (métro, bus, etc.). Par ailleurs, il existe des **modèles combinant plusieurs étapes** : des modèles combinés distribution - répartition modale, ou des modèles dits directs regroupant les trois premières étapes dans un seul modèle.

Au niveau désagrégé, des **modèles de choix discrets** sont appliqués. On modélise alors le comportement de chaque individu (par la probabilité qu'il sélectionne un mode), ce qui permet de tenir compte de ses caractéristiques spécifiques plutôt que de considérer un individu moyen. Cette méthode est basée sur une approche probabiliste de l'utilité supposée aléatoire (l'hypothèse sur la forme de la distribution du terme aléatoire conduisant à différents types de modèles). De plus, l'univers de choix de chaque individu est supposé fini, ce qui implique la définition (non évidente) des modes disponibles et considérés par chaque individu (le « *choice set* »). La méthode permet d'obtenir des prévisions individuelles qui passent ensuite généralement par un processus d'agrégation pouvant entraîner des biais.

Les autres limites de cette troisième étape sont notamment liées à la formalisation de la fonction d'utilité : la forme sélectionnée est généralement additive mais ce choix n'est pas clair. De plus, cette fonction d'utilité implique la sélection de différentes variables explicatives. Toutefois, du fait

de la complexité du phénomène, il est difficile de prendre en compte toutes les variables qui expliquent le choix modal, incluant des facteurs psychologiques (les informations ne sont pas toujours disponibles). Au niveau désagrégé, le zonage a également un impact non négligeable sur les caractéristiques des déplacements. Enfin, les modèles de répartition modale n'incluent le plus souvent que deux modes (l'automobile et le transport en commun) et la condition d'indépendance des alternatives rend difficile la prise en compte de la complémentarité modale (combinaison de deux modes dans un même déplacement).

2.3.1.4 Affectation

Finalement, la dernière étape de la procédure vise à affecter les déplacements sur les réseaux en vue d'obtenir des volumes (ou des charges) sur les routes et les lignes de transport collectif. Cette étape implique l'application en amont d'une méthode de codification des réseaux. De même, quelques traitements préliminaires sont généralement nécessaires. En effet, les matrices obtenues en sortie de l'étape de choix modal exprimées en déplacements de personnes doivent être converties en véhicules pour pouvoir être affectées. Cette conversion est réalisée par le biais d'hypothèses sur le taux d'occupation des véhicules. En outre, l'étape d'affectation est souvent effectuée pour une période d'une heure : il est donc nécessaire de distribuer préalablement les déplacements dans le temps et de choisir la « pire » heure de la journée, c'est-à-dire la plus achalandée.

Le principe de cette étape est le même que dans les étapes précédentes : l'individu choisit son itinéraire en minimisant le coût généralisé de son déplacement. Les différentes méthodes d'affectation ont été classées par Bonnel (2004) en fonction de la prise en compte des contraintes de capacité des liens (les modèles ne considérant pas cette contrainte n'étant applicables que dans le cas de réseaux faiblement congestionnés) et en fonction du type d'utilité (déterministe ou probabiliste : dans le premier cas, tous les individus perçoivent de manière identique le coût d'un itinéraire, dans le deuxième cas, des facteurs supplémentaires sont considérés pour expliquer les choix individuels). Le Tableau 2.3 ci-dessous reprend cette classification. Les méthodes prenant en compte la congestion (méthodes d'équilibre) nécessitent plusieurs itérations : une étape d'initialisation est d'abord réalisée, conduisant à une distribution très hétéroclite (liens vides versus liens saturés), puis le processus est itéré pour redistribuer les véhicules en fonction du ratio débit/capacité des liens et des retards engendrés par la congestion.

Tableau 2.3 Classification des modèles d'affectation d'après Bonnel (2004)

		Procédure de choix de l'itinéraire	
		Déterministe	Probabiliste
Contrainte de capacité	Non	Méthode tout-ou-rien ou plus court chemin	Procédure stochastique pure (multichemin Dial, Burrell...), prix-temps
	Oui	Méthode d'équilibre (équilibre de Wardrop, affectation par tranche)	Méthode d'équilibre avec utilité stochastique

Les principales limites associées à l'étape d'affectation concernent le manque de données et d'indicateurs statistiques pour calibrer et valider le modèle, la sensibilité des résultats au zonage et à la définition des réseaux, ou encore les hypothèses des méthodes sur la capacité (supposée constante), la procédure de choix de l'itinéraire ou la convergence (critères). Cette convergence n'est d'ailleurs pas toujours garantie. La notion d'équilibre ne représente pas non plus nécessairement la réalité : elle implique une situation stabilisée qui ne tient pas compte de la variabilité des comportements.

Comme le montre la Figure 2.1, une rétroaction (de l'étape d'affectation sur les étapes de distribution et de répartition modale) est parfois introduite dans le modèle à quatre étapes. En effet, la charge des réseaux, et en particulier la congestion, a une incidence sur les temps de parcours. Or, ces derniers sont nécessaires aux étapes de distribution et de répartition pour le calcul des coûts généralisés et des utilités. Par conséquent, une procédure itérative doit être mise en place pour boucler sur les résultats de l'étape d'affectation. Pour ce faire, une première affectation peut être réalisée au début du processus (calage du modèle à quatre étapes) afin de calculer une première matrice de coûts généralisés. Cette affectation initiale, réalisée à partir des matrices de demande de référence issues des données d'une enquête sur les déplacements, permet d'obtenir les premières matrices origine-destination de coûts généralisés pour chaque mode de transport, ensuite utilisées comme données d'entrée pour les modèles de distribution et de choix modal. Les matrices de demande issues des étapes de génération-distribution-choix modal sont à leur tour affectées sur les réseaux. De nouvelles matrices de coûts généralisés par mode de transport peuvent alors être calculées. Celles-ci sont ainsi réinjectées dans la procédure jusqu'à ce que les matrices de coûts généralisés obtenues soit stables, c'est-à-dire jusqu'à ce que la procédure converge. Cependant, cette convergence est souvent difficile à obtenir (McNally, 2007).

2.3.1.5 Limites générales de la modélisation classique

De manière générale, plusieurs limites de la procédure séquentielle classique sont dues au niveau d'agrégation de la demande de transport, qui empêche la prise en compte de la variabilité réelle des comportements dans le temps et dans l'espace. Au niveau temporel, la demande n'est en effet prédite que pour une journée moyenne de semaine, voire seulement pour l'heure de pointe. Pourtant, la représentativité du jour typique de semaine a largement été remise en question (Verreault & Morency, 2011). De même, plusieurs paramètres de la méthode de modélisation classique sont supposés constants dans le temps, négligeant ainsi les phénomènes tendanciels de la mobilité. Ces tendances ont d'ailleurs été mises en évidence par Morency et Trépanier (2013) dans le contexte montréalais. C'est pourquoi le modèle de prévision de la demande de transport à Montréal (le MOTREM) est basé sur une approche tendancielle, mise en œuvre par un modèle d'ajustement des facteurs de pondération de l'enquête Origine-Destination. Pour tenir compte des tendances observées, ces derniers sont projetés à une date ultérieure en étant multipliés par des facteurs de croissance calculés pour différentes strates de la population. Des prévisions démographiques fournies par l'Institut de la Statistique du Québec sont ainsi utilisées (Tremblay, 2002). Toutefois, les tendances qui ne sont pas projetées sont supposées constantes (Morency & Trépanier, 2013). Au niveau spatial, les résultats de la procédure séquentielle classique sont très sensibles au découpage du territoire étudié (nombre de zones et frontières choisies). De plus, la segmentation du territoire est contrainte par un nombre minimum d'observations dans chaque zone, ce qui conduit généralement à des zones de taille assez grande. Ces limites liées aux niveaux d'agrégation sont essentiellement dues aux données sur lesquelles se base cette procédure. En effet, les enquêtes de déplacements étant transversales et échantillonnées, elles n'apportent pas nécessairement les précisions exigées par l'exercice de modélisation.

Par ailleurs, la procédure séquentielle classique ne permet pas de prendre en compte toutes les interactions présentes dans un système urbain. Bonnel (2004) le décompose en trois sous-systèmes : le sous-système des infrastructures de transport, le sous-système de localisation des activités et le sous-système des relations sociales (comportements d'activités et de mobilité). De plus, l'auteur différencie les interactions entre ces trois sous-systèmes dans le temps (à court, moyen, long terme) et dans l'espace (du quartier à l'agglomération en passant par l'échelle de la zone). À l'échelle du quartier et à court terme, les interactions entre les trois sous-systèmes peuvent être négligées, et les sous-systèmes des transports et des activités peuvent être considérés comme

figés. À l'échelle de la zone et du moyen terme, les sous-systèmes des transports et des activités évoluent mais leurs interactions ne sont pas considérées. Ainsi, ces sous-systèmes ne sont que des données exogènes et ne sont pas modifiés par la procédure, mais ils ont un effet (univoque) sur les comportements. Le modèle à quatre étapes s'applique dans ces dimensions. Toutefois, en pratique, le manque de rétroaction entre les quatre étapes du modèle et le besoin d'adapter les paramètres de coûts généralisés rendent difficile la simulation des impacts de différentes mesures d'investissement (telles que l'implantation de nouvelles infrastructures ou de politiques). De même, l'agrégation spatiale discutée précédemment ne permet pas d'analyser les impacts d'interventions locales. Par ailleurs, à l'échelle de l'agglomération et à long terme, toutes les interactions (réciproques) entre les trois sous-systèmes devraient être prises en compte.

2.3.2 Nouvelles approches de modélisation

Afin de dépasser plusieurs des limites de la procédure séquentielle classique, de nouvelles approches de modélisation ont été développées, notamment des approches intégrées visant à prendre en compte différents types d'interactions et à rendre endogènes un certain nombre de paramètres. Un projet réalisé pour le MTQ a permis de dresser une revue de littérature sur le sujet (Morency & Trépanier, 2013). En particulier, ce projet s'intéresse aux modèles d'activités, qui permettent de générer des programmes d'activités individuels plutôt que des déplacements agrégés par zone, en partant du principe que la demande de transport est une demande dérivée du besoin de réaliser des activités. Le module TASHA (Travel Activity Scheduling Model for Household Agents), qui fait partie du modèle de micro-simulation basé agents ILUTE développé à l'Université de Toronto, est ainsi appliqué au cas de Montréal. Toutefois, ce modèle ne permet toujours pas de prendre en compte les tendances observées et, étant encore basé sur les données transversales d'une enquête de déplacements, il ne permet de modéliser qu'une journée de semaine moyenne. Cette limite des modèles d'activités a également été reconnue par Rasouli et Timmermans (2014) qui ont appelé à développer des méthodes dynamiques pour différents horizons temporels. Face à cette problématique persistante, il apparaît pertinent de recenser des approches longitudinales qui permettraient de tenir compte de la variabilité des comportements de mobilité (section 2.3.2.1).

Bien que la modélisation de la demande de transport ait historiquement été réalisée pour un nombre limité de modes, différentes recherches ont été menées sur le choix modal afin de préciser les fonctions d'utilité de chaque mode et d'intégrer de nouveaux modes de transport dans les modèles.

Cela a notamment permis aux modèles d'activités et multi-agents d'être appliqués dans des réseaux multimodaux, pour les modes inclus dans les données d'enquêtes utilisées (Atmani, 2015; Chow & Djavadian, 2015; Li, Q. et al., 2018). Toutefois, même s'il tient compte de la structure des chaînes d'activités et des interactions avec les autres membres du même ménage, le module de choix modal qui vient après la génération des programmes d'activités reste basé sur des modèles de choix discrets et des coûts généralisés. Cette approche ne permet pas facilement de modéliser les déplacements multimodaux, ni d'évaluer les impacts de stratégies d'intervention visant à favoriser la mobilité multimodale. Quelques réflexions sur le sujet sont rapportées dans la section 2.3.2.2.

2.3.2.1 Modélisation longitudinale

La modélisation basée sur des données longitudinales fait appel à des méthodes de prévision de séries temporelles. Ces dernières, utilisées dans de multiples disciplines scientifiques, sont définies ci-après (section 2.3.2.1.1). Ces séries peuvent servir à différentes tâches d'exploration de données telles que la classification, la détection d'anomalies ou encore la décomposition en motifs (sous-séquences), mais aussi à la prévision (Esling & Agon, 2012). Deux livres publiés récemment recensent diverses méthodes de prévision de séries temporelles avec des exemples de mise en pratique dans le logiciel statistique R (Brockwell & Davis, 2016; Hyndman & Athanasopoulos, 2018). S'inspirant principalement de ces deux ouvrages, additionnés à d'autres références complémentaires de la littérature, les principales techniques de modélisation longitudinale sont répertoriées dans la section 2.3.2.1.2. Finalement, quelques exemples d'application en prévision de la demande de transport sont fournis dans la section 2.3.2.1.3.

2.3.2.1.1 Séries temporelles : définition, composantes structurelles et indicateurs

Une série temporelle (ou série chronologique) est définie comme une séquence ordonnée de valeurs issues de mesures répétées dans le temps (Esling & Agon, 2012). Dans le cas du transport, cette série peut par exemple prendre la forme d'une séquence de 0 et de 1 indiquant si l'utilisateur s'est déplacé ou non pendant une période de temps donnée (Agard et al., 2006), ou encore d'une suite ordonnée de nombres de déplacements par mode (Morency et al., 2018).

Une série temporelle est généralement décrite par trois types de phénomènes structurels observés dans les données : les termes « tendance », « saisonnalité » ou « cycle » sont ainsi souvent entendus

(Hyndman & Athanasopoulos, 2018). Une **tendance** existe lorsqu'une augmentation ou diminution est observée à long terme dans les données. Elle peut également désigner un changement de direction quand une augmentation devient une diminution ou inversement. Une **saisonnalité** est détectée lorsque la variable d'intérêt est affectée par des facteurs saisonniers liés au calendrier tels que la période de l'année ou le jour de la semaine. Cette saisonnalité est caractérisée par une fréquence (ou une période, inverse de la fréquence) fixe. On dit qu'un **cycle** est présent dans les données lorsque la variable analysée augmente et diminue sans avoir une périodicité constante. Par exemple, les fluctuations dues à l'économie peuvent générer des cycles dans les données. Un cycle est généralement plus long et a un effet plus variable qu'un phénomène saisonnier.

Différents moyens peuvent permettre de visualiser les variations des données d'une série temporelle (Hyndman & Athanasopoulos, 2018). Aussi, la plupart des applications basées sur ces séries nécessitent des mesures de similarité. Il existe ainsi différents types de métriques de distance entre deux séries temporelles (Esling & Agon, 2012; He et al., 2018). Un des indicateurs les plus connus est la **corrélation**, qui évalue la force de la relation linéaire entre deux séries temporelles. Toutefois, des relations non linéaires peuvent parfois être mises en évidence (Hyndman & Athanasopoulos, 2018). L'**autocorrélation** est la corrélation d'une série temporelle avec elle-même décalée dans le temps (d'un ou plusieurs pas de temps). En anglais, on parle de « *time lags* » pour désigner ces décalages. Cet indicateur, visualisé via un **corrélogramme** (représentant la mesure d'autocorrélation en fonction du décalage temporel), permet de mettre en évidence la fréquence d'une saisonnalité car l'autocorrélation sera plus élevée pour les décalages multiples de la période du phénomène. Une manière alternative d'obtenir cette fréquence est de tracer un **périodogramme** en estimant la densité spectrale, c'est-à-dire la force explicative associée à chaque fréquence par une transformation de Fourier (Brockwell & Davis, 2016). Le **spectrogramme**, basée sur la transformée en ondelettes (*wavelet transform*), permet quant à lui d'analyser la variabilité temporelle de la force spectrale associée à chaque fréquence (Torrence & Compo, 1998).

2.3.2.1.2 Techniques de modélisation longitudinale

La prévision d'une série temporelle peut être réalisée à l'aide d'une multitude de méthodes. Les plus basiques comprennent la méthode de la **moyenne historique**, qui considère que les valeurs futures prédites sont égales à la moyenne des données passées, ou la **méthode naïve**, parfois aussi appelée méthode de marche aléatoire (*random walk*), qui consiste à prédire une valeur au temps

$t + 1$ en la supposant égale à la valeur au temps t . Cette méthode naïve peut également prendre en compte une saisonnalité en supposant que la valeur au temps $t + 1$ est égale non pas à la valeur précédente immédiate mais à la dernière valeur observée à la même saison de l'année, par exemple au même mois de l'année précédente si la saisonnalité considérée est annuelle (méthode naïve saisonnière). La **méthode de Drift** est une variante de la méthode naïve qui permet aux prévisions d'augmenter ou de diminuer dans le temps, l'ampleur du changement dans le temps devant correspondre au changement moyen observé dans les données historiques.

Les **modèles de régression linéaire** ordinaires, souvent utilisés sur des données transversales, peuvent également être appliqués aux séries temporelles. Le principe consiste alors à prévoir une série temporelle (variable dépendante) en supposant une relation linéaire avec d'autres séries temporelles (variables indépendantes qui évoluent dans le temps, en décalage ou non par rapport à la variable dépendante) et des variables explicatives (variables indépendantes fixes). La tendance de la série temporelle prédite (variable dépendante) est alors mesurée à l'aide d'un terme linéaire de premier ordre représentant le temps. Pour prendre en compte une saisonnalité telle que l'effet du jour ou du mois, des variables factices (*dummy variables*) ou des termes de Fourier peuvent être ajoutés dans l'équation. Cependant, la modélisation de séries temporelles par une régression linéaire entraîne généralement une violation de l'hypothèse selon laquelle les termes d'erreurs sont indépendants. En effet, la présence de saisonnalité dans la série et donc d'autocorrélation fait en sorte qu'une variable prédite à une période donnée pourra être similaire à sa valeur dans la période précédente, ou à sa valeur décalée de deux ou plusieurs périodes.

Au contraire, les **modèles autorégressifs AR** (*AutoRegressive*) permettent, comme leur nom l'indique, l'inclusion de termes autorégressifs. En effet, le terme *autorégression* signifie que la variable dépendante est régressée sur elle-même. Ainsi, dans un modèle autorégressif, la variable d'intérêt est prédite à partir d'une combinaison linéaire de valeurs passées de cette même variable. Un modèle autorégressif d'ordre p , noté $AR(p)$, contient les variables « décalées » y_{t-1} , y_{t-2} , ..., y_{t-p} pour prédire la variable y_t . Plutôt que de faire intervenir les valeurs passées de la variable dépendante, les **modèles à moyenne mobile MA** (*Moving Average*) utilisent une combinaison linéaire des erreurs de prévision passées. Ainsi, un modèle d'ordre q , noté $MA(q)$, contient les erreurs ε_{t-1} , ε_{t-2} , ..., ε_{t-p} en plus de l'erreur ε_t . Le **modèle ARMA** (*AutoRegressive Moving Average*) d'ordre p et q , soit $ARMA(p, q)$, est une combinaison des modèles $AR(p)$ et $MA(q)$.

Tous les modèles précédents (AR, MA et ARMA) supposent des séries stationnaires, c'est-à-dire des séries temporelles dont les propriétés ne dépendent pas du temps. Toutefois, il est possible de les généraliser à des données non-stationnaires en différenciant la série temporelle initiale, c'est-à-dire en calculant la différence entre des observations consécutives ($y'_t = y_t - y_{t-1}$) une ou plusieurs fois. Les **modèles ARIMA** (*AutoRegressive Integrated Moving Average*) modélisent alors cette série temporelle différenciée plutôt que la série temporelle initiale, le terme « *integrated* » étant ajouté pour désigner l'étape préalable de différenciation (même si l'intégration est en fait plutôt l'inverse de la différenciation). Les modèles ARIMA sont désignés par $ARIMA(p, d, q)$ où p est l'ordre du modèle AR, d le nombre de différenciations de la série temporelle initiale et q l'ordre du modèle MA (Box et al., 2015). Le **modèle ARIMAX** est une extension du modèle ARIMA, le X étant ajouté pour indiquer que le modèle permet d'inclure des variables exogènes. Une autre alternative consiste à combiner le modèle ARIMA et le modèle de régression linéaire : un terme d'erreur autorégressif qui suit un modèle ARIMA est ajouté dans le modèle de régression pour tenir compte de l'autocorrélation des résidus. On parle alors de **modèles dynamiques de régression**. Les modèles ARIMA permettent de considérer l'autocorrélation dans les données mais ils sont moins appropriés pour décrire la structure de la série temporelle. Au mieux, ils prennent en compte une seule saisonnalité (**modèle SARIMA**), mais cela implique de déterminer un grand nombre de paramètres supplémentaires. Par conséquent, d'autres méthodes plus adaptées ont été développées.

Les méthodes de décomposition permettent de diviser une série temporelle en plusieurs composantes, chaque composante pouvant représenter un phénomène structurel sous-jacent (une tendance, une saisonnalité ou un cycle). En général, la tendance et les cycles sont considérés dans la même composante. Par conséquent, la série temporelle est décomposée en trois parties : une composante tendancielle-cyclique, une composante de saisonnalité et une composante restante (d'erreurs) qui contient tout ce qui n'est pas expliqué par les phénomènes temporels. Cette décomposition peut être additive ou multiplicative, la deuxième étant plus appropriée quand l'ampleur des fluctuations varie autour de la tendance. La méthode de décomposition classique est basée sur un calcul de moyennes (moyennes mobiles pour la composante tendancielle-cyclique et moyennes par saison pour la composante de saisonnalité). Toutefois, des méthodes plus évoluées et plus robustes ont été développées, comme les **décompositions X11** et **SEATS** (*Seasonal Extraction in ARIMA Time Series*), expliquées entre autres par Dagum et Bianconcini (2016), ou

encore la **décomposition STL** (*Seasonal and Trend decomposition using Loess*). Cette dernière méthode, inventée par Cleveland et al. (1990), présente plusieurs avantages sur les décompositions précédentes. En particulier, elle permet de prendre en compte tous les types de saisonnalités (et non seulement des saisonnalités mensuelles ou trimestrielles comme dans la décomposition X11) et elle est moins sensible aux valeurs aberrantes. De plus, la (ou les) composante(s) de saisonnalité peut varier dans le temps et la régularité de la composante tendancielle-cyclique peut être contrôlée. Ces méthodes de décomposition permettent de mieux comprendre une série temporelle, mais elles peuvent aussi servir à faire des prévisions. Pour ce faire, la composante de saisonnalité et la partie désaisonnalisée de la série temporelle sont prédites séparément. Les phénomènes saisonniers étant souvent supposés constants dans le temps, on prédit la composante désaisonnalisée (par exemple, via un modèle ARIMA) et on applique les mêmes coefficients que ceux de l'année précédente pour prendre en compte l'effet des différentes saisonnalités considérées (méthode naïve saisonnière). En outre, ces différentes composantes sont également modélisées séparément dans les **modèles ETS** (modèles d'espaces d'état, *state space models* en anglais), qui sont une généralisation de méthodes de lissage exponentiel de Holt-Winters (Gardner, 2006; Hyndman et al., 2008). Ces méthodes impliquent plusieurs équations : une équation de prévision ou de mesure, et trois équations d'état (une équation de lissage, une équation de tendance et une équation de saisonnalité).

Par ailleurs, des méthodes spécifiques existent dans le cas de la modélisation de séries temporelles hiérarchiques afin de tenir compte de la structure imbriquée des données. Généralement, les prévisions agrégées (par exemple, pour des groupes d'éléments) sont estimées en sommant les prévisions individuelles (par exemple, par élément) obtenues à partir des séries temporelles désagrégées situées au niveau le plus bas; c'est l'approche « *bottom-up* ». Les **modèles multiniveaux** peuvent également être manipulés avec des séries temporelles hiérarchiques. Ces modèles sont surtout connus pour leur utilisation dans le domaine de l'éducation, afin de comparer les résultats de tests entre des élèves de différentes classes provenant de plusieurs écoles, mais ils peuvent également être appliqués sur des données longitudinales : on les appelle alors parfois modèles de croissance (*growth models*). Ces modèles sont principalement utilisés en sociologie et en psychologie pour étudier l'évolution des comportements individuels (Singer & Willett, 2003). Des mesures répétées décrivant chaque individu constituent un modèle hiérarchique à deux niveaux (individus au niveau 2 et observations au niveau 1), composé de deux équations. L'équation de niveau 1 modélise le changement de chaque individu dans le temps (changement intra-individuel)

et l'équation de niveau 2 explique les différences entre les individus (hétérogénéité inter-individuelle). Contrairement aux modèles linéaires classiques, les modèles multiniveaux ne supposent pas que les observations sont indépendantes : l'autocorrélation dans les données de chaque individu est considérée. Ils tiennent également compte des différences entre les individus, même si les variables qui expliquent ces différences ne sont pas toutes contrôlées. Par conséquent, cette approche permet une inférence au bon niveau et réduit les biais liés à l'omission de variables. Les caractéristiques des individus peuvent être modélisées par des effets fixes ou aléatoires. Dans le premier cas, l'ordonnée à l'origine est propre à chaque individu (une variable binaire est ajoutée dans le modèle pour chaque individu). Dans le deuxième cas, des covariables (attributs individuels) et un terme d'erreur aléatoire (qui suit une distribution normale) sont inclus dans le modèle. Les modèles fixes permettent de mieux capturer l'effet de variables omises, mais les modèles aléatoires sont plus adaptés lorsque le nombre d'individus est très élevé (le terme aléatoire permet de réduire le nombre de paramètres) et ils offrent également une plus grande flexibilité, par exemple pour modéliser l'autocorrélation dans les données (Allison, 2009; Huang, F. L., 2018).

La prévision de séries temporelles est un sujet de recherche en pleine évolution et de nouvelles approches ont récemment fait leur apparition. En particulier, des modèles multivariés (avec plusieurs variables dépendantes) ont été développés afin de pouvoir prédire un vecteur de deux ou plusieurs séries temporelles. Par exemple, le **modèle VAR** (*vector autoregression*) est une généralisation multivariée du modèle univarié AR (Pfaff, 2008). Dans la même catégorie de modèles permettant de mesurer diverses relations entre plusieurs variables, les modèles d'équations structurelles sont également en cours de développement pour pouvoir introduire dans les équations des termes autorégressifs : cela devient des **modèles d'équations structurelles dynamiques** (Asparouhov et al., 2018). Enfin, des méthodes non-paramétriques, notamment des **méthodes d'apprentissage machine** (*machine learning*), ont aussi leur place parmi les méthodes de prévision de séries temporelles. Des réseaux de neurones autorégressifs ont par exemple été proposés (Hyndman & Athanasopoulos, 2018).

2.3.2.1.3 Quelques exemples d'application en transport

Dans le domaine du transport, l'utilisation de ces méthodes pour prévoir l'achalandage d'un mode à court ou moyen/long terme est de plus en plus courante car encouragée par la disponibilité croissante de données passives longitudinales, notamment des données de cartes à puce pour le

transport en commun. Le plus souvent, des modèles de régression linéaire sont appliqués à des séries temporelles afin d'examiner l'influence de différentes variables explicatives variant dans le temps (par exemple, la météo ou le prix du carburant) sur la demande (Morency, Trépanier, Faucher, et al., 2017; Zhou, M. et al., 2017).

Dans le but de prévoir l'achalandage mensuel d'une société de transport en commun, Chiang et al. (2011) ont introduit un terme autorégressif dans leur régression afin de corriger l'autocorrélation temporelle des erreurs. Ils ont comparé ce modèle dynamique de régression à deux autres modèles : un modèle de réseau de neurones et un modèle ARIMA. Les résultats ont révélé que le modèle de régression et le réseau de neurones avaient des performances assez similaires. En revanche, le modèle ARIMA a du mal à prévoir la demande sur un long horizon temporel. Cette dernière conclusion, confirmée par Lepage et Morency (2021), expliquerait pourquoi l'utilisation des modèles ARIMA a été privilégiée dans des analyses à court terme (Ding et al., 2018; Tao et al., 2018). De même, des modèles ETS ont été entraînés dans un contexte de transport pour prédire les flux de passagers sur les prochaines tranches de 15 minutes (Noursalehi et al., 2018) ou les prochains jours (Wang, Xiao, 2019). De plus, des modèles multiniveaux ont été appliqués avec des séries temporelles nichées dans des divisions géographiques ou des éléments du réseau, par exemple pour déterminer les facteurs affectant l'utilisation d'un service d'autopartage (De Lorimier & El-Geneidy, 2013) ou de vélopartage (El-Assi et al., 2017) à l'échelle d'une station, ou pour étudier les déterminants de l'achalandage du transport en commun à l'échelle d'une ligne (Diab, DeWeese, et al., 2020) ou d'une ville (Boisjoly et al., 2018; Erhardt et al., 2022).

L'application des nouvelles approches de prévision de séries temporelles reste encore relativement limitée dans le domaine du transport. Toutefois, quelques auteurs ont prouvé l'utilité des modèles basés sur des ensembles d'arbres de décision (compilés par *random forests* ou *gradient boosting*) et sur des réseaux de neurones LSTM (*Long-Short Term Memory*) pour prévoir l'achalandage à court et long termes de différents modes de transport collectif (Ding et al., 2016; Lin, S. & Tian, 2020; Toque et al., 2017). En outre, ces modèles d'apprentissage machine ont aussi été combinés avec des méthodes de décomposition et d'agrégation (approche « bottom-up ») de séries temporelles pour réaliser des prévisions à moyen terme (Egu & Bonnel, 2021; Qin et al., 2019).

Cependant, tous les modèles recensés n'ont été entraînés que pour un seul mode de transport. La prévision longitudinale de la demande de transport est donc majoritairement unimodale.

2.3.2.2 Modélisation multimodale

Basée sur les modèles classiques à quatre étapes, la planification des transports a longtemps été axée sur un seul mode : l'automobile privée. Aujourd'hui, on reconnaît la nécessité de se tourner vers une planification multimodale et intégrée des transports, c'est-à-dire une planification qui considère tous les modes de transport ainsi que les possibles connexions entre ces modes. Cela exige d'abord de pouvoir évaluer les impacts (incluant la demande induite) de toutes les stratégies d'implémentation de mesures de mobilité, et non seulement de celles qui concernent l'ajout de capacité routière (Litman, 2017). Toutefois, l'aptitude des modèles existants à répondre à ce besoin a été questionnée, en particulier du fait de la non prise en compte des déplacements multimodaux (Van Eck et al., 2014). Un modèle de transport multimodal devrait pouvoir prédire l'utilisation de tous les modes mais aussi de toutes les séquences de modes faisables (en fonction des modes disponibles à chaque point de jonction), ce qui implique d'abord de déterminer toutes les combinaisons possibles. Ensuite, leur représentation mathématique à l'aide d'un modèle de choix est complexe : plus de choix doivent être considérés (pour chacun des segments de déplacement, l'attractivité et les coûts associés à chaque mode pouvant être différents selon sa position dans la séquence de modes et selon s'il est utilisé seul ou en combinaison avec d'autres modes), et le choix du mode et le choix du trajet deviennent intrinsèquement liés (les modes sont combinés en fonction de la présence d'infrastructures le long du trajet). En outre, l'interaction offre-demande entre les différentes alternatives (dépendantes) doit être prise en compte (par exemple, une forte demande pour le vélo aura un impact sur les places de stationnement vélo disponibles à proximité des arrêts de transport en commun, potentiellement utilisés pour faire la connexion entre les deux modes), et des coûts supplémentaires doivent être attribués aux correspondances.

Face à ces exigences, Van Eck et al. (2014) regroupent les modèles existants en trois types de modèles, rangés de l'approche la moins adaptée à l'approche la plus adaptée pour représenter la complexité des comportements multimodaux :

- Les **modèles classiques** sont ceux de la section 2.3.1. Ces modèles sont basés sur une séparation stricte entre les modes privés et collectifs puisque les deux réseaux sont représentés de manière distincte. Généralement, on considère l'automobile d'un côté, et le transport en commun de l'autre. La demande du vélo en tant que mode à part entière est ignorée, et il n'y a pas de transfert possible entre les modes (puisque les modes d'accès et de

sortie ne sont pas modélisés). Les parts modales sont calculées à partir d'un modèle logit multinomial (*multinomial logit*) basé sur la maximisation d'une utilité.

- Les **modèles à la pointe de la pratique** (*state-of-the-practice model*) sont une extension aux modèles classiques. Les transferts entre modes privés (voiture et vélo) et modes collectifs sont autorisés aux arrêts de transport en commun, et les liens de connexion entre les nœuds du réseau routier et les arrêts de transport collectif sont inclus dans la représentation du réseau. Des séquences de deux ou plusieurs modes sont ainsi prédéterminées et modélisées comme une alternative artificielle à part entière. La répartition modale est généralement obtenue par un modèle logit niché (*nested logit*).
- Les **modèles à la pointe de la recherche** (*state-of-the-research model*) correspondent à une famille de modèles qui supposent trois conditions principales : la construction d'un réseau multimodal intégré (*supernetwork*) composé de plusieurs couches reliées par des liens qui représentent de possibles correspondances entre les modes, la génération a priori d'un ensemble d'alternatives (filtrées préalablement à l'étape d'affectation pour ne garder que les alternatives réalisables et compétitives) et la modélisation simultanée (et itérative) du choix du mode et de l'itinéraire. Les parts de la demande de déplacement pour chaque alternative mode x itinéraire sont ensuite calculées avec un modèle logit combinatoire couplé (*paired combinatorial logit*).

Toutefois, dans tous les exemples d'application de ces modèles, la demande totale de déplacements est supposée fixe et inélastique. Cette limite est probablement due à un manque d'informations sur le nombre de déplacements générés par l'implantation d'une nouvelle infrastructure. En effet, dans le cadre de l'évaluation de l'incidence de l'aménagement du territoire sur le transport, une revue de littérature internationale sur l'estimation des taux de génération de déplacements a montré que la plupart des études réalisées jusqu'alors ont été centrées sur la génération de véhicules (automobiles) uniquement (De Gruyter, 2019). Cette focalisation peut entraîner des infrastructures insuffisantes pour permettre l'utilisation des autres modes de transport. Ainsi, les auteurs insistent sur la nécessité de dépasser ces pratiques passées et de s'orienter vers des taux de génération multimodaux, c'est-à-dire des nombres de déplacements de personnes (et non des véhicules) générés pour chaque mode de transport, incluant les modes émergents. Cependant, cette génération de déplacements-modes (combinaison des étapes de génération et de répartition modale de la

procédure classique) s'applique seulement dans le cas de la modélisation basée déplacements. D'autres méthodes devront être développées pour les modèles d'activités.

Par ailleurs, la modélisation multimodale requiert la prise en compte d'une multitude de facteurs permettant d'expliquer les comportements d'utilisation des différents modes de transport. Beaucoup de recherches ont été menées afin d'identifier ces facteurs.

2.3.3 Facteurs pouvant avoir un effet sur la répartition modale

Divers facteurs peuvent avoir un impact sur la répartition modale de la mobilité, tant au niveau agrégé que désagrégé. De nombreux travaux se sont intéressés aux facteurs qui influencent le choix modal individuel : un court rappel basé sur des revues de littérature existantes est d'abord présenté (section 2.3.3.1). Toutefois, le sujet de cette recherche porte plutôt sur les facteurs qui ont une influence sur le changement de modes ou sur le choix d'un ensemble de modes : le fait d'être multimodal ou non, et le type de combinaisons de modes choisis (dans le temps ou dans le même déplacement). Les variables ayant un impact sur la multimodalité des comportements individuels sont donc ensuite rapportées (section 2.3.3.2). Aussi, des facteurs influant sur la demande globale à un niveau plus systémique sont répertoriés (section 2.3.3.3).

2.3.3.1 Facteurs ayant un effet sur le choix modal individuel

Le choix modal, défini comme le choix entre différents modes de transport, est un sujet de recherche majeur dans l'étude des comportements de mobilité individuels. De connivence avec l'évolution de la modélisation en transport, trois types d'approches peuvent être distingués : l'approche rationaliste, l'approche sociogéographique et l'approche socio-psychologique (De Witte et al., 2013). L'**approche rationaliste**, la plus répandue, est fondée sur l'hypothèse que les voyageurs prennent des décisions rationnelles face à différentes alternatives impliquant la connaissance de toutes sortes d'informations. La personne choisit ainsi le mode de transport qui maximise une fonction d'utilité. Dans cette approche, les déterminants du choix modal considérés sont principalement des composantes objectives, rationnelles et économiques, comme le temps et le coût de déplacement. L'**approche sociogéographique** introduit explicitement la dimension spatiale dans le processus de choix modal. Elle s'appuie sur le principe que les individus se déplacent pour réaliser des activités qui s'enchaînent dans le temps et dans l'espace. Dans cette approche, des indicateurs spatiaux sont donc ajoutés à la définition du choix modal. Enfin,

l'**approche socio-psychologique** explique le choix modal par les attitudes et les perceptions des individus à l'égard des modes de transport disponibles. La procédure qui encadre le choix modal comporte donc à la fois des facteurs objectifs et subjectifs. Dans cette approche, des variables psychologiques sont incluses.

Ainsi, le choix modal peut être influencé par une multitude de déterminants qui font écho à différentes facettes de la mobilité. De Witte et al. (2013) les regroupent en quatre catégories : les indicateurs socio-démographiques, les indicateurs spatiaux, les caractéristiques des déplacements et les indicateurs socio-psychologiques. Martel-Poliquin (2012) segmente quant à lui ces facteurs en cinq groupes : les variables de la personne, les variables du ménage, les variables du milieu bâti, les variables des déplacements et les autres variables. Ces deux classifications sont finalement assez similaires et leur croisement permet de recenser de manière quasi exhaustive l'ensemble des facteurs impliqués dans le choix modal individuel. Ces facteurs peuvent être répartis comme suit :

- **Les variables socio-démographiques de la personne et du ménage**

Les variables de la personne peuvent être décomposées en propriétés intrinsèques comme l'âge et le genre, et en propriétés extrinsèques telles que le niveau d'éducation, l'occupation principale, la possession d'un permis de conduire ou la possession d'autres outils de mobilité (abonnement au transport en commun, vélo, etc.). Les variables du ménage incluent la structure du ménage, en termes de composition mais aussi de taille, la présence d'enfants étant une caractéristique particulièrement importante. D'autres variables comme la possession automobile ou le revenu peuvent être définies à la fois à l'échelle de la personne et du ménage. La culture, l'ethnicité et la religion des personnes pourraient également avoir un impact sur certains déplacements.

- **Les variables spatiales du milieu bâti et de la géographie**

Ewing et Cervero (2010) résument les variables du milieu bâti en 5D : densité, diversité, design, destinations (accessibilité des) et distances d'accès au transport. Ces cinq catégories font respectivement référence aux densités de population et d'emplois, à la mixité des usages au sol (par exemple : résidentiel, commercial, institutionnel, etc.), à l'organisation de la ville, à l'accessibilité et à la proximité des systèmes d'activités et de transport (incluant la présence de stationnements et la fréquence des services de transport en commun). L'organisation de la ville peut, entre autres, être mesurée par la connectivité du réseau routier (souvent exprimée en termes de densité d'intersections). De plus, des propriétés géométriques et géographiques comme la tortuosité (ratio

entre la distance réseau d'un trajet et la distance euclidienne dite à vol d'oiseau), la taille et la topographie de la ville sont d'autres éléments pouvant être examinés. Par ailleurs, des variables comme l'âge de la ville ou la distance au centre-ville peuvent servir de proxy pour plusieurs caractéristiques spatiales : l'âge de la ville peut permettre de tenir compte des propriétés urbanistiques de l'époque à laquelle la ville a été construite (généralement conçue en privilégiant un certain mode de transport), et une distance plus courte au centre-ville est souvent synonyme d'une meilleure accessibilité et d'une meilleure desserte en transport et en activités.

- Les variables caractéristiques des déplacements

Ces variables sont celles que l'on retrouve le plus souvent dans les modèles agrégés au niveau du déplacement. Elles comprennent le motif de déplacement, la distance, la durée, le coût, les horaires (notamment considérés par l'individu pour fixer son heure de départ), la complexité de la chaîne de déplacements à réaliser durant la journée et la présence de correspondances.

- Les variables socio-psychologiques

Cette catégorie fait référence aux expériences passées, à la familiarité (liée aux connaissances que les utilisateurs ont développé sur les systèmes de transport à leur disposition), au style de vie, aux habitudes et aux perceptions de l'usager. La méta-analyse réalisée par Lanzini et Khan (2017) permet entre autres de replacer ces variables dans le cadre de différents modèles théoriques. Le phénomène psychologique de la dépendance à l'automobile est également un facteur de plus en plus modélisé (Moody & Zhao, 2019).

- Autres variables

Les conditions météorologiques, la qualité et la quantité des informations disponibles sur les différentes alternatives de transport sont d'autres variables à considérer dans le choix modal individuel.

Par ailleurs, ces mêmes facteurs peuvent également expliquer les changements d'un mode de transport à un autre. Ainsi, Diana (2010) a proposé d'étudier la diversion modale plutôt que le choix modal. Dans cette perspective, plusieurs travaux ont été réalisés afin de modéliser (au niveau de l'individu ou du déplacement) les substitutions provoquées par l'implantation des nouveaux services de mobilité (Ceccato & Diana, 2021; Fukushima et al., 2021; Loa et al., 2021). De même, plutôt que d'analyser le choix d'un seul mode, il est possible de modéliser le choix d'une

combinaison de modes sur une période (ou dans un même déplacement), c'est-à-dire le fait qu'un individu (ou un déplacement) soit multimodal ou non. C'est ainsi que plusieurs recherches ont été menées afin de mettre en évidence les facteurs ayant un effet sur la multimodalité des comportements individuels.

2.3.3.2 Facteurs ayant un effet sur la multimodalité des comportements

Plusieurs **variables socio-démographiques** de la personne et du ménage semblent être reliées à la multimodalité des individus. En particulier, les personnes plus jeunes ont généralement une plus grande probabilité d'être multimodales (Buehler & Hamre, 2014, 2015; Heinen & Chatterjee, 2015; Nobis, 2007). En étudiant la relation entre multimodalité et mode de vie, Nobis (2007) révèle une diminution drastique de la multimodalité des comportements à l'entrée dans la vie professionnelle. Cette tendance semble être accentuée par la présence d'enfants (Buehler & Hamre, 2015; Nobis, 2007). Au contraire, les personnes vivant seules ont plus de chances d'utiliser différents modes de transport pendant une période donnée (Kuhnimhof et al., 2006), en particulier les hommes (Buehler & Hamre, 2015). De même, le niveau d'éducation et les revenus sont corrélés positivement avec la multimodalité des individus (Buehler & Hamre, 2014, 2015; Heinen & Chatterjee, 2015; Kuhnimhof et al., 2006; Nobis, 2007). Néanmoins, l'occupation d'un travail à temps plein a un impact négatif sur la diversité de modes utilisés (Heinen & Chatterjee, 2015; Nobis, 2007), les personnes sans emploi étant ainsi caractérisées par une plus grande multimodalité (Buehler & Hamre, 2015). En outre, l'autre extrémité de la population (les personnes âgées) démontrent un niveau de multimodalité plus élevé que la moyenne (Kuhnimhof et al., 2006; Nobis, 2007). Toutefois, il n'est pas clair que cette observation soit due à un changement de comportement ou à un phénomène générationnel, impliquant que les personnes aient toujours eu ce comportement (Nobis, 2007). Aussi, la présence d'un handicap impactant la mobilité de la personne ou le fait d'appartenir à une minorité ethnique réduisent la variabilité modale (Heinen & Chatterjee, 2015). Par ailleurs, l'accessibilité à la voiture et la possession automobile font diminuer fortement la probabilité qu'un individu soit multimodal (Buehler & Hamre, 2014, 2015; Heinen & Chatterjee, 2015; Nobis, 2007). Par exemple, les ménages montréalais ayant accès à une automobile sont ceux qui l'utilisent le plus et qui ont la répartition modale la moins diversifiée (Sioui et al., 2013). Au contraire, une plus grande variabilité d'utilisation des modes a été trouvée chez les personnes possédant un vélo et un abonnement au transport collectif (Heinen & Chatterjee, 2015).

De même, la multimodalité des déplacements est reliée aux caractéristiques socio-démographiques des personnes. Souhaitant déterminer les facteurs qui influencent les comportements de déplacement intermodaux, Gebhardt et al. (2016) ont défini un « usager intermodal » comme un individu ayant fait au moins un déplacement multimodal (combinant plusieurs modes) durant la période enquêtée. Comparativement aux usagers monomodaux, la proportion de femmes et de jeunes est plus élevée parmi les usagers intermodaux. Ces derniers sont composés en majorité de personnes qui travaillent, qui ont un plus haut niveau de ressources et de compétences, qui ont moins accès à l'automobile mais qui possèdent généralement un abonnement au transport en commun. Ces résultats sont en partie confirmés par Oostendorp et Gebhardt (2018). Les auteurs différencient cet effet selon le type de combinaison de modes utilisés. Ainsi, les personnes qui réalisent des déplacements TC + TC et vélo + TC ont tendance à être plus jeunes que la moyenne (avec une plus grande proportion d'étudiants), alors que les personnes combinant voiture et transport en commun dans un même déplacement sont plus âgées (avec une plus grande part de retraités). La composition du ménage et les outils de mobilité disponibles diffèrent aussi entre les combinaisons. En particulier, un plus fort accès à l'automobile et un moins grand nombre de personnes abonnées au transport collectif sont repérés parmi les personnes combinant la voiture et le transport en commun (en comparaison aux autres combinaisons).

Des **variables spatiales** sont également corrélées à la multimodalité : les personnes qui résident dans des zones géographiques différentes ont tendance à avoir des comportements différents. Plusieurs travaux ont montré que la multimodalité des comportements était encouragée dans les centres urbains à forte densité de population, en particulier lorsqu'un accès à un réseau de métro est fourni (Buehler & Hamre, 2014, 2015; Diana, 2012; Heinen & Chatterjee, 2015). Nobis (2007) a aussi constaté que le pourcentage de personnes multimodales augmentait avec la taille de la ville (mais aussi probablement le nombre d'alternatives présentes). De même, d'après les données des enquêtes nationales allemandes, la proportion des déplacements intermodaux est plus élevée dans les villes plus grosses (Gebhardt et al., 2016). De plus, les différents types de combinaisons de modes ne sont pas distribués de manière uniforme dans la ville (Oostendorp & Gebhardt, 2018).

Les **variables des déplacements** peuvent aussi participer à prédire l'intermodalité. En particulier, les déplacements multimodaux ont tendance à être plus longs que les déplacements monomodaux, à la fois en termes de distance et de durée (Gebhardt et al., 2016; Kagerbauer et al., 2015). Le déplacement doit être suffisamment long pour qu'un ou plusieurs changement(s) de mode vaille(nt)

la peine (en termes de temps total de trajet). En effet, la pénibilité des correspondances est un frein majeur à la combinaison de différents modes de transport dans un même déplacement. Dans l'enquête réalisée par Oostendorp et Gebhardt (2018), les participants ont dû jauger dans quelle mesure certains critères liés aux correspondances étaient importants pour eux (sur une échelle de 5 points). Trois critères sont ressortis : la fiabilité du transfert, le temps d'attente et la distance à parcourir dans le changement de mode.

Les **variables socio-psychologiques**, encore peu étudiées pour la question de la multimodalité des comportements, ont aussi probablement un impact sur la mobilité multimodale. Par exemple, Molin et al. (2016) ont montré que les différents groupes d'utilisateurs multimodaux produits étaient caractérisés par des attitudes et des perceptions différentes.

Parmi les **autres variables** pouvant influencer la multimodalité des comportements individuels, on peut citer l'accessibilité à des moyens d'informations. En effet, dans l'enquête de Hilgert et al. (2017), les utilisateurs d'applications mobiles affichent une plus grande variété de modes utilisés : ils sont plus multimodaux, plus flexibles dans leur choix modal, moins dépendants à l'automobile et plus ouverts aux nouveaux services de mobilité. En outre, la réglementation en matière de correspondance et le manque de stationnement peuvent être un frein à la combinaison du vélo et du transport en commun dans un même déplacement : les vélos ne sont pas toujours autorisés à bord des bus ou dans les métros/trains. Même lorsque le vélo respecte les restrictions en matière de poids et de taille d'un bagage à main (par exemple un vélo pliant), il peut être encore difficile de le rentrer dans des véhicules bondés (Pan et al., 2010).

En remarquant ces relations entre le comportement modal, les caractéristiques socio-démographiques et l'environnement urbain, plusieurs auteurs ont parlé de « **modality styles** » (styles de modalités) dans la littérature de la multimodalité (Carrel et al., 2011; Heinen & Chatterjee, 2015; Lavery et al., 2013; Olafsson et al., 2016; Vij et al., 2011). Ce concept suppose la prédisposition d'une personne à choisir un ensemble de modes de déplacement en fonction de ses habitudes et de son mode de vie. C'est pourquoi différents travaux ont cherché à classer les usagers en fonction de leur utilisation de différents modes plutôt que d'un seul mode, puis à analyser les caractéristiques associées à chaque groupe (Buehler & Hamre, 2015; Nobis, 2007; Vij et al., 2013). Pour aller plus loin, certains travaux différencient les usagers en fonction de leur usage actuel des modes de transport mais aussi en fonction de leurs perceptions ou attitudes envers les

autres modes (Diana & Mokhtarian, 2009; Lavery et al., 2013; Molin et al., 2016). Les résultats de ces études ont montré que les automobilistes qui empruntaient aussi d'autres modes de transport et qui vivaient en zone urbaine dense avaient une perception moins biaisée et une attitude plus favorable à l'égard des modes alternatifs. En outre, les habitudes multimodales affectent de manière positive la volonté de changer de mode, même vers un service de transport non existant faisant appel à des concepts nouveaux (Diana, 2010). Par conséquent, plusieurs auteurs croient en l'importance de changer de perspective dans les politiques stratégiques de transport, en encourageant non plus les usagers à remplacer l'utilisation d'un mode de transport par un autre, mais plutôt en les incitant à changer leur mode de vie au complet et à utiliser un ensemble de modes de transport (Fuller et al., 2013; Heinen & Chatterjee, 2015). De même, ces « styles de mobilité » pourraient être utilisés dans les modèles de choix comportementaux. Cela permettrait d'éviter le recours à des données attitudinales, plus difficiles à collecter et à analyser (Diana & Pirra, 2016).

2.3.3.3 Facteurs ayant un effet sur la demande multimodale au niveau systémique

Tous les facteurs précédemment évoqués (au niveau individuel) ont une incidence sur le choix modal et la multimodalité des personnes : ils ont donc indirectement un effet sur la répartition modale de la demande à une échelle plus macroscopique. Il suffit de regarder l'influence de ces facteurs sur les parts modales de plusieurs villes pour s'en rendre compte (Santos et al., 2013). Aussi, d'autres recherches agrégées réalisées au sein d'une même ville ont permis de mettre en évidence des variables exogènes ayant un effet sur la demande globale et sur les interactions entre les modes de transport au niveau systémique.

En particulier, les impacts d'une augmentation de l'offre de transport sur la demande ont fait l'objet de multiples recherches. Toutefois, c'est dans le secteur routier qu'elles ont principalement été menées, pour évaluer l'effet d'un ajout de capacité. Ainsi, plusieurs auteurs (Cervero, 2002, 2003; Cervero & Hansen, 2002; Hymel et al., 2010; Weis & Axhausen, 2009) se sont intéressés au phénomène de la demande induite, qui peut prendre la forme d'une demande redistribuée (redistribution spatiale et temporelle des déplacements) et d'une demande générée. Cette dernière provient de différents changements : suite à une augmentation de l'offre, de nouveaux déplacements peuvent être créés, des déplacements plus longs peuvent être réalisés, mais des transferts modaux peuvent aussi être déclenchés. Par ailleurs, ce phénomène est souvent dissocié en un effet à court terme (changements comportementaux) et un effet à long terme (changements

structuraux). Parmi les modèles utilisés pour les quantifier, on retrouve les méthodes d'inférence causale recensées dans la section 2.2.1.6 (avec des données de trafic avant/après sur une route améliorée par rapport à une route non améliorée comparable), ainsi que des modèles d'équations simultanées ou d'équations structurelles. Ce genre de méthodes est nécessaire afin de réduire les biais déjà évoqués dans l'effet endogène entre l'offre et la demande : des investissements peuvent à la fois induire la croissance de la demande et être engagés pour répondre à une augmentation de la demande (causalité inversée).

Cependant, malgré le fait que ce sujet soit largement théorisé dans le domaine routier, la littérature sur la demande induite (ou de manière plus générale sur l'impact d'une bonification de l'offre) est beaucoup moins documentée pour les autres modes de transport. En outre, la plupart des études réalisées ne permettent pas de déterminer la source des déplacements ajoutés, notamment la part qui est due aux reports modaux. En effet, peu de travaux ont évalué l'effet des interventions en transport sur les transferts modaux, et encore moins en dehors du domaine routier. Néanmoins, dans le contexte où le transport collectif est souvent présenté comme un moyen de réduire la congestion automobile, l'effet de l'augmentation des investissements dans le transport en commun sur la demande automobile a parfois été modélisé (Beaudoin et al., 2015; Beaudoin & Lin Lawell, 2018; Yang, J. et al., 2018). Un effet de substitution (report modal et diminution de la congestion automobile) à court terme et un effet d'équilibre (croissance de la demande automobile) à long terme ont été distingués. De plus, les impacts de l'implantation de nouvelles infrastructures pour un mode sur l'utilisation d'un autre mode ont été analysés dans les cas du transport en commun et du taxi (Li, M. et al., 2017), du vélo et du transport en commun (Cervero et al., 2013), ainsi qu'entre les modes émergents et les modes classiques (tous les exemples de la section 2.2.1 dans lesquels des données avant/après l'arrivée d'un nouveau mode étaient disponibles).

Par ailleurs, les transferts modaux sont compensés (la demande induite est diminuée) lorsque le mode dans lequel on a investi est finalement utilisé en complémentarité du mode potentiellement substitué, ce qui arrive notamment lorsque des infrastructures (par exemple, des stationnements) sont ajoutées pour le vélo (Geurs et al., 2016; Martens, 2007; Pucher & Buehler, 2009), mais aussi entre le taxi et le métro (Jiang et al., 2018; Li, M. et al., 2017) et dans tous les cas de combinaisons entre modes émergents et modes classiques vues précédemment. On revient ainsi à ce qui avait été mis en évidence dans la relation entre modes classiques et modes émergents et qui s'applique finalement à tous les modes : le transfert modal observé n'est jamais strict.

Outre ces changements permanents dans l'offre de transport, des changements temporaires peuvent également survenir et avoir une influence sur la demande et la répartition modale. Un des thèmes étudiés dans la littérature est l'impact des perturbations des services de transport collectif (retards, annulations ou fermetures sur une plus ou moins longue durée). Bien que peu de recherches aient été menées pour analyser les effets d'un problème survenu dans le réseau d'un certain mode sur l'utilisation des autres modes d'un système multimodal (Diab & Shalaby, 2018), les résultats de plusieurs enquêtes ont montré qu'une bonne partie des usagers du transport en commun transféraient vers un autre mode lors d'une interruption de service (Lin, T. Y.-T., 2017; Pnevmatikou et al., 2015; Saxena et al., 2019). Au niveau agrégé, la perturbation du service d'un mode de transport collectif (par exemple, le métro) entraîne une diminution des vitesses d'opération sur les lignes des autres services (par exemple, tramway et bus) (Diab & Shalaby, 2018) mais une augmentation des temps de congestion automobile (Anderson, 2014; Bauernschuster et al., 2017) et de l'utilisation des services de taxi (Lepage & Morency, 2021) et de vélopartage (Saber et al., 2018; Younes et al., 2019). Ces effets à court terme (transferts modaux en réponse à une perturbation donnée) peuvent s'annuler une fois le service revenu (Younes et al., 2019), mais la qualité du service peut aussi avoir un effet à long terme sur les comportements de déplacement : les usagers pourraient décider de ne plus emprunter le transport en commun et de changer de mode définitivement si la fréquence des perturbations devenait trop élevée (Lin, T. Y.-T., 2017).

La plupart du temps, des effets au niveau local seulement sont mesurés. Pourtant, les impacts de certaines mesures stratégiques ne peuvent pas être évalués localement, notamment lorsqu'une politique est appliquée à toute la ville, par exemple une modification des tarifs. Ce phénomène est généralement étudié selon le principe d'élasticité, d'après lequel une diminution du tarif va faire augmenter la demande pour un certain mode de transport, notamment en attirant les usagers d'autres modes de transport. Encore une fois, les recherches ont surtout porté sur les impacts sur un mode en particulier, mais des « élasticités croisées » entre plusieurs modes ont parfois été estimées (Litman, 2004; Paulley et al., 2006). Outre ces calculs d'élasticités, des méthodes statistiques de détection de points de rupture peuvent également être appliquées pour mesurer des changements structuraux dans les séries temporelles (Yeh & Lee, 2019). Contrairement aux méthodes d'inférence causale vues précédemment (section 2.2.1.6), ce genre de méthodes permet d'évaluer les effets d'interventions touchant une population entière (sans groupe contrôle).

Des facteurs temporels exogènes peuvent également impacter la répartition de la demande de transport et les interactions entre les modes. Peu nombreux sont les travaux qui ont analysé l'effet des conditions météorologiques sur plusieurs modes de transport en même temps. Néanmoins, il a été montré que la météo affectait peu le nombre total de déplacements alors qu'elle entraînait des transferts entre les modes (Sabir, 2011). Globalement, des températures plus élevées et des précipitations plus faibles conduisent à une augmentation de l'utilisation du vélo et du vélopartage, mais à une diminution de l'utilisation du transport en commun, du taxi et de l'automobile (Lepage & Morency, 2021; Morency et al., 2018; Sabir, 2011). Ainsi, des phénomènes saisonniers liés à la météo mais aussi aux habitudes de déplacement peuvent être détectés dans l'évolution de la répartition modale à différents horizons temporels : des transferts modaux entre la voiture, le transport en commun et le taxi vers des modes actifs ont potentiellement lieu en fonction des saisons (Bergström & Magnusson, 2003; Morency et al., 2018), mais aussi en fonction du type de jour (ouvrable versus non ouvrable) (Morency et al., 2018). De surcroît, différents types d'événements (programmés ou non) peuvent influencer l'utilisation des modes. En particulier, la pandémie de COVID-19 pourrait avoir eu un impact important sur la répartition modale : une revue de littérature spécifique sur le sujet sera présentée dans le Chapitre 8 (article n°6).

2.3.4 Synthèse spécifique

Finalement, deux grandes limites de la modélisation classique ont été mises en évidence. En premier lieu, les comportements modélisés pour un jour moyen (voire seulement l'heure de pointe) sont supposés fixes dans le temps. Ni les tendances ni les rythmes hebdomadaires ne sont considérés car ces informations ne sont pas disponibles dans les données d'enquêtes utilisées (transversales). En deuxième lieu, un nombre limité de modes est inclus dans ces modèles, et les interactions entre offre et demande de même que les relations entre les modes sont très peu prises en compte, ce qui ne permet pas d'évaluer les impacts de la plupart des stratégies d'intervention. Bien que la création d'un modèle complet de prévision de la demande de transport soit au-delà des objectifs de cette thèse, des modèles de séries temporelles, basés sur des données longitudinales, et des méthodes d'évaluation des impacts seront développés pour mieux comprendre la demande de transport multimodale et les interactions entre mobilité et systèmes de transport. Cela nécessitera de prendre en compte de nombreux facteurs qui peuvent avoir un effet local ou global sur l'utilisation des modes, aussi bien au niveau individuel qu'au niveau systémique.

CHAPITRE 3 DONNÉES ET MÉTHODOLOGIE GÉNÉRALE

Ce chapitre présente la démarche générale adoptée pour répondre aux besoins méthodologiques, analytiques et stratégiques mis en évidence grâce à la revue de littérature. Tout d’abord, un cadre conceptuel est défini afin de fournir une vision d’ensemble de toutes les interactions modales qui ont été abordées précédemment (section 3.1). Ces interactions peuvent être mesurées à partir de deux grands types de données : des données d’enquête ou des données passives. Les sources de données rassemblées dans cette recherche pour l’étude du cas de Montréal sont décrites et comparées dans la section 3.2. Enfin, la méthodologie générale utilisée pour analyser, expliquer et modéliser les différents types d’interactions est exposée dans la section 3.3. Les investigations proposées sont découpées en six articles.

3.1 Cadre conceptuel

Le cadre conceptuel permet de regrouper tous les types d’interactions entre modes en fonction de différents niveaux d’analyse et de différentes situations de complémentarité/compétition.

3.1.1 Différents niveaux d’analyse

Les interactions entre les modes de transport peuvent être analysées à différents niveaux selon les données disponibles. Avec des données d’enquêtes, des interactions au niveau individuel peuvent être étudiées, soit par individu (multimodalité des personnes), soit par déplacement (intermodalité ou multimodalité des déplacements). L’analyse porte alors sur la combinaison de modes utilisés par chaque individu pendant une période donnée ou sur la séquence de modes empruntés dans chaque déplacement. Bien que cette possibilité n’ait pas encore été avancée dans la littérature, la multimodalité des comportements de mobilité individuels pourrait aussi être examinée à un niveau plus agrégé, par ménage ou autre groupe de personnes, ou à un niveau intermédiaire entre le déplacement et l’individu, par exemple par chaîne de déplacements. Avec des données passives, les interactions sont explorées au niveau systémique, soit de manière agrégée pour le système complet, soit de manière désagrégée par déplacement (ou segment de déplacement). À l’échelle du système, des corrélations d’usage entre les modes peuvent être mesurées. Au niveau du déplacement, il s’agit d’évaluer si l’utilisation d’un mode est faite en complémentarité ou en

concurrence avec un autre mode. Différentes situations de complémentarité ou de compétition seront discutées dans la section suivante.

En outre, ces interactions varient dans le temps et dans l'espace. Plusieurs des travaux précédemment recensés ont montré qu'il était nécessaire de distinguer des effets à court et à long terme, ou des changements temporaires versus permanents. Les impacts observés directement après un changement dans l'offre de transport ou l'implantation d'un nouveau service de mobilité peuvent finalement changer de sens (Beaudoin & Lin Lawell, 2018; Singleton & Clifton, 2014), stagner dans le temps (Cervero et al., 2007) ou même s'annuler et revenir au point de départ (Younes et al., 2019). Par conséquent, l'horizon temporel doit être clairement précisé. De même, l'effet mesuré peut être différent selon la localisation et l'environnement urbain. Des interactions différentes ont par exemple été mesurées entre différentes villes selon leur taille ou leur densité de population (Beaudoin & Lin Lawell, 2018; Hall et al., 2018; Shaheen et al., 2013), de même que des effets locaux ont pu être distingués, par exemple entre le centre-ville et la périphérie rurale (Ma, T. & Knaap, 2019; Martin & Shaheen, 2014; Tang et al., 2021). La composante spatiale, qui est parfois négligée par manque d'informations, doit donc aussi être considérée dans l'évaluation des relations entre les modes.

3.1.2 Différentes situations de complémentarité/compétition

Différents types d'interactions peuvent également être définis en fonction des impacts sur chaque mode, qui sont soit positifs (+), négatifs (-) ou neutres (0). La terminologie proposée est basée sur une analogie avec les six types d'interactions biologiques étudiés dans les processus écologiques (Odum & Barrett, 2005). Ces dernières, précisées dans le Tableau 3.1, sont les différentes interactions possibles entre deux espèces qui coexistent dans un même écosystème. Cette typologie d'interactions a déjà été transposée aux interactions entre différentes technologies (Sandén & Hillman, 2011) et est ici appliqué dans un contexte de transport pour décrire les interactions entre deux modes de transport A et B au niveau systémique. Bien que les interactions modales puissent être exprimées en termes de ressources (concurrence en termes d'espace, d'infrastructures, d'énergie consommée, etc.), elles sont ici définies à partir des impacts sur les parts de marché et l'achalandage de chaque mode. Ainsi, un signe (+) indique que l'interaction est bénéfique pour le mode (augmentation de l'utilisation du mode), un signe (-) signale une interaction néfaste (diminution de l'utilisation du mode) et un signe (0) est synonyme d'aucun changement.

En outre, comme pour les interactions entre les technologies, il faut distinguer des interactions dynamiques (à long terme) et des interactions quasi-statiques (à court terme).

Tableau 3.1 Interactions biologiques appliquées à deux modes de transport A et B en fonction de l'impact sur leur utilisation respective : positif (+), négatif (-) ou neutre (0)

INTERACTION	Impact sur le MODE A (ancien mode)	Impact sur le MODE B (nouveau mode)
Compétition	-	-
Mutualisme	+	+
Neutralisme	0	0
Parasitisme	-	+
Commensalisme	0	+
Amensalisme	0	-

Une interaction dynamique fait référence à l'arrivée d'un nouveau mode B (par exemple, un service de vélopartage) lorsqu'un ancien mode A (par exemple, le transport en commun) est déjà bien établi, ou à des changements à long terme dans les parts de marché de chaque mode. Le mode B peut attirer la part de marché du mode A, entraînant un transfert modal entre les deux modes, également appelé **parasitisme** en biologie, ou cannibalisation en économie (Moorthy & Png, 1992). Cependant, le mode B peut bénéficier d'un marché de niche créé par le mode A sans impacter l'achalandage de l'ancien mode. Par exemple, les taxis collectifs à la demande (mode B) peuvent être utilisés pour le transport adapté, un type particulier de transport en commun (mode A). Dans ce cas, il s'agit d'une situation de **commensalisme**. L'**amensalisme** se produit si le marché est fermé par le mode A (car tous les usagers de ce mode sont déjà satisfaits) ou si le mode B souffre de la mauvaise réputation du mode A. De cette façon, le mode B est inhibé, tandis que le mode A n'est pas affecté. Il y a **mutualisme** lorsque les deux modes (par exemple, le transport en commun et le vélopartage) se complètent (le mode B comble les lacunes du mode A en matière de mobilité) et créent un nouveau marché en provoquant un transfert modal depuis un autre mode (par exemple, la voiture particulière) ; c'est la situation la plus souhaitable pour une transition vers des modes de vie moins auto-dépendants. Le mode B peut également avoir une relation neutre (**neutralisme**) avec le mode A s'il développe un marché indépendant (en faisant concurrence à un autre mode ou en générant une nouvelle demande) sans interagir avec le mode A. Enfin, deux technologies émergentes peuvent également se disputer un marché commun (**compétition**).

Les interactions quasi-statiques peuvent prendre la forme de l'un des trois types d'interactions symétriques (concurrence, mutualisme ou neutralisme) et correspondent à une situation quasi-stable à court terme (le nouveau mode B est établi depuis un certain temps, et le marché partagé par les deux modes est supposé être presque fixe). Les deux modes peuvent évoluer indépendamment sans s'influencer mutuellement (pas de chevauchement de la demande) ; c'est le cas du **neutralisme**. Les deux modes peuvent également être en concurrence pour les mêmes clients (**compétition**). Dans ce cas, lorsque les individus se déplacent, ils doivent choisir entre le mode A et le mode B, en fonction de la disponibilité du service (pour une heure et un lieu spécifiques), de la qualité du trajet et de leurs préférences. Au contraire, les deux modes peuvent être utilisés en complémentarité, jusqu'à peut-être devenir mutuellement dépendants (**mutualisme**). Les deux modes peuvent notamment être combinés au sein d'un même déplacement, avec un mode principal et un autre mode utilisé pour faire la connexion du premier/dernier kilomètre ; on revient alors à la notion d'intégration modale ou d'intermodalité, qui a un impact positif sur l'utilisation de chaque mode. En réalité, les deux types d'interactions (compétition et mutualisme) se produisent à l'échelle du déplacement et leur cumul au niveau systémique tendra soit vers une augmentation, une diminution ou aucun changement dans l'utilisation des modes.

Toutefois, d'autres types de complémentarité peuvent exister sans affecter la demande du mode A, lorsque le mode B complète le mode A dans des trajets différents (indépendants) : on glisse alors vers du **commensalisme**. Ce sera notamment le cas si le trajet n'est pas réalisable avec le mode A parce que l'origine et la destination sont en dehors de la zone de desserte (complémentarité dans l'espace) ou que le début du trajet est en dehors des horaires de service, par exemple la nuit (complémentarité dans le temps). De plus, si l'alternative avec le mode A n'est pas compétitive avec celle du mode B, c'est-à-dire si le trajet n'est pas réalisable dans une durée ou une distance raisonnable, ou si l'utilisation du mode A n'est pas confortable, par exemple parce que le service est saturé, on pourra également parler de complémentarité dans la qualité du service. Ces trois types de complémentarité (dans l'espace, le temps et la qualité) représentent des situations de commensalisme à court terme mais pourront se transformer en mutualisme à long terme. En effet, en complétant les lacunes spatio-temporelles et/ou en déchargeant le réseau du mode A (synergie entre les deux modes), le mode B pourra rendre le mode A plus fiable et attirera potentiellement de nouveaux usagers vers le mode A.

3.2 Données utilisées

Les principales données utilisées dans cette recherche sont les enquêtes ménages déplacements (dites enquêtes Origine-Destination) de Montréal et des données passives sur l'achalandage de différents modes de transport.

3.2.1 Enquête Origine-Destination de Montréal

L'enquête Origine-Destination de Montréal est une enquête téléphonique (et web depuis 2018) organisée tous les cinq ans dans la région métropolitaine de Montréal. Le territoire de la dernière enquête (2018) est représenté dans la Figure 3.1. Cette enquête permet de recenser les déplacements d'environ 4 à 5% des ménages résidents au cours d'une journée moyenne d'automne. Les ménages sont échantillonnés aléatoirement (échantillonnage stratifié sans remplacement) chaque jour sur une période de quatre mois, entre septembre et décembre. Chaque répondant interrogé doit fournir les déplacements de la veille ainsi que les informations sociodémographiques de tous les membres du ménage. Les carnets de déplacements sont finalement fusionnés de sorte à représenter une « journée type » et les enregistrements sont pondérés afin de permettre une inférence statistique à l'ensemble de la population de référence (Secrétariat à l'enquête Origine-Destination, 2010, 2015, 2020). Pour ce faire, des facteurs de pondération sont d'abord calculés par ménage à l'aide d'une méthode de calage sur marge basée sur la strate géographique du lieu de domicile, la taille et la composition (par groupe d'âges) des ménages, les marges (totaux par catégorie) étant tirées des données du recensement canadien. Ensuite, les facteurs s'appliquant aux personnes sont déterminés en ajustant les facteurs des ménages pour tenir compte des cohortes d'âges et du sexe. Les déplacements sont également estimés via les facteurs des personnes. En 2018, les facteurs ont aussi été corrigés pour compenser le mois de septembre (qui a été sous-échantillonné), en particulier afin de mieux décrire l'utilisation du vélo.

Dans cette recherche, les trois dernières enquêtes sont exploitées : 2008, 2013 et 2018 (77 095, 78 731 et 75 266 ménages respectivement). Les déplacements ont été collectés à la fois en semaine et en fin de semaine en 2008 (grâce à deux enquêtes indépendantes basées respectivement sur 66 124 et 10 971 ménages), mais seuls les jours de semaine sont disponibles dans les enquêtes de 2013 et 2018. Les modes de transport suivants sont distingués dans la codification des segments de déplacement : auto conducteur, auto passager, taxi, métro, autobus STM (Montréal), autobus

STL (Laval), autobus RTL (Longueuil), autobus CIT/OMIT/MRC (Couronnes) - ou Exo dans l'enquête de 2018 -, autobus scolaire, autre bus, train, vélo, marche, moto, mode interurbain, transport adapté, autre. Toutefois, à chaque déplacement a été associée une classe modale, qui dépend du mode dominant dans la séquence de déplacements : AC (Auto-conducteur), AP (Auto-passager), TC (Transport en commun), BS (Bus scolaire), MV (Marche Vélo), PR (Park and Ride), KR (Kiss and Ride), AJ (Autre jonction) ou AU (Autre).

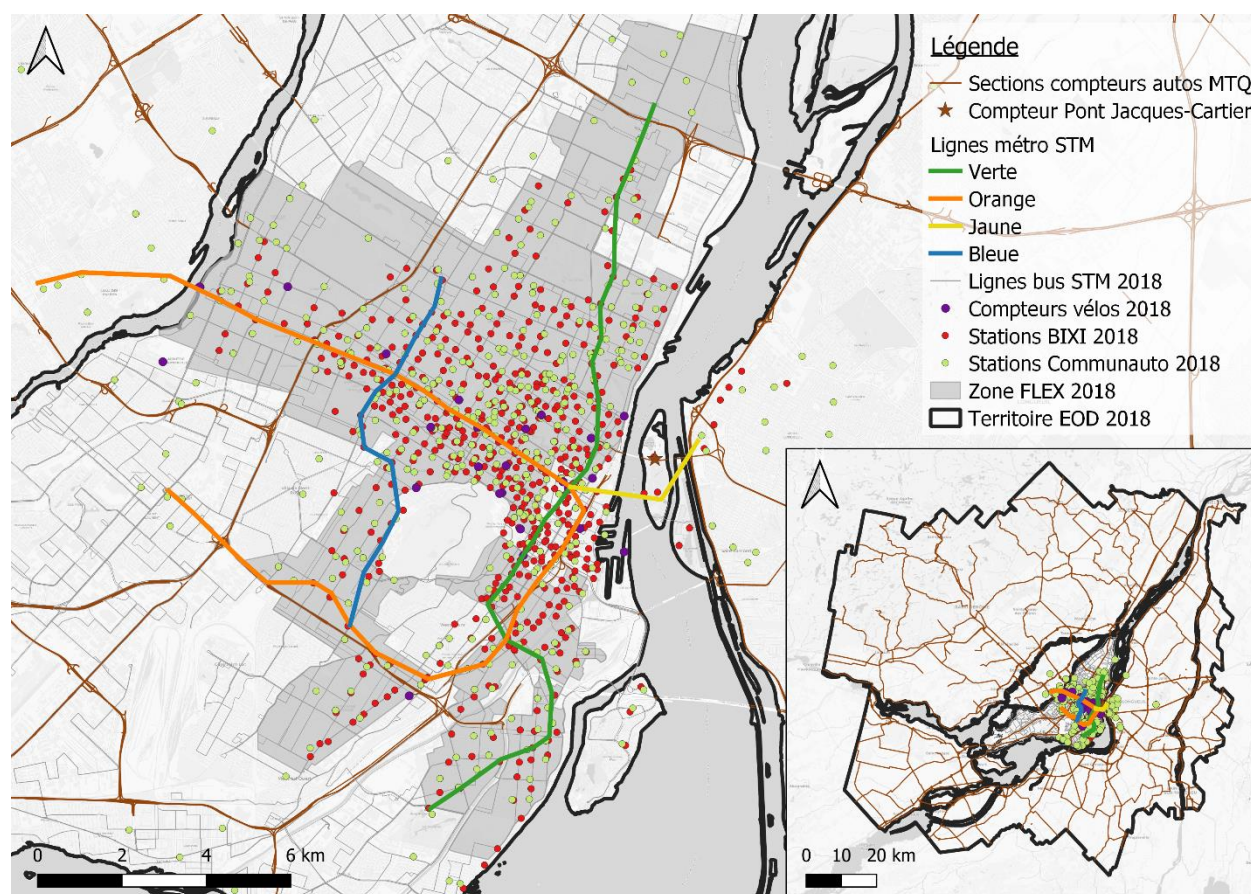


Figure 3.1 Étude du cas de Montréal : territoire de l'enquête Origine-Destination et des sources de données passives

3.2.2 Données passives

Le Tableau 3.2 décrit les données passives rassemblées pour cette recherche, chaque base de données se rapportant à des modes de transport particuliers. Ces données ont pour la plupart été fournies par les opérateurs des services sous l'égide de différentes ententes de confidentialité, mais certaines données (marquées par une croix dans la quatrième colonne du Tableau 3.2) sont

également publiques. Les périodes sur lesquelles les données sont disponibles sont également indiquées (jusqu’à fin 2021, les données de 2022 n’ayant pas pu être considérées pour cette thèse), mais les échantillons utilisés dans chaque article seront précisés dans les chapitres suivants.

Tableau 3.2 Description sommaire des données passives utilisées pour chaque mode

Mode	Types de données	Fournisseur	Données ouvertes	Période de couverture
Automobile privée	Comptages classifiés	MTQ		2013-01-01 à 2018-12-31
		PJCCI		2018-01-01 à 2021-12-31
Transport en commun (métro et bus)	Validations (embarquements seulement)	STM		2013-01-01 à 2021-12-31
Vélo personnel	Comptages du système Eco-compteur	Ville de Montréal	x	2009-01-01 à 2021-12-31
Taxi	Données transactionnelles (OD des courses)	Taxi Diamond		2014-03-01 à 2016-12-31
Vélopartage	Données transactionnelles (OD des emprunts)	BIXI	x	2014-04-15 à 2021-11-15
Autopartage basé stations et en libre-service	Données transactionnelles (OD des emprunts)	Communauto		2018-01-01 à 2021-12-31

La majorité des bases de données recensées sont des données transactionnelles. En particulier, ce projet bénéficie de toutes les validations (quel que soit le support : carte à puce, carte magnétique ou ticket papier) recueillies entre janvier 2013 et janvier 2022 sur le réseau de la STM. Ce réseau, composé de 5 lignes de métro et d’environ 200 lignes de bus, est représenté sur la Figure 3.1. Les validations, collectées (à l’embarquement seulement) via un système de perception automatique de titres de transport appelé le système OPUS, ont été transmises aux auteurs dans un format agrégé et anonymisé. Les autres sources de données transactionnelles amassées retracent les courses de taxi de l’intermédiaire Taxi Diamond, ainsi que les emprunts réalisés dans les services de vélopartage (BIXI) et d’autopartage (Communauto). Les stations BIXI, de même que les stations et la zone FLEX (véhicules partagés en libre-service) de Communauto, sont localisées sur la Figure 3.1 (à partir des données de 2018).

En revanche, seuls des comptages sont disponibles pour évaluer l’utilisation de l’automobile et du vélo. Pour le vélo, les informations proviennent du système Eco-compteur de la Ville de Montréal. Pour l’automobile, des données ont été obtenues via deux fournisseurs : le MTQ (système CIR) pour les compteurs situés sur les autoroutes et les ponts gérés par le Ministère, ainsi que la société

PJCCI pour les compteurs du pont Jacques-Cartier. Ces différents compteurs sont identifiés dans la Figure 3.1. Chaque compteur du MTQ est associé à une section de trafic (unidirectionnelle ou bidirectionnelle) qui peut être longue de plusieurs kilomètres, mais la position exacte du compteur le long de cette section n'est pas connue. Un seul point est représenté sur le pont Jacques-Cartier, mais différents compteurs sont en réalité répartis sur toutes les voies à l'entrée et la sortie du pont. Tous les comptages sélectionnés dans cette recherche sont des compteurs classifiés, c'est-à-dire qu'ils permettent de distinguer les voitures particulières (sans les camions).

3.2.3 Comparaison des deux principales sources de données

Ces deux sources de données (enquêtes Origine-Destination et données passives) présentent de nombreuses différences. Plusieurs de leurs caractéristiques sont comparées dans le Tableau 3.3. Tout d'abord, des divergences existent au niveau de la population de référence : l'enquête Origine-Destination s'adresse seulement aux résidents de la région métropolitaine, alors que les données passives capturent les déplacements de toute la population présente à Montréal et faisant l'usage d'un des modes (personnes résidentes et touristes). Toutefois, le territoire sur lequel les données passives ont été recueillies dépend de la zone de desserte du mode de transport considéré ou de la zone de couverture du système de collecte (par exemple, pour les systèmes de compteurs). La taille de cette population collectée n'est pas non plus la même : seul un échantillon plus ou moins représentatif (mais pondéré) de la population est connu avec l'enquête Origine-Destination tandis que les données transactionnelles passives donnent généralement accès à des inventaires complets. Néanmoins, les données de comptage peuvent ne pas être représentatives de l'utilisation totale des modes comptabilisés puisque l'information collectée dépend de l'emplacement des compteurs.

Par ailleurs, les données des enquêtes Origine-Destination sont des données transversales puisque les déplacements ne sont collectés que sur une seule journée d'automne tous les cinq ans. Au contraire, les données passives sont des données longitudinales dont on peut suivre l'évolution de manière continue. En outre, l'enquête Origine-Destination permet d'enquêter plusieurs modes de transport dans la même base de données (données multimodales) alors que dans les données passives, une base de données est généralement associée à un seul mode. La codification des modes de transport de l'enquête Origine-Destination ne correspond pas toujours non plus aux données passives. En particulier, les nouvelles mobilités partagées ne sont pas différenciées des modes classiques dans les enquêtes Origine-Destination. Un enjeu lié aux indicateurs d'intensité

d'utilisation de ces modes est également à signaler : l'unité de mesure (ou métrique), uniforme dans l'enquête Origine-Destination (déplacement), présente au contraire différentes formes dans les données passives (transaction, compte, course, réservation, etc.), menant à un problème de commensurabilité des données.

Toutes ces différences rendent a priori ces deux sources de données incomparables. Toutefois, ce projet mettra en lumière des similarités dans les comportements de mobilité enregistrés et permettra de combiner les avantages de chaque source, notamment la longitudinalité des données passives et la multimodalité des enquêtes Origine-Destination. Ainsi, les deux types de données seront utilisés en complémentarité dans un processus de fusion de données.

Tableau 3.3 Comparaison de l'enquête Origine-Destination et des données passives

		ENQUÊTE OD	DONNÉES PASSIVES
Population collectée		Échantillon pondéré Personnes résidentes du territoire d'enquête	Inventaire complet Personnes résidentes et touristes
Territoire		Région métropolitaine de Montréal	Zone de desserte du mode de transport ou zone de couverture du système de collecte
Période d'observation		1 journée d'automne (semaine)	Continue
Modes de transport	Nombre	Plusieurs	1 seul
	Types	Modes classiques	Modes classiques et émergents
Indicateur d'utilisation		Déplacement	Transaction, compte, etc.

3.2.4 Autres données

Par ailleurs, d'autres bases de données sont également exploitées dans ce projet pour contextualiser les analyses ou servir de variables explicatives dans les modèles. Des informations sur l'offre de transport, telles que la description des infrastructures des différents systèmes de transport ou de la fréquence des services du transport en commun (fichiers GTFS), ont été réunies. Ces données ont été récupérées en ligne ou fournies directement par les opérateurs. De plus, les données du recensement canadien (Statistiques Canada) ont permis de compléter les enquêtes Origine-Destination, notamment pour avoir des informations sur les revenus des ménages. Les fichiers du réseau routier du recensement ont également été téléchargés. En outre, plusieurs sources de données géoréférencées ont été rassemblées, notamment les données d'utilisation du sol de la

Communauté métropolitaine de Montréal (CMM), qui attribuent à chaque parcelle du sol une fonction parmi 11 classes d'utilisation du sol, ainsi que les données CanMap® GIS de l'entreprise DMTI Spatial, qui comprennent des renseignements sur divers points d'intérêts. Des informations sur le calendrier civil ont été codées manuellement, et des informations sur les conditions météorologiques ont été tirées des bases de données climatiques historiques d'Environnement Canada (à partir de la station météo de l'aéroport international Pierre-Elliott-Trudeau). Une base de données recensant les événements socio-culturels ayant eu lieu à proximité des stations de métro entre 2015 et 2017 a également été fournie par la STM. Enfin, des informations sur la pandémie de COVID-19 ont été obtenues sur le site de l'Institut national de santé publique du Québec (INSPQ).

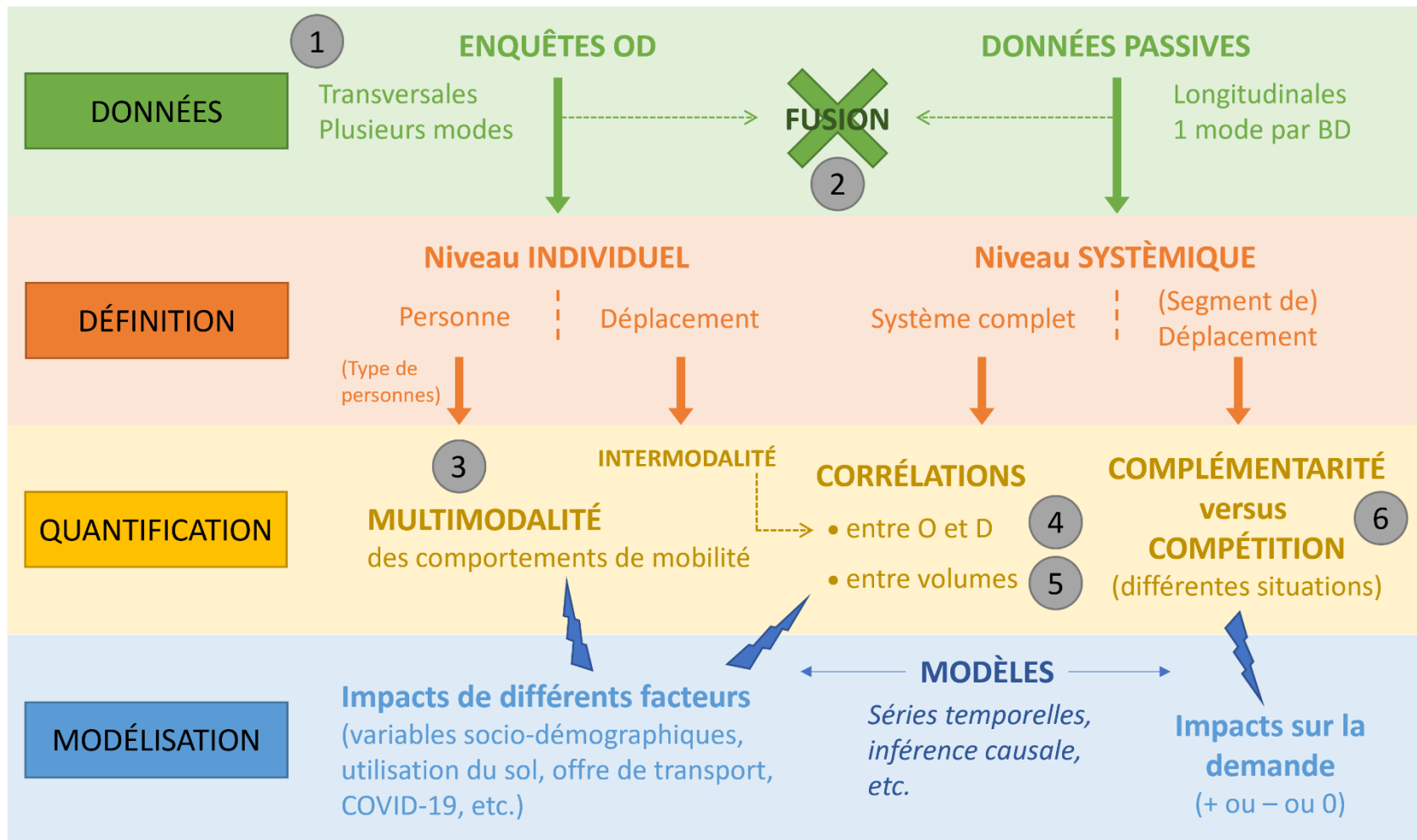
3.3 Démarche méthodologique

La démarche méthodologique générale de cette recherche est schématisée dans la Figure 3.2. Plutôt que d'être considérées séparément, les deux principales sources de données décrites précédemment (enquêtes Origine-Destination et données passives) sont d'abord utilisées ensemble. Dans un premier temps, les deux types de données sont comparés afin d'identifier des points communs dans l'analyse des comportements d'utilisation de plusieurs modes de transport (**article 1**). Cette comparaison permet également de préciser le rôle complémentaire de chaque type de données dans l'étude des interactions entre les modes. Dans un deuxième temps, une méthode de fusion des deux types de données est développée afin de suivre l'utilisation de différents modes dans le temps (**article 2**). Les résultats obtenus permettent ainsi des analyses longitudinales et multimodales.

Ensuite, les deux sources de données sont valorisées pour qualifier et quantifier différents types d'interactions. Les enquêtes Origine-Destination permettent d'examiner des interactions au niveau individuel, par type de personnes (**article 3**). Le niveau de multimodalité des comportements de mobilité et son évolution à moyen-long terme (sur cinq ans) sont évalués. Les données passives sont quant à elles exploitées pour des analyses au niveau systémique à moyen-court terme. Des interactions sont d'abord examinées au niveau agrégé : des corrélations sont mesurées dans l'utilisation des modes, localement entre les destinations et les origines pour estimer de potentielles relations intermodales (**article 4**), ou globalement entre les volumes (**article 5**) de plusieurs modes. Une étude plus désagrégée à l'échelle du déplacement est également réalisée afin de détecter différentes situations de complémentarité ou de compétition entre deux modes (**article 6**).

Modélisation des interactions entre les modes de transport par l'intégration de sources diversifiées de données

6 articles



OD = Origine-Destination; BD = Base de données

Figure 3.2 Démarche méthodologie générale de la thèse et positionnement des 6 articles

Les effets de différents facteurs (variables socio-démographiques, utilisation du sol, offre de transport, etc.) sur la multimodalité des comportements et les corrélations d’usage entre les modes sont également évalués (articles 3 et 4). En particulier, l’impact de la pandémie de COVID-19 sur les interactions entre les modes est exploré dans l’article 5. Par ailleurs, l’influence de l’intermodalité et de la compétition entre les modes sur la demande de transport est modélisée dans les articles 4 et 6. Tous ces effets sont mesurés à l’aide de méthodes longitudinales et de techniques d’inférence causale. Des effets locaux (sur certains individus ou certaines composantes du réseau par rapport à un groupe contrôle) mais aussi des effets systémiques (par exemple, les impacts de la crise sanitaire) sont investigués.

La démarche méthodologique de chacun des six articles est décrite plus précisément ci-après.

3.3.1 Article 1 : Comparaison entre enquête OD et données passives

L’article 1, intitulé “Cross-analysis of the variability of travel behaviors using one-day trip diaries and longitudinal data”, est présenté au Chapitre 4. Il a été publié le 7 septembre 2022 dans le journal *Transportation Research Part A*. Cet article vise à valoriser deux types de données (enquêtes ménages déplacements et flux passifs) pour comparer les variations temporelles systémiques dans les comportements d’utilisation de trois modes de transport : l’automobile privée, le métro et le vélo. Ces analyses sont réalisées au niveau agrégé (utilisation des modes par jour) puisqu’il n’est pas possible de suivre des individus dans les données de comptages (automobile et vélo) et de validations (métro) exploitées.

Tout d’abord, une méthode de repondération de l’enquête Origine-Destination de Montréal (décrite dans la Figure 4.3) est développée afin de considérer chaque jour d’enquête comme une enquête à part entière. Cette segmentation des journées est possible grâce au grand nombre d’observations collectées par jour. Les facteurs de pondération sont alors corrigés de sorte que l’échantillon de personnes enquêtées chaque jour soit représentatif de la population de référence. La méthode proposée permet ainsi de calculer un nombre de déplacements par mode par jour sur toute la période d’enquête. Par conséquent, elle permet ensuite de comparer les variations mesurées dans les données d’enquête à celles observées dans les données passives sur plusieurs mois d’automne. Pour ce faire, les déplacements, comptages ou validations sont sommés par jour, par mode et par base de données. Cette agrégation aboutit à la création de deux séries temporelles par mode, une à partir de l’enquête et une autre à partir des données passives.

Les comparaisons entre les deux types de données sont d'abord réalisées graphiquement sur deux mois et demi, ainsi qu'au cours d'une semaine moyenne, à l'automne 2013 et l'automne 2018. Ensuite, des indicateurs de comparaison sont proposés afin de quantifier trois types de phénomènes dans la structure des séries temporelles : la tendance, la variabilité par type de jour (pour tous les lundis, les mardis, etc.) et la périodicité des comportements. Ces indicateurs servent à estimer si les mêmes variations sont observées dans les deux types de données. Ils permettent des comparaisons systématiques et peuvent être appliqués quel que soit l'indicateur de volume dans les données (déplacements, comptages ou validations).

3.3.2 Article 2 : Fusion entre enquête OD et données passives

L'article 2, intitulé "Combining a regional household survey and passive data streams for longitudinal monitoring purposes", est exposé au Chapitre 5. Il a été présenté à la 12^{ème} conférence ISCTSC (International Conference on Transport Survey Methods) en mars 2022 puis a été accepté pour publication dans le journal *Transportation Research Procedia*. L'objectif de cet article est de fusionner l'enquête Origine-Destination de Montréal avec des données passives afin de permettre un suivi longitudinal de l'utilisation de plusieurs modes de transport (ainsi que des parts modales correspondantes) sur une période de cinq ans (entre deux enquêtes).

La méthode développée, basée sur la décomposition de séries temporelles, est schématisée dans la Figure 5.1. Elle consiste, pour chaque mode de transport, à appliquer les variations longitudinales détectées dans les données passives au jour moyen de semaine observé dans l'enquête Origine-Destination. Ces variations sont mises en évidence dans les données passives (sélectionnées de 2013 à 2018) à l'aide de l'algorithme STL, qui permet d'isoler la tendance long-terme et les saisonnalités (annuelle et hebdomadaire) dans les données de chaque mode. Ces différentes composantes sont ensuite normalisées par le jour moyen de semaine d'automne et appliquées à un indicateur d'utilisation modale moyen calculé à partir de l'enquête Origine-Destination, par exemple le nombre moyen de déplacements faits avec chaque mode.

Cette procédure automatisée est mise en pratique sur les données de l'enquête 2013 pour projeter le nombre de déplacements de trois modes (automobile privée, métro et vélo). La projection est validée à l'aide de l'enquête 2018. De plus, les résultats sont comparés à ceux obtenus avec les facteurs de projection fournis par le MTQ. Les parts modales interpolées entre 2013 et 2018, ainsi que les parts modales saisonnalisées de 2018, sont finalement estimées et analysées.

3.3.3 Article 3 : Multimodalité des comportements de mobilité

L'article 3, intitulé "Measuring changes in multimodal travel behavior resulting from transport supply improvement", est retranscrit au Chapitre 6. Il a été présenté à la 100^{ème} conférence annuelle du TRB (Transportation Research Board) en janvier 2021 et a été publié le 15 avril 2021 dans le journal *Transportation Research Record*. Cet article cherche à évaluer l'effet d'une bonification de l'offre de transport sur l'évolution du niveau de multimodalité des comportements de déplacement entre 2013 et 2018 à Montréal. Pour ce faire, des informations sur l'offre et les données des deux dernières enquêtes Origine-Destination sont exploitées.

La méthodologie proposée est découpée en trois grandes étapes (voir Figure 6.1). Dans un premier temps, une typologie de personnes est construite à partir de trois ensembles d'attributs : les caractéristiques du voisinage du lieu de domicile, les caractéristiques du ménage dans lequel vit la personne, et les caractéristiques de la personne elle-même (groupe d'âges et genre). Les participants des enquêtes OD 2013 et 2018 sont ainsi regroupés par combinaison de ces variables et par année. Cette agrégation permet de pallier la nature transversale des données en permettant de faire correspondre les mêmes types de personnes en 2013 et en 2018. Il s'agit d'une alternative (empirique) à la méthode d'appariement par score de propension permettant de comparer des personnes semblables à deux périodes différentes. En outre, cette agrégation permet également de compenser le manque d'observations sur l'utilisation des modes (chaque individu n'est observé que sur une seule journée dans l'enquête). Les intensités d'utilisation de plusieurs modes sont ici sommées par type de personnes.

Par conséquent, la multimodalité des comportements est mesurée par type de personnes plutôt que par personne. L'indicateur de multimodalité choisi (la variante de l'indice de Dalton proposée par Diana et Pirra (2016)) est appliqué sur une série d'intensités agrégées et pondérées par type. De plus, des analyses de sensibilité sont produites pour préciser l'effet de l'unité de l'intensité d'utilisation et de la classification des modes sur la mesure de la multimodalité. Finalement, l'évolution du niveau de multimodalité moyen entre 2013 et 2018 est étudié, et les variables explicatives de cette évolution sont identifiées. En particulier, l'impact causal d'une bonification de l'offre de transport entre 2013 et 2018 (en nombre de passages-arrêts de transport en commun cumulés sur 24 heures, en nombre d'ancrages de vélopartage et en nombre de véhicules d'autopartage dans les stations) est modélisé à l'aide de la méthode des doubles différences.

3.3.4 Article 4 : Interactions systémiques intermodales

L'article 4, intitulé "Modeling interactions between the Montreal subway and other urban transportation modes at the station level", est présenté au Chapitre 7. Il a été accepté pour publication dans le journal *Tunnelling and Underground Space Technology* et est disponible en ligne depuis le 23 mars 2023. Cet article propose d'analyser les interactions systémiques entre le métro de Montréal et l'environnement urbain (incluant les interactions avec d'autres modes de transport) au niveau des stations.

Un modèle multiniveau à effets mixtes (avec effets fixes et aléatoires) est développé pour modéliser le nombre de validations par jour dans le métro sur une période de deux mois. La Figure 7.2 illustre la structure hiérarchique du modèle en deux niveaux (61 observations journalières nichées dans chaque station), qui permet d'expliquer la variabilité inter- et intra-station. Ce modèle longitudinal estime l'influence de différentes variables spatio-temporelles (liées au calendrier civil, aux conditions météo, aux attributs socio-économiques des personnes vivant à proximité, à la connectivité au réseau, à l'utilisation du sol et à l'environnement bâti) sur l'utilisation de chaque station de métro. De plus, les corrélations entre les entrées dans le métro et les destinations de plusieurs modes de transport (vélopartage, autopartage avec véhicules en libre-service et taxi) sont mesurées afin de détecter des relations intermodales potentielles.

Le modèle est d'abord appliqué avec des variables continues variant en fonction de la journée et/ou décrivant le voisinage de chaque station. Par ailleurs, le modèle proposé est comparé à d'autres types de modèles de régression à l'aide d'un exercice de prévision. Ensuite, chaque jour est décomposé en plusieurs groupes d'heures et les stations sont regroupées par type de stations. Les variables catégorielles introduites dans le modèle permettent alors de préciser les interactions entre les modes en fonction de la période de la journée et de la position de la station dans le réseau.

3.3.5 Article 5 : Impacts de la pandémie de COVID-19

L'article 5, intitulé "Assessing the impacts of the COVID-19 pandemic on subway ridership and on the interactions with other transportation modes", est présenté au Chapitre 8. Il a été soumis pour publication dans le journal *Transportation* le 26 novembre 2022. Cet article vise à évaluer les impacts de la pandémie de COVID-19 sur l'achalandage quotidien du métro de Montréal et sur ses interactions avec quatre autres modes de transport : l'automobile privée, le vélo, l'autopartage basé

stations et l'autopartage avec véhicules en libre-service. La méthodologie mise en œuvre est constituée de trois étapes.

Des points de rupture liés à des changements structuraux dans les données (entre janvier 2019 et décembre 2021) sont d'abord détectés dans l'achalandage du métro en croisant les résultats de deux méthodes : une méthode de régression linéaire par morceaux et une méthode bayésienne. Les dates associées à ces points permettent de découper la série temporelle du nombre de validations par jour en plusieurs segments ou phases. Cette première étape sert à identifier quels mesures et événements ont eu le plus d'impacts sur l'utilisation du métro, et les effets correspondants sont quantifiés (direction, ampleur des impacts, et délai entre l'évènement et l'impact).

Ensuite, la variabilité des impacts par phase est examinée par type de jour (ouvrable versus non ouvrable) et par période de la journée en sélectionnant les sous-groupes de déplacements correspondants. Des analyses en temps et en fréquence sont également réalisées en traçant le spectrogramme de la série temporelle du nombre de validations par jour. Cette deuxième étape permet de préciser les changements dans les comportements de déplacement, notamment dans les rythmes journaliers et hebdomadaires typiques.

Finalement, les interactions systémiques entre le métro et les autres modes étudiés sont estimées à l'aide d'un modèle de série temporelle interrompue (avec prise en compte de l'autocorrélation des résidus). Les impacts directs des mesures et événements mis en évidence précédemment pour le métro sont évalués sur les autres modes (analyse croisée des impacts). De plus, la corrélation entre l'utilisation du métro et celle des autres modes est modélisée (via des termes d'interaction par phase introduits dans le modèle) afin d'explorer les changements dans les relations entre les modes.

3.3.6 Article 6 : Complémentarité et compétition au niveau du déplacement

L'article 6, intitulé "Assessing complementary and competing interactions between transit and shared transportation modes", est présenté au Chapitre 9. Il a remporté une bourse doctorale « Prestige » dans le cadre d'un concours de la conférence *World Conference on Transport Research Society (WCTRS Young Initiative II)* et il a été accepté pour présentation et recommandé pour publication dans le journal *Transportation Research Procedia* (article initialement soumis le 1er décembre 2022, puis révisé une première fois le 22 avril 2023). Cet article s'intéresse aux interactions quasi-statiques (à court terme) entre deux modes partagés (le vélopartage et

l'autopartage en libre-service) et le transport en commun (bus et métro) à partir de données transactionnelles totalement désagrégées sélectionnées sur une période d'un mois.

Pour ce faire, un algorithme (décrit dans la Figure 9.2) basé sur différentes règles et valeurs seuils est d'abord appliqué afin de classer chacun des déplacements (ou segments de déplacement) des données historiques du vélopartage et de l'autopartage dans différentes catégories de relations potentielles avec le transport en commun : relation neutre (pas d'interaction, compétition avec la marche), différents types de complémentarité (en temps, espace ou qualité du service), ou compétition. Cette classification dépend de la faisabilité spatio-temporelle du déplacement en transport en commun (déterminée par la couverture du réseau et les horaires planifiés), ainsi que de la qualité de l'alternative en transport en commun (en temps d'attente, en nombre de correspondances, en niveau de tortuosité et en niveau de compétitivité du temps total de trajet) estimée par le calculateur de chemins de la plateforme Transition.

Les déplacements de chaque catégorie d'interactions sont ensuite analysés dans le temps et dans l'espace afin de déterminer où et quand les deux modes partagés tendent à compléter ou substituer le transport en commun (selon les hypothèses de l'algorithme proposé). Au niveau temporel, des distributions par jour de la semaine et par heure de la journée sont produites. Au niveau spatial, des densités de paires Origine-Destination (ou lignes de désir) sont représentées par catégorie d'interactions à travers une grille hexagonale.

Finalement, l'impact de la compétition avec le vélopartage sur l'achalandage du transport en commun est modélisé par jour et par ligne à l'aide d'un modèle des doubles différences (avec doubles effets fixes pour tenir compte de la structure nichée des données). La période contrôle correspond à la fermeture de BIXI pendant l'hiver (après le 15 novembre). En outre, quatre définitions de traitement sont testées pour répartir les lignes de transport en commun en deux groupes (groupe traité et groupe contrôle). Celles-ci permettent de distinguer l'impact des déplacements en compétition de l'impact global de l'usage du vélopartage. La séquence de lignes qui aurait été empruntées en l'absence du service de vélopartage est également considérée dans une des quatre définitions.

CHAPITRE 4 ARTICLE 1 : CROSS-ANALYSIS OF THE VARIABILITY OF TRAVEL BEHAVIORS USING ONE-DAY TRIP DIARIES AND LONGITUDINAL DATA

Elodie Deschaintres, Catherine Morency, Martin Trépanier

Publié le 7 septembre 2022 dans *Transportation Research Part A : Policy and Practice*

Volume 163, Pages 228-246, DOI : 10.1016/j.tra.2022.07.013

Abstract

Many travel surveys used in transportation planning still collect information over a single weekday while mobility and activity patterns fluctuate from day to day. However, the emergence of new continuous data sources has led to the generation of large amounts of longitudinal data. Therefore, this paper aims to value these two data types to cross-analyze the temporal variability of aggregated travel behaviors. A weighting method based on large-scale single-day survey samples was first developed to extend the representativeness of the daily trip diaries available in the three most recent Origin-Destination household travel surveys of Montreal, Canada, to the entire data collection period (four months in the fall). This method was then validated using passive data streams (smart card and count data) for three modes: car, subway, and bicycle. The variations in the daily use of these modes measured independently by the surveys and passive data were compared over several months of the fall period and throughout an average week. Indicators were also proposed to evaluate whether the two data sources capture the same trend and seasonality, especially the weekly rhythms. The results show that it is possible to accurately infer day-to-day variability in travel behavior from a large-scale cross-sectional single-day survey. Furthermore, this paper demonstrates the potential for complementarity between traditional surveys and emerging data (at least over the survey period), as well as the possibility to combine them since there are large similarities in the aggregated travel patterns they allow monitoring.

Keywords: travel survey; longitudinal data; typical weekday; day-to-day variability; weighting method; data comparison

4.1 Introduction

Although new data sources have become available over the past few years, a lot of transportation-related decisions are still based on data from travel surveys. Indeed, they contain contextual and attitudinal variables that are generally not collected in other data sources (Bayart et al., 2009). However, most surveys are cross-sectional and travel information is often recorded for only one weekday. Such data is not enough to represent the complexity of travel and activity behaviors since their variability has been demonstrated, especially throughout the week (Verreault & Morency, 2011). Therefore, further investigation is needed to go beyond the concept of a “typical weekday”. For instance, longitudinal data could be used to enrich the daily information collected in travel surveys. In line with that, this work explores potential complementarity between traditional surveys and passive data. The two data types are compared in order to find similarities in observed travel demand and its fluctuations over time, despite all inherent differences between the datasets that will be discussed and that require data aggregation. Such an analogy could be further exploited for data fusion purposes, especially to legitimize the application of variability detected in passive data to survey data. This kind of complementarity would be very useful for transportation planners since longitudinal surveys over a week or more are expensive and time-consuming. Although the development of methods to merge traditional and emerging data is beyond the scope of this study, some approaches are suggested in the light of our findings.

More specifically, this paper aims to develop a weighting method to expand a household travel survey composed of single-day samples, and then to cross-analyze the observed variability in aggregated mobility behaviors using longitudinal data. To this end, the Origin-Destination travel survey of Montreal, Canada, and passive data streams are harnessed. The variations in the use of three modes (car, subway, and cycling), measured with the two data sources, are compared over several months in the fall and throughout an average week. Indicators are also calculated to evaluate whether the two data sources capture the same trend and seasonality, especially the weekly rhythms. As a result, the methodological contribution of this paper is twofold. First, a method for extending the use of large-scale single-day travel surveys at the longitudinal level is provided. Such a method is useful to make survey data comparable to passive data, and it could also allow transportation stakeholders to monitor changes in representative mobility behaviors over several months (at the aggregate level). Second, this paper proposes formalized indicators (instead of

descriptive and qualitative analyses only) which allow for systematic comparisons, and which can be applied regardless of the volume metric in the data (trips, validations, counts, or others). Therefore, the proposed framework makes it possible to cross-analyze variations in travel patterns using both traditional surveys and passive data, rather than using only one of these two data sources as is generally the case in the literature.

This paper is organized as follows. First, some background about the variability of mobility behaviors (limitedly) measured in survey data, the emerging data landscape, and the related data fusion issues are documented. The various datasets used in this work and their intrinsic differences are then described in the subsequent section. The weighting method used to broaden the representativeness of the travel survey over time is also presented, as well as the indicators proposed to compare the different data sources. The results are reported in three parts: expansion of the travel survey, graphical comparison with passive data and calculation of variability indicators. Finally, practical implications are discussed, and further research directions are proposed in the conclusion section.

4.2 Background

4.2.1 Variability of mobility behaviors in survey data

The analysis of the variability (or regularity) of mobility behaviors has received a lot of attention in the literature. Indeed, transportation planning and modeling have long been focused on the peak hours of an average working day, but the related concept of “typical” or “representative” weekday has been much debated in the past. Many researchers claimed that such a scale fails to capture variations in travel demand: see (Schönfelder & Axhausen, 2010) for a detailed review. In particular, the weekly variability of individual activity-travel patterns was examined in some studies with trip diaries spanning from one to six weeks. Mobility behaviors vary between persons, but the intra-personal day-to-day variations (within persons over time) explain a large part of the total variability observed over a week (Kang & Scott, 2010; Kitamura et al., 2006; Pas, 1987; Raux et al., 2016; Watanabe et al., 2021). Diverse approaches were investigated to evaluate the within-person variability of trip generation (Pas, 1987; Raux et al., 2016), travel habits (Kitamura et al., 2006; Minnen et al., 2015), action space (Susilo, Yusak O. & Kitamura, 2005), activity scheduling (Habib & Miller, 2008; Lee & McNally, 2003) and time allocation (Chikaraishi et al., 2010; Jara-

Díaz & Rosales-Salas, 2015; Kang & Scott, 2010). The difference between weekdays and weekends was largely established (Astroza et al., 2018; Kang & Scott, 2010; Schlich & Axhausen, 2003), but significant deviations were also reported between the weekdays (from Monday to Friday), even at a more aggregated level (Verreault & Morency, 2011). In fact, complementary relationships and competing trade-offs were found in activity scheduling and time use across the days of the week (Astroza et al., 2018; Watanabe et al., 2021).

All these studies demonstrate that one-day data are not enough to capture the work-leisure weekly cycles in activity-travel behaviors. However, many travel surveys are cross-sectional in nature, hence executed at a specific point in time, and trip diaries of the respondents are often collected over a single day. Although they are typically conducted over several weeks or months, large regional household travel surveys are still considered cross-sectional as the single-day samples are pooled to represent a “typical day”. They are sometimes called large-scale cross-sectional surveys (Ortúzar et al., 2011; Verreault & Morency, 2011) or repeated cross-sectional surveys (Miller et al., 2018), but they are not continuous since there is a gap of several years between two surveys. The underlying assumption of such surveys is that activity-travel patterns are similar from one day to another. Their variability, although amply demonstrated in the previous papers, is not considered. Therefore, the ability of traditional surveys to represent mobility behaviors is questioned. Moreover, topical issues such as measuring multimodality (i.e., variability in mode choice) need multiday information (Deschaintres et al., 2021). One-week (seven days) travel diary data is a benchmark (Kuhnimhof, 2009) while weighted two- and three-day diaries may be a good surrogate (Jara-Díaz & Rosales-Salas, 2015) to depict weekly cycles, but two weeks at least are required to better measure variations (Schlich & Axhausen, 2003).

In the past few years, there was a growing interest in analyzing activity-travel patterns that cover even longer periods of time (several years). Multiday panel surveys such as the German Mobility Panel (MOP) (Hilgert et al., 2018) and continuous travel surveys (El-Assi et al., 2020) were then used to study variability in mobility behaviors (e.g., seasonality and long-term trends); see a definition and review of these different types of longitudinal surveys in (Ortúzar et al., 2011) and (Ampt, 2013). However, the advantages of such surveys have not been sufficiently established for most analysis and modeling purposes (Miller et al., 2018), and the temporal variations of travel behavior may be only visible at the aggregated level (El-Assi et al., 2020). Furthermore, longitudinal surveys involve expensive and difficult collection processes. The loss of continuity

and motivation of the interviewers was one of the reasons argued to interrupt several ongoing surveys (Ortúzar et al., 2011). These surveys also face representativeness issues because the length of the surveys but also the exposure to multiple survey requests cause fatigue effects (Porter et al., 2004). This is traduced by less complete and accurate surveys as participants tend to underreport some specific trips, e.g., short walking trips and less complex chains (Chlond et al., 2013), and by rising immobility rates due to soft refusal, i.e., trips deliberately not reported (De Haas et al., 2017). Respondent fatigue also entails attrition in panel surveys, defined by the loss of participants over time, and this phenomenon is non-random since some people are more likely to drop out than others; thus this leads to self-selection bias (Hoogendoorn et al., 2018). Nevertheless, the emerging data landscape opens opportunities to overcome these limitations.

4.2.2 An emerging data landscape: opportunities and challenges

Recent technological developments have made new operational and collective data available in transportation, such as GPS, count, smartphone, social media and smart card or other transactional data (Cherchi & Bhat, 2018). These data are said passive as they gather information without the active participation of the subject, thus reducing the respondent burden. The information may be continuously collected over long periods of time, hence longitudinal. This leads to large databases, that is why they are also called Big Data. The latter are often characterized by 4 V's: volume, velocity, variety and veracity (Lovelace et al., 2016).

Many researchers see in this emerging data landscape an opportunity to complement (but not substitute) traditional data, especially surveys (Bonnell & Munizaga, 2018; Miller et al., 2018). Indeed, the most commonly shared view is that the two data sources (traditional surveys and emerging data) should be used together to maximize their respective value (Callegaro & Yang, 2018). The pros and cons of each data source highlight their complementary. Passive data, initially intended for functions other than the analysis and modeling of mobility behaviors, provide very precise spatial-temporal information. They are therefore particularly suitable for studying the variability of travel behaviors over long periods: see examples based on smart card data in the transit context (Briand et al., 2017; Deschaintres et al., 2019). However, passive data do not contain any contextual data (due to confidentiality and privacy concerns). Rather, traditional surveys include comprehensive information on the socio-demographic characteristics of the respondents, which are necessary to explain mobility behaviors (Bayart et al., 2009), but they are often based

on a single-day travel diary and cross-sectional, while emerging data are longitudinal and can generally be reached at any time (sometimes even in real time). Therefore, unlike passive data, traditional surveys can quickly become dated and less representative of current conditions (Miller et al., 2018). In addition to these differences in content and nature, another argument for why either surveys or passive data will not replace the other is related to concerns about the representativeness of each data source. Indeed, significant limitations are recognized in the scope and quality of data that can be collected from any single data source (Polak & Cornelis, 2009). Beyond the well-known stated response biases, methodological issues such as decreasing response rates and underrepresentation of some population subgroups are representing increasing challenges in traditional surveys (Bayart et al., 2009; Miller et al., 2018). However, even if available in huge quantity, passive data can also be incomplete and imperfect (Bonnell & Munizaga, 2018). Lovelace et al. (2016) reminded that the fourth V (veracity) should always be checked. Other reasons for why passive data will not replace survey data in the foreseeable future include the fact that they are massive and therefore difficult to store and manipulate, and that passive data collection methods are still-evolving (Miller et al., 2018).

Data fusion is one solution put forward in the literature to take advantage of all data sources and to reduce their respective drawbacks, but this raises many challenges (Bayart et al., 2009). The first one is to obtain emerging big data because the different databases (generally one per transportation mode or system, while all modes are considered in survey data) may belong to different private owners (Miller et al., 2018). Consequently, few authors have access to enough passive datasets to analyze simultaneously different transport services (Morency, Trépanier, Saunier, et al., 2017). This paper is one of the exceptions since the authors capitalize on data for three modes (car, subway, and cycling). Other difficulties are linked to the heterogeneity of data formats and usage metrics (or travel indicators) among the various datasets. This is especially true with survey and passive data since the former contains trips whereas transactions or counts or GPS tracks are recorded in the later (Bonnell & Munizaga, 2018). The resolution as well as the data aggregation levels can also be distinct in the datasets. Additional data comparability issues include differences in the catchment areas (the whole region versus the network coverage of a given service) and the corresponding reference population. Non-resident travel by visitors and service/business-related travel are not covered in traditional surveys while all movements are captured in passive data (Miller et al., 2018). All these disparities, which may be explained by the different initial role of

the two data sources, can be classified into semantic, temporal or spatial incompatibilities (Bayart et al., 2009). To compensate them and allow valuing both data types, data aggregation is generally needed, entailing a lack of precision. Some data preprocessing steps may also be required, and the algorithms must be kept up to date despite the rapid evolution of most of the passive data collection technologies. Furthermore, there is a lack of fusion methods as well as validation processes, and there are few examples of data combination in the transportation field. Therefore, more research efforts are needed to benefit from these various data sources, as well as to better define the role of both traditional and emerging data for complementarity purposes (Cherchi & Bhat, 2018).

4.2.3 Data fusion between traditional and emerging transportation data

Applying the framework of D'Orazio et al. (2006) and D'Ambrosio et al. (2007) in a transportation context, the data fusion problem was conceptualized as the combination of a donor database and a receptor database in order to convey additional information (specific variables observed in the donor database but not in the receptor database) basing on variables that are common across both databases (Bayart et al., 2009). Data fusion procedures are often classified in two categories: exact matching, when two records (one from each database) can be matched thanks to a unique identifier or, more commonly, statistical matching which is based on models. They are two types of data fusion models: explicit and implicit. Explicit modeling consists in training classical models (multiple regression, logistic regression, or nonparametric models such as decision trees) on the donor data source and then applying them to the receptor data source. Implicit modeling occurs when each record of the donor dataset is linked to one (constrained matching) or several (unconstrained matching) records of the receptor datasets which are most similar, for instance using nearest neighbor procedures (Bayart et al., 2009; Zhu et al., 2018).

Combining several surveys is a long tradition in transportation, for instance to merge revealed and stated preference data (Hensher, D. et al., 1998), or census data and travel surveys (Venigalla, 2004). Most recent examples include the combination of several survey modes (Verreault & Morency, 2018), or the integration of a special-purpose satellite survey (e.g., targeting students) with a core household travel survey (Verreault & Morency, 2018; Wang, Kaili et al., 2021). Data fusion methods were also developed for applications in road traffic engineering and intelligent transportation systems using sensor data (El Faouzi & Klein, 2016). However, examples of data integration between traditional surveys and passive data are much sparser. This is in part due to the

lack of common sociodemographic attributes between the two data types; the common variables (needed in most data fusion methods) must be constructed from temporal/spatial information, mobility indicators, or a combination of them (Bonnell & Munizaga, 2018). Nevertheless, some works can be found in the transit context where the two types of data sources, namely smart card data and travel origin-destination surveys, have been compared in order to validate the collected information through spatial (OD matrixes) and temporal distributions (Egu & Bonnell, 2020; Munizaga et al., 2014; Spurr et al., 2014; Spurr et al., 2015). Furthermore, enrichment procedures (probability-based or not) were proposed to add additional information from surveys to passive data, such as origin/destination places (Long & Thill, 2015) and activity locations/trip purposes (Kusakabe & Asakura, 2014; Long & Thill, 2015) in smart card data, or trip departure times in mobile phone data (Alexander et al., 2015). Explicit models were also developed to match smart cards to individuals who can then be described by sociodemographic attributes (Grapperon et al., 2016; Spurr et al., 2015).

The previous works exploited traditional surveys (donor) to complete information about trip or user attributes in big data (receptor) or both data sources (both as donor) were used for validation/comparison purposes, but none of them tried to apply the variations captured in longitudinal data (donor) to cross-sectional surveys (receptor). Considering the cross-sectional survey as receptor is a relatively new idea and echoes with the core-satellite data collection approach proposed in some papers, where the household survey is the core and passive datasets are ancillary (Miller et al., 2018; Wang, Kaili et al., 2021). This work is a first step in that direction by checking the potential of finding similar fluctuations in both data sources despite their different initial roles and natures. This similarity could then be further investigated in data fusion procedures for longitudinal monitoring purposes. Such a validation is indeed necessary before trying to apply variations detected in passive data to survey data, for instance to annualize (including annual and weekly seasonality) the average patterns observed in a survey, or to fill the gap between two cross-sectional surveys with longitudinal data (thanks to the long-term trend observed in longitudinal data). More specifically, this paper aims to develop an expansion method for “typical” one-day trip diaries, and then to cross-analyze the variations in travel demand using passive data. The term cross-analysis is preferred to validation because the latter would assume passive data provide the truth about mobility behaviors while there are still concerns about their quality and completeness. Moreover, the above-mentioned differences between traditional surveys and passive datasets (in

format and metric, catchment area and reference population, etc.) mean that the observed fluctuations are not obviously similar. Comparisons between the two data sources will be carried out at the longitudinal level (over several months and throughout a week) instead of a single day as it was already done for Montreal (Spurr et al., 2014; Spurr et al., 2015), and for several transportation modes instead of only one. Going beyond simple descriptive comparisons, indicators will be proposed to quantify the differences between the two data sources.

4.3 Data and methodology

4.3.1 Data description

Two types of data are harnessed for this work: the Origin-Destination (OD) regional household travel survey of Montreal, Canada, and several passive data streams.

On the one hand, the OD survey of Montreal is a telephone (and web since 2018) survey conducted every-five years in the metropolitan area of Montreal. Resident households are randomly sampled every day over the fall to be interviewed (stratified sampling without replacement, thus each household is surveyed no more than once). Each respondent must report the trips of all household members made the previous day. Sociodemographic information is also requested for the household and for each person living in the household. The collection period spans over 16 weeks (78 weekdays and 30 weekend days when the latter are collected) from September to December, public holidays (Labor Day in September and Thanksgiving in October) not being surveyed. Nevertheless, the data of all single-day samples are merged to represent a so-called “typical day”. In this way, the global sampling rate reaches about 4-5% of all Montreal resident households. The records are weighted using the Canadian census according to sociodemographic and geographic settings to allow statistical inference to the entire reference population, representative of the metropolitan area of Montreal (Secrétariat à l'enquête Origine-Destination, 2010, 2015, 2020). Weighting factors are first calculated for the households using a margin calibration method based on the geographic stratum of their home, as well as their size and composition (by age group). Category totals are taken from the Canadian census. Then, weights are determined for the persons by adjusting the household factors to consider age cohorts and gender. Trips are estimated using the person factors. In 2018, the factors were also corrected for the month of September (which was under sampled), especially for bicycle use. In this paper, the last three surveys are handled: 2008,

2013 and 2018 (77,095, 78,731 and 75,266 households respectively). Trips were collected on both weekdays and weekends in 2008 (thanks to two independent surveys based on 66,124 and 10,971 households respectively) but only weekdays are available in the 2013 and 2018 surveys.

On the other hand, several passive data streams are mined to describe the use of various modes of transport: smart card data for the subway, and count data for cycling and private cars. The subway network of Montreal is composed of 4 lines (68 stations) and is operated by the STM (*Société de Transport de Montréal*) transit authority. Transactional information is collected by a smart card automatic fare collection (SCAFC) system called the OPUS system. This system gathers tap-in validations (at boarding only) in the subway stations and onboard buses, but only the subway data is used in this paper because each validation corresponds to one trip (while first boardings and transfers should have been differentiated for the bus validations). There are several supports in the STM fare system: traditional smart cards, used by more than 90% of the users, but also occasional smart cards and disposable magnetic cards (paper tickets). However, all validations (regardless of the support) are collected by the OPUS system and are therefore considered in this paper. Cycling data comes from the Eco-counter system and was provided by the municipal organization of the city of Montreal (*Ville de Montréal*). The counts include personally owned bicycles but also (station-based) bikesharing, operated by BIXI in Montreal. Car data is derived from highway and bridge traffic counters which are managed by the Ministry of Transport of Quebec (*Ministère des Transports du Québec*). Each counter is associated with a section of traffic that can be several kilometers long (but the exact position of the counter along this section is not known). Most of these counters are classified; hence it was possible to select only private car counts (without trucks). The subway network of Montreal, as well as the cycling counters and the highway and bridge traffic sections associated with counters and located in the OD survey area are displayed in Figure 4.1 below. Note that all passive initial databases that the authors received have already been cleaned by the data providers and aggregated by date x time interval (one hour or 15 minutes) x station or counter-direction.

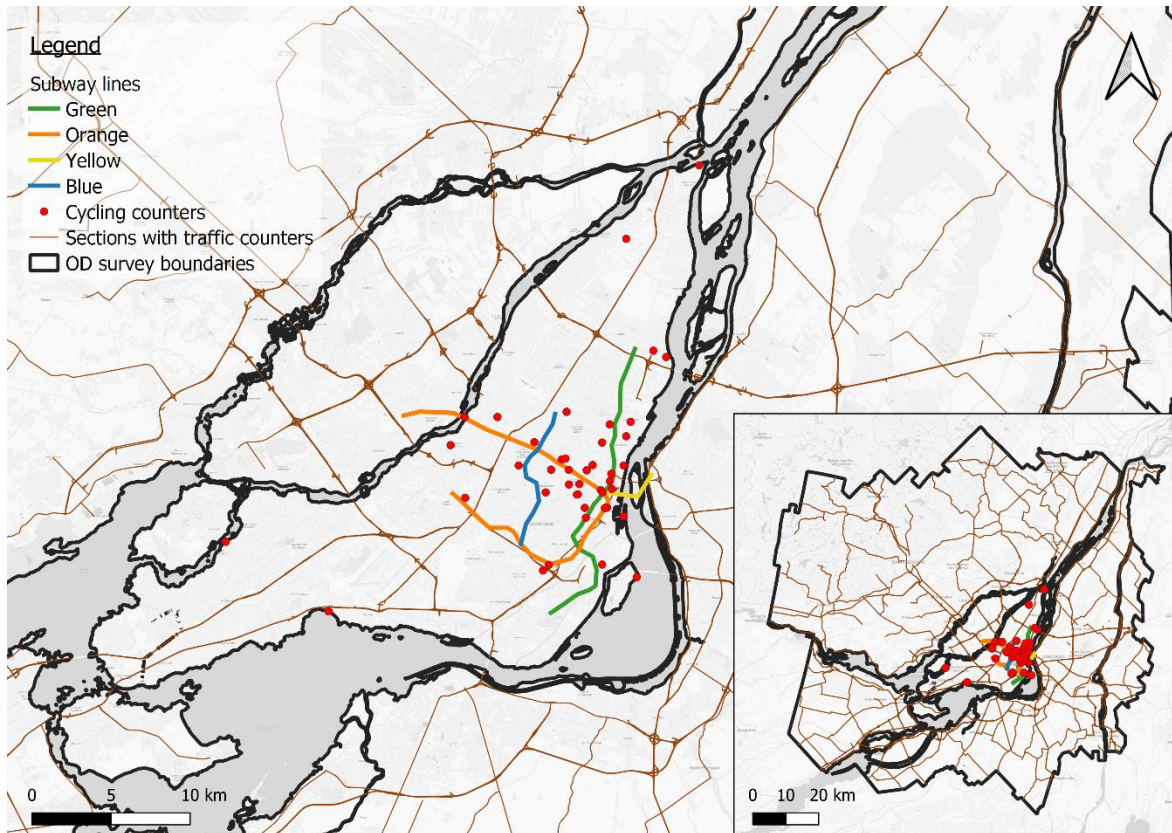


Figure 4.1 Study area (metropolitan area of Montreal) with the subway network, cycling counters, and highway and bridge traffic sections

All these datasets have many differences in the spatial dimension and reference population, as well as in the collected information and their quality/completeness. Indeed, the covered area is not the same: the whole Montreal metropolitan region is surveyed in the OD survey, while only the catchment zone of the considered transportation mode or the data collection system coverage is available in each passive dataset. The traffic counters are spread out all over the OD survey area, but the subway network and the cycling counters are located on the Montreal Island only. Moreover, the reference population accounts for all residents of the Montreal region in the OD survey, whereas non-residents (visitors) are also included in the passive datasets (but only the users of each transportation system are considered). As participants are recruited based on landline and cell phone listings in the OD survey, some subgroups of the population may also be underrepresented while they are automatically observed in the passive datasets. A weighted sample of the reference population is reported in the OD surveys and an unweighted sample of car and bicycle use is gathered in count data, whereas an almost complete inventory is available in subway

data since all validations (regardless of the support) are comprised (but some validations may have been lost due to fraud and errors in the OPUS SCAFC system). The travel indicators or metrics differ too: trips versus counts or validations are recorded in the OD survey versus passive data respectively. This semantic incompatibility partly explains why the comparisons between survey and passive data made in this paper are done at the aggregate level, and not at the individual level: we do not have users in passive data but transit cards and tickets or counts, and there are no direct links or common attributes to match each card, ticket or count with a survey respondent. Regarding data accuracy, the well-know biases due to respondent stated information are probably present in the OD survey, whereas observed data in the passive databases should be closer to the “ground truth”. Based on the authors’ knowledge of the data, there may be errors (due to the collection and/or preparation processes) in the count data used for car and bicycle, while smart card data should indeed be more reliable.

Therefore, all these disparities made the two data sources studied in this work not necessarily similar a priori. In the following section, no additional preprocessing procedures will be applied to convert the passive data metrics to trips, or to match the spatial coverage of each database, but the variations in aggregate travel behavior will still be compared. This is a cross-analysis or a double validation: passive data are used to validate the estimated variability of travel behavior (or to validate the possibility to represent weekly travel behavior) from one-day travel diaries, but survey data are also leveraged to validate the variability observed in the imperfect and/or incomplete and apparently incompatible passive datasets. Moreover, the indicators and methods developed to compare the two data sources should be able to overcome these differences, basing on variations only and not on the absolute values which are fundamentally different as they refer to different contexts and objects.

4.3.2 Methodological steps

The proposed method is composed of three main steps depicted in Figure 4.2 and explained in the rest of this section. Data processing is the most challenging.

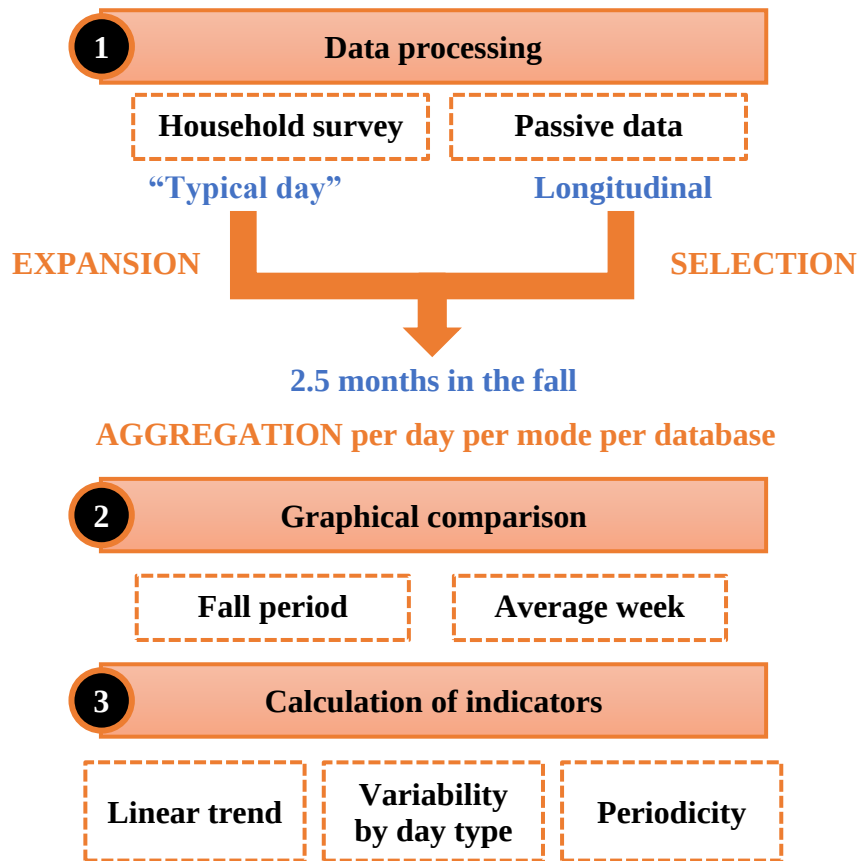


Figure 4.2 Method overview

4.3.2.1 Data processing

First, the two types of data described above are processed to make them comparable and to explore the variability of mobility behaviors over a period of 2.5 months. An expansion method is applied to the household survey to go beyond the notion of “typical day”. This process is illustrated in Figure 4.3. The input (top of the figure) is an OD survey of Montreal, in which the sum of the initial person weighting factors (F_{PER_i} for individual i) is equal to the reference population (Pop_{ref}). The respondents are first segmented according to the day on which their trips were recorded, for both week and weekend when the information is available. For each day, an expansion factor (F_{EXP_j} for day j) is calculated and applied to the initial weights so that the daily number of people is now equal to the reference population. This method was developed by Verreault et Morency (2011) and automated by the authors of this work. Mathematically, this amounts to

calculating a corrected person weighting factor by multiplying the initial person factor (which corrects for representativeness) by an expansion factor (which corrects for daily variability).

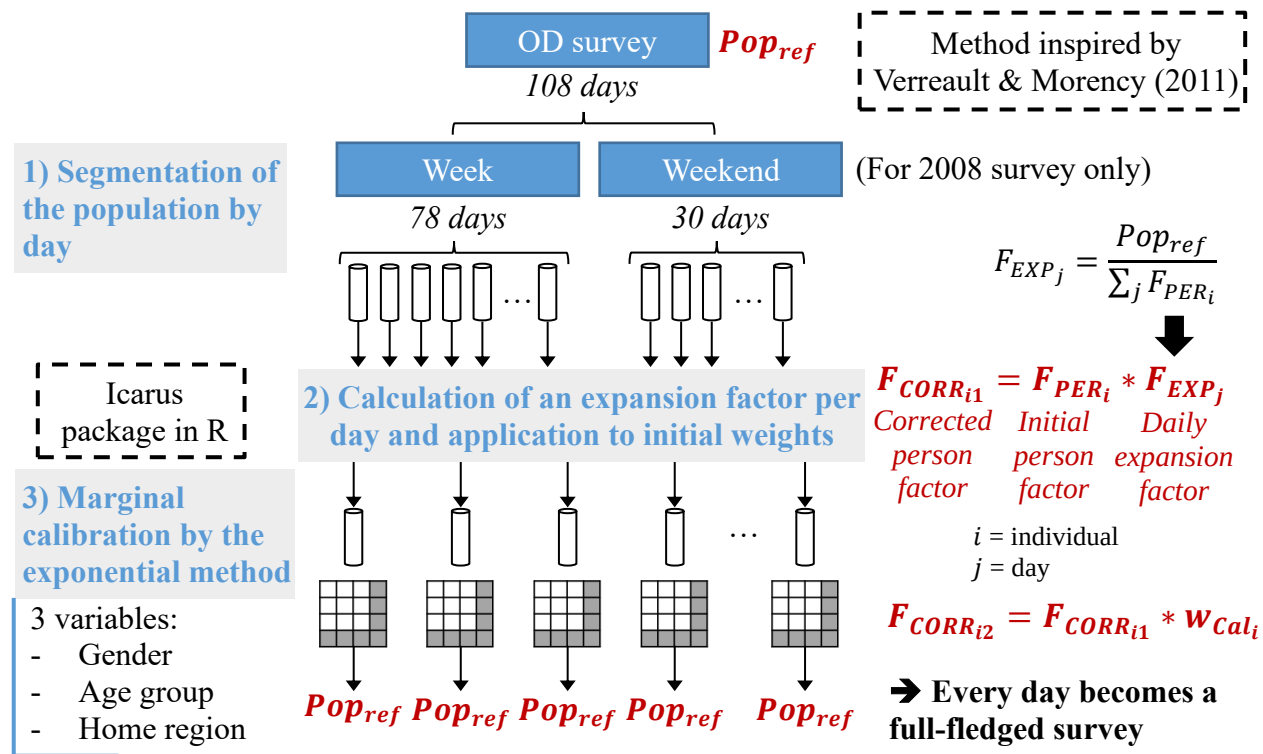


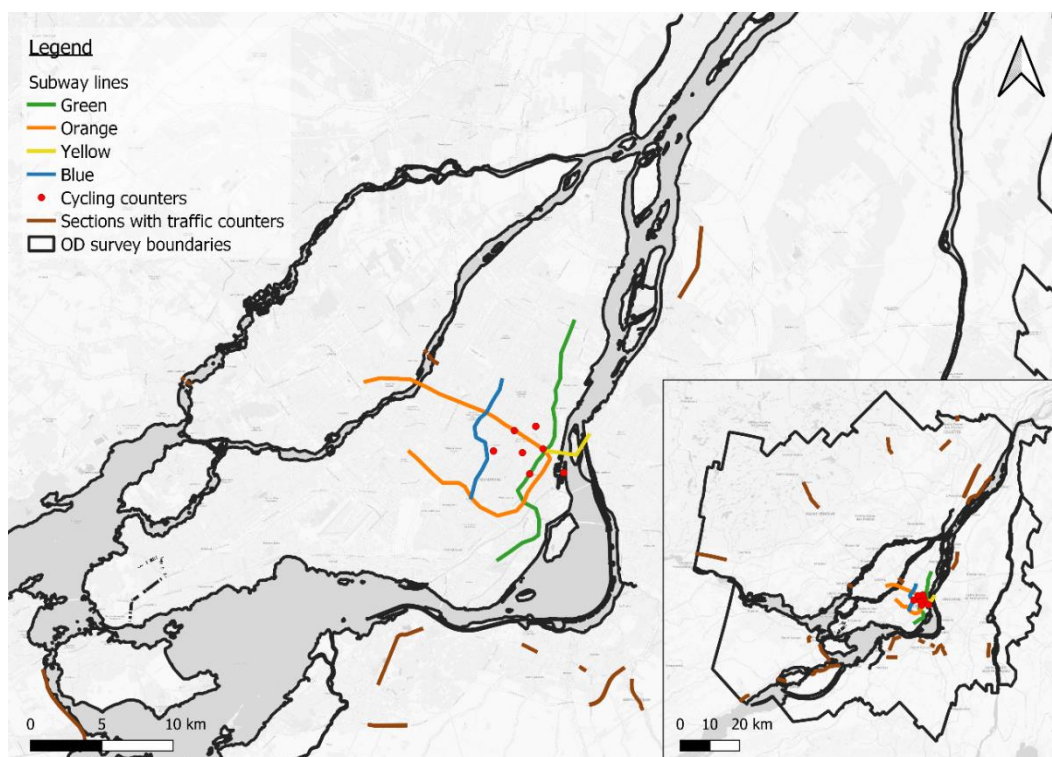
Figure 4.3 Expansion method applied to the Montreal Origin-Destination (OD) survey

In this paper, a third original part is added because the composition of the single-day samples is not uniform (for instance, some regions were more surveyed than others on certain days), and this may cause artificial differences in travel behaviors over time. Therefore, a margin calibration method is applied for each day to have an aggregated population with the same socio-demographic characteristics as the initial population. Note that a marginal calibration method (instead of a simple segmentation method) is chosen in this third step in order to overcome the “zeros” (or low values) that would be obtained in some categories by crossing several socio-demographic attributes. The package *Icarus* in R (Rebecq, 2016) is used with the exponential method (or raking ratio estimation). Three calibration variables are considered: gender, age group, and geographic stratum of home. These three attributes are selected because they also are the variables used in the initial weighting method of the respondents in the OD survey of Montreal. Regarding the third variable (geographic stratum), municipal tracts are used in the weighting of the week OD surveys whereas regions are used in the weighting of the weekend OD surveys; thus, we keep the most aggregate

area, i.e., region. In this way, the proposed methodology is consistent with the underlying assumptions applied in Montreal transportation planning. Moreover, these variables are widely considered in the literature to explain travel behavior. Therefore, this third step is equivalent to multiplying the previous corrected factor by a calibration factor (w_{cal_i}) to correct for the heterogeneity of the single-day samples. Finally, the initial reference population is found in each day. In this way, every day becomes a full-fledged survey, which allows calculating a representative number of trips on all days over the survey collection period (four months in the fall), and thus longitudinal monitoring of aggregated daily travel behaviors.

Regarding the second data source used in this paper (passive data), they are already longitudinal. However, a common study period must be chosen to compare the two data types. Therefore, the passive data corresponding to the OD survey collection periods, namely the 2013 and 2018 fall periods (the passive data used in this paper was not available for 2008), are selected. However, some sampling issues caused high calibration factors in the 2018 OD survey (see the results section for more details), which led us to not consider the whole month of September and the third week of December in the final comparison. The common study period then spans over 2.5 months, from early October to mid-December. This means all smart card validations gathered on this period (in all stations) are considered, but only the counters that were active almost continuously on the whole period (per year, i.e., 2013 or 2018) are selected. More specifically, we select the counters for which less than 24 hours of observations are missing in each direction (i.e., the total of NA values in our hourly raw data must be inferior to 24 in each direction per selected counter). In this way, 7 counters for bicycles and 23 counters for private vehicles are considered in fall 2013 (21 and 37 counters respectively in fall 2018). These counters are displayed in Figure 4.4. The selected traffic counters cover different parts of the study area, while the selected cycling counters are concentrated in the center of Montreal Island. Nevertheless, all most important bicycle lanes are in this zone.

a) 2013



b) 2018

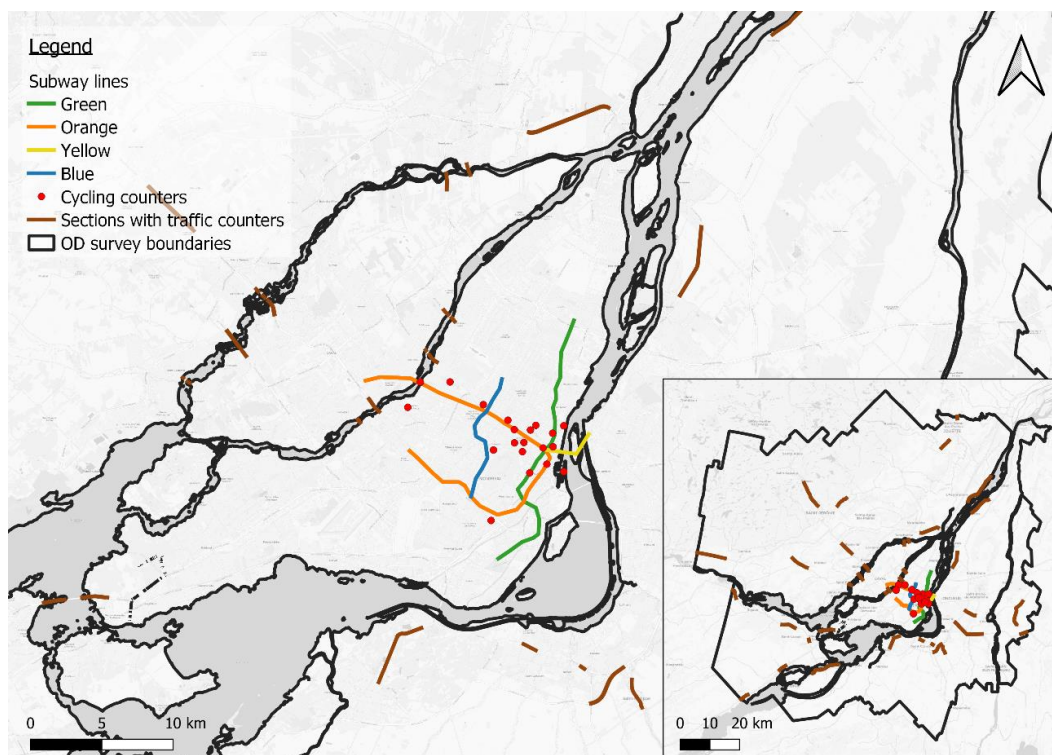


Figure 4.4 Selected cycling and traffic counters in a) 2013 and b) 2018 passive datasets

Following this first step of expansion and selection, times series per mode are built to track variations in the travel demand over 2.5 months in the fall from both household survey and passive data. The trips (survey data) or validations (smart card data) or counts (count data) are aggregated by day and hour from 4:00 a.m. to 28:00 p.m. (4:00 a.m. the next day): this is the definition of a day in the OD survey, where trips collected from 24:00 to 28:59 are trips that began the day after the one being surveyed and that typically allow for the conclusion of a travel chain. The smart card validations are summed (because one validation is equivalent to one trip) whereas the mean of the selected counters (without distinguishing the direction, i.e., the counts of both directions were summed per counter before) is calculated for the car and the bicycle. In fact, a sum could also be done for the counters because the variations only are of interest in this paper, neither the absolute value nor the intensity measurement metric. As a result, hourly time series for each mode (car, subway, and bicycle) are derived from each database (2008 OD survey, 2013 OD survey, 2018 OD survey, 2013 passive data and 2018 passive data) in total number of trips or total number of validations or average number of counts. The data are also aggregated by day, hence 77 days-time series per mode per database are obtained.

4.3.2.2 Graphical comparison

The previous data processing steps make it possible to compare the variability of the use of three transportation modes (car, subway and bicycle) measured from the two data types at the longitudinal and aggregated level, over 2.5 months in the fall or across an average week. Such comparisons are first made visually thanks to line and bar charts. The curves representing the daily number of trips or validations or counts over the fall are drawn for each mode and each database. To make the values of all time series comparable, the trips/validations/counts are distributed over the whole period (in proportion instead of absolute value). The percentages of use of each mode per day are estimated using as denominator the total number of trips or validations or counts on weekdays only because some databases do not contain information for the weekends. Average weeks are also calculated and compared across all modes and all databases. In this way, the weekly variability of the aggregated travel behaviors is cross-analyzed using the two data types.

4.3.2.3 Calculation of indicators

To go beyond qualitative comparisons, indicators are also developed to measure and compare (across all databases) three patterns in the time series: the trend, the variability by day type and the

periodicity. The following indicators aim to check if the same variations are observed in the two data types, regardless of the volume metric.

- **Linear trend:** First, we try to detect a linear trend (increase or decrease) in the use of each mode over the fall and then to compare this trend across all databases. To this end, simple linear regressions are estimated from the time series of trips or validations or counts (converted in proportions) on working weekdays only. Moreover, a variance analysis (one-way ANOVA) is used to test the interaction of the trend and the database (to check if the choice of the database influences the slope obtained). When this interaction is significant, Tukey post-hoc tests allow doing pairwise comparisons of the slopes calculated with each database. These comparisons are made by applying the *lstrends* function of the *lsmeans* package in R: see (Lenth, 2016) for more details.
- **Variability by day type:** Secondly, we seek to evaluate variability in the use of each mode by day type (from Monday to Sunday). To this end, daily ratios between the time series inferred from passive data (numerator) and OD survey (denominator) are estimated using the following equation:

$$Ratio_j = \frac{Passive\ data_j}{OD\ survey_j} \text{ for the day } j$$

Such ratios remove the trend in the data (assuming the trend is similar in both survey and passive data) and allow assessing the similarity of the variations detected in the two data sources. Three types of ratios are calculated: between the 2018 passive data and the 2018 OD survey (for weekdays only), between the 2013 passive data and the 2013 OD survey (for weekdays only), and between the 2013 passive data and 2008 OD survey (for weekdays and weekends). The stability of these ratios (equivalent to three daily time series) over the whole study period is further examined. Moreover, the average ratios and the corresponding coefficients of variations are calculated by day type.

- **Periodicity:** Thirdly, we look at the hourly time series to highlight periodicity in the data. Like for other signals in physics, the Fourier transform can be used to detect frequencies that explain the variations in the time series. Moreover, the power spectral density (based on the Fourier transform) gives the relative strengths of these frequencies. In this paper, the power spectral density of each hourly time series per mode is estimated using the

spec.pgram function in R (which first allows detrending the curves) and the corresponding periodograms are drawn to reveal the most important frequencies (or periods, inverse of the frequencies) in each database, to be compared between survey and passive data.

4.4 Results

Results are reported in three parts. First, the new person weighting factors of the 2008, 2013 and 2018 Montreal OD surveys are presented, as well as the curves of all aggregated travel behaviors extended over the entire collection period. The use of three modes (car, subway, and bicycle) described by time series built from both reweighted OD surveys and passive data is then visually compared over 2.5 months in the fall and across an average week. Finally, indicators are calculated to quantify these comparisons.

4.4.1 Expansion of the Montreal origin-destination survey

The single-day samples of the last three Montreal OD surveys are made representative over the entire collection period using the expansion method explained above. Some statistics associated with the margin calibration (third step of the method, intended to correct for the heterogeneity of the samples) are first reported. Table 4.1 gives the mean and the distribution of the calibration factors (w_{cal_i}), defined as the ratio between the calibrated weight (corrected weighting factor obtained in step 3) and the previous weight (corrected weighting factor obtained in step 2). Most of these factors have a value close to 1: this is desirable since it means that there is not too much imbalance in the composition of the samples regarding the three socio-demographic variables considered. However, there are some outliers, especially in the 2018 OD survey. Figure 4.5 shows that the highest and lowest calibration factors are in fact concentrated on a few specific days, notably during September and the last week of December. This is because some regions were much more sampled than others on these days. A bounded calibration method may have reduced the outliers, but we chose to apply the traditional unbounded method to make convergence simpler and faster (in particular because the method had to be applied to each observation day).

Table 4.1 Descriptive statistics on the calibration factors for each Montreal OD survey

Year		2008	2008	2013	2018
Type of day		Week	Weekend	Week	Week
Mean		1.0007	1.0026	1.0009	1.0240
Quantiles	0%	0.182	0.341	0.428	0.072
	1%	0.363	0.495	0.709	0.495
	10%	0.758	0.679	0.829	0.754
	25%	0.863	0.803	0.898	0.871
	50%	0.973	0.969	0.988	0.995
	75%	1.102	1.128	1.082	1.143
	90%	1.248	1.349	1.183	1.309
	99%	1.818	1.955	1.464	1.890
	100%	15.825	6.371	5.577	9.270

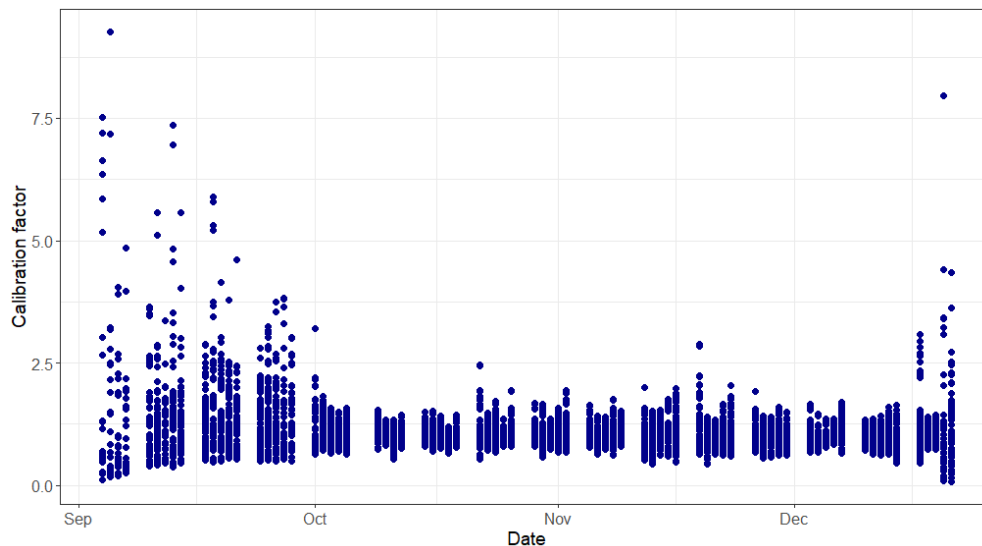


Figure 4.5 Distribution of the calibration factors over the fall period (2018 Montreal OD survey)

The application of the expansion method allows us to calculate a total daily number of trips during the four months of the fall (by summing all final corrected person factors by day), for the whole Montreal region and all modes of transport. This was only possible on a “typical day” with the initial database. The resulting weighted curves are plotted in Figure 4.6 for each survey. Only the 2008 curve is complete as both weekdays and weekends were surveyed. Instead, gaps are visible in the 2013 and 2018 curves because data are not available on weekends. However, the number of trips on weekends is artificially brought down at the beginning and at the end of every week to better differentiate them. Weekly cycles are observed, and they are repeated more or less regularly throughout the study period. Some peaks are noticeable on Fridays in all surveys, especially in

December. They could correspond to Christmas parties, which are quite popular in Quebec. However, the third week of December, as well as the first week of September, look abnormal. This is where the outliers of the calibration factors were found. Moreover, the blue histogram displayed in Figure 4.6 shows that the number of households sampled in the 2018 OD survey on these days was much smaller. To avoid biases in the results, the four weeks of September and the third week of December are removed in the subsequent analyses.

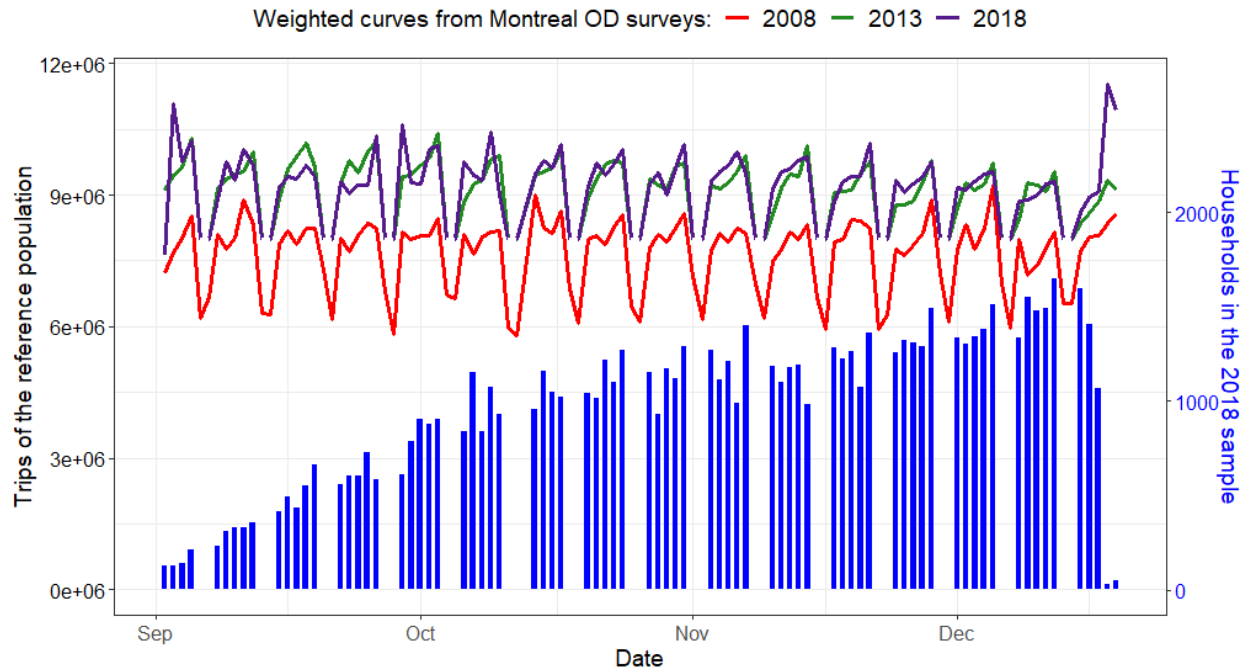


Figure 4.6 Total number of trips per day (all modes of transport) during the entire survey period

4.4.2 Description of the use of three modes from the two data sources

The previous curves (or time series) can also be estimated by mode and then compared with the same curves obtained (after temporal selection and daily aggregation) with passive data. Three modes for which both survey and operational data were available are considered: car, subway, and bicycle. Figure 4.7 provides the curves produced for these modes from the 2013 OD survey (weekdays only) and passive data, whereas the 2018 databases are used in Figure 4.8. The use of each mode is measured as a proportion of the total number of trips or validations or counts for the selected fall period on weekdays only. Average weeks are also calculated for all modes and all databases. The differences (in percentages) between the average use by day (excluding holidays) and the average working day are reported in Figure 4.9 for both weekdays and weekends

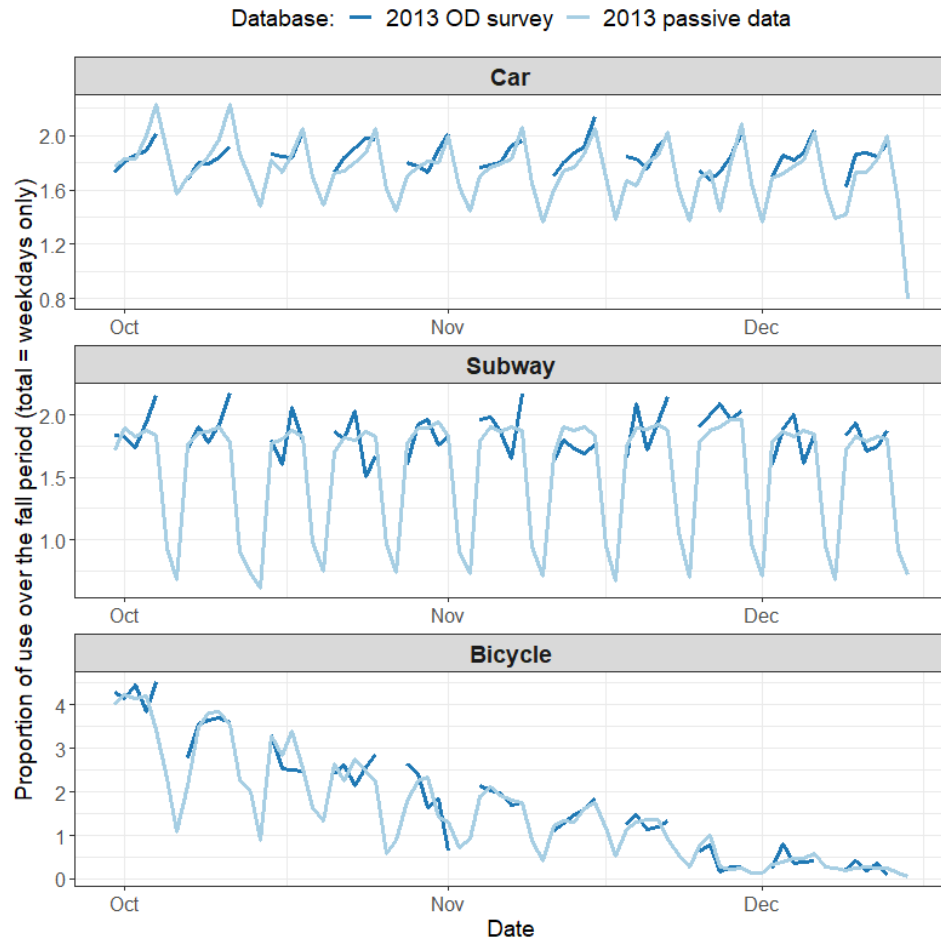


Figure 4.7 Proportion of use of three modes (car, subway, and bicycle) over the fall period from 2013 OD survey and passive data

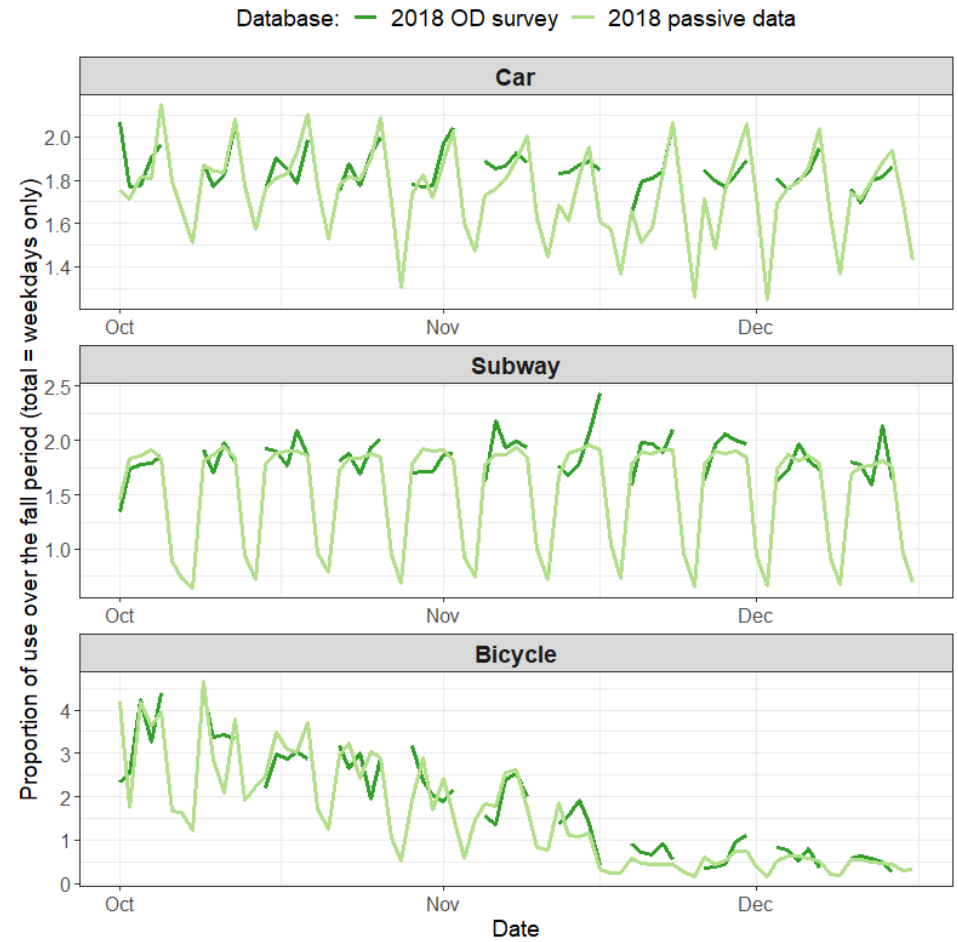


Figure 4.8 Proportion of use of three modes (car, subway, and bicycle) over the fall period from 2018 OD survey and passive data

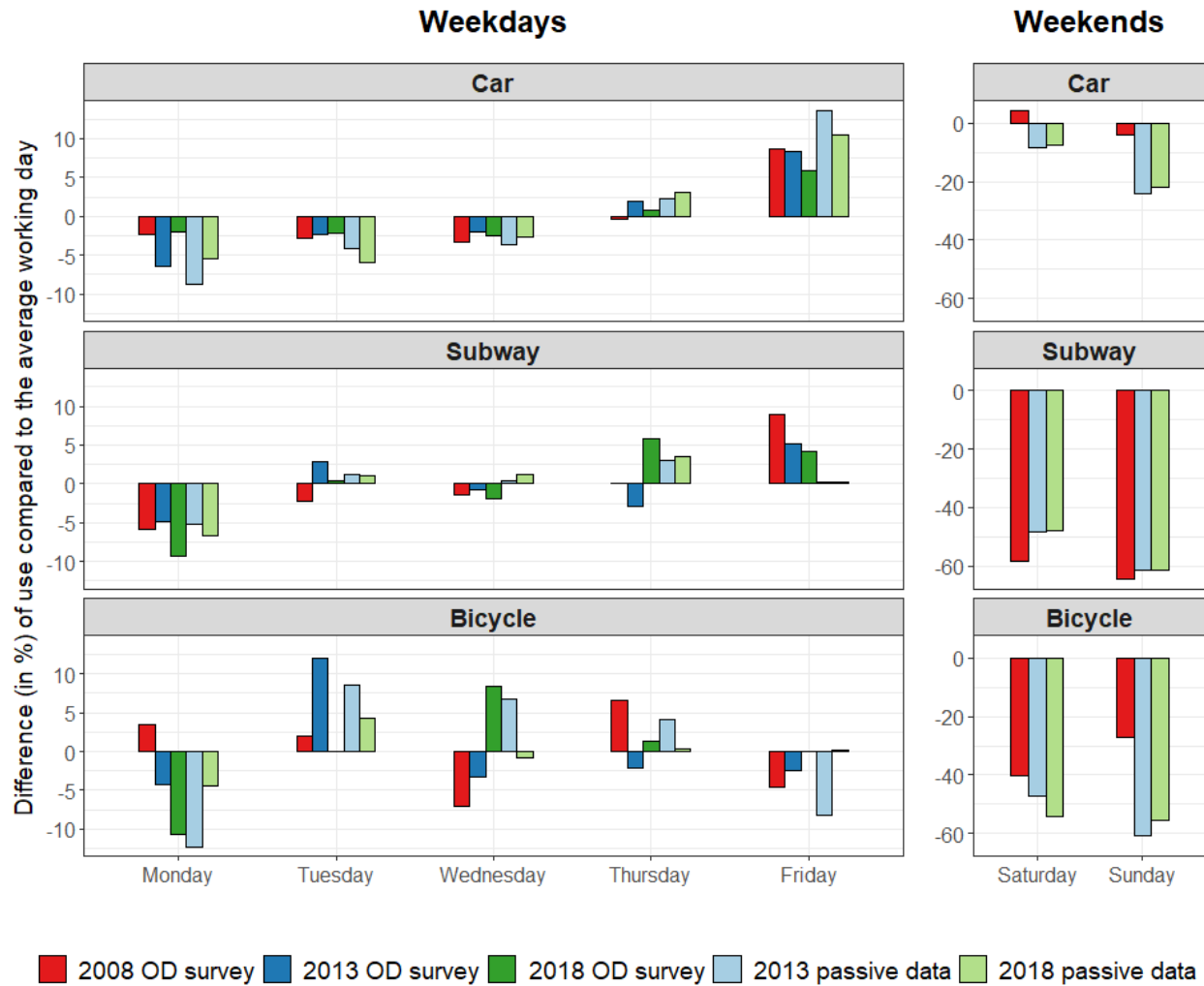


Figure 4.9 Difference in the average use of three modes (car, subway, and bicycle) by day type compared to the average working day from all databases (2018, 2013, 2008 OD surveys and 2018, 2013 passive data)

For the car, the shape of the weeks is similar in the OD surveys and passive data: both in 2013 and 2018, the usage of this mode tends to increase over the week, especially on Fridays. This is confirmed by the average week in Figure 4.9: the average use of the car on Thursdays and Fridays is higher than on the average working day, while it is lower on Mondays, Tuesdays, and Wednesdays. The variations observed on weekdays (in comparison with the average working day) are very much alike in all databases. However, the drop on weekends is greater in the passive data than in the 2008 OD survey, especially on Sundays. The decreases obtained for the car on weekends are nevertheless the smallest differences from the average working day among all three modes.

Regarding the subway, the weekly cycles are more regular in the curves drawn from the passive data than from the OD surveys, but the average weeks are quite analogous on both weekdays and weekends. A lower use of the mode on Mondays is detected in all databases. However, a slight increase is detected on Thursdays in the passive data, while this increase is rather found on Fridays in the OD surveys (except for the 2018 OD survey, where the increase is more important on Thursdays too). All data sources capture the decline in bicycle use over the study period as winter approaches and temperatures drop. Therefore, the global trend is akin in the two types of data. However, the average weeks calculated from the different databases are more divergent.

4.4.3 Comparison indicators of the two data sources per mode

The previous analyses showed visual similarities between all data sources in the relative total use and in the average weekly use of three modes over the fall, but this likeness needs to be quantified. The two data types are then compared with indicators according to three concepts: linear trend, variability by day type and periodicity.

4.4.3.1 Linear trend

First, simple linear regression models are trained on the previous curves (in percentages of use of each mode) for the working weekdays of the 2.5-month selected fall period. Based on the results, there is a trend in the bicycle data only (the coefficient associated with the time variable is not significant for the other modes). The linear trend detected in each database, as well as the related slope and model accuracy (coefficient of determination R^2), are given in Figure 4.10. A similar negative coefficient (decreasing use) is observed in both surveys and passive data. The question raised now is: are the slopes significantly different between these data sources? The interaction term between the database (as categorical variable) and the trend in the use of the mode (bicycle) is significant (at 5%) according to the one-way ANOVA test. However, the pairwise comparisons of the slopes (Tukey post-hoc tests) in Table 4.2 reveal that slopes are only significantly different between the 2008 OD survey and the 2013 databases (the null hypothesis that they are equal is rejected at the 10% significance level since p-value less than 0.10). Rather, the deviations are not significant at all between OD surveys and passive data in both 2013 and 2018 (the equality of the slopes cannot be rejected). This confirms that similar trends are found in the two types of data.

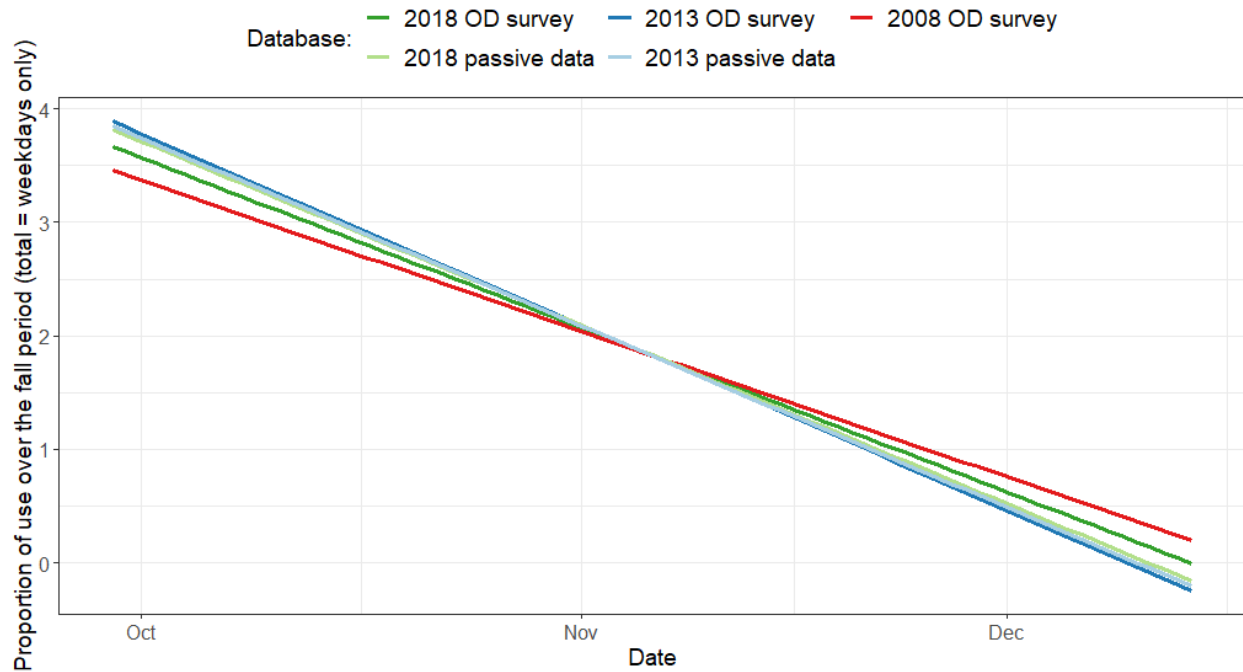


Figure 4.10 Linear trends in the bicycle databases

Table 4.2 Pairwise comparisons of the linear trends in the bicycle databases (p-values of the Tukey tests)

	2008 survey	2013 survey	2018 survey	2013 passive	2018 passive
2008 survey	1	0.0511	0.7010	0.0953	0.1662
2013 survey		1	0.5994	0.9992	0.9889
2018 survey			1	0.7516	0.8735
2013 passive				1	0.9993
2018 passive					1

4.4.3.2 Variability by day type

Ratios are also calculated between the survey and passive data time series to determine if these two types of data capture the same fluctuations in ridership over the fall, especially the weekly rhythms. These ratios are illustrated in Figure 4.11 for each mode over the entire study period. Moreover, the average ratios by day type and the associated coefficients of variations are provided in Table 4.3. The absolute values of the ratios are not of interest in this paper (since the metric and the volumes observed in the two data types are different); only the variations are.

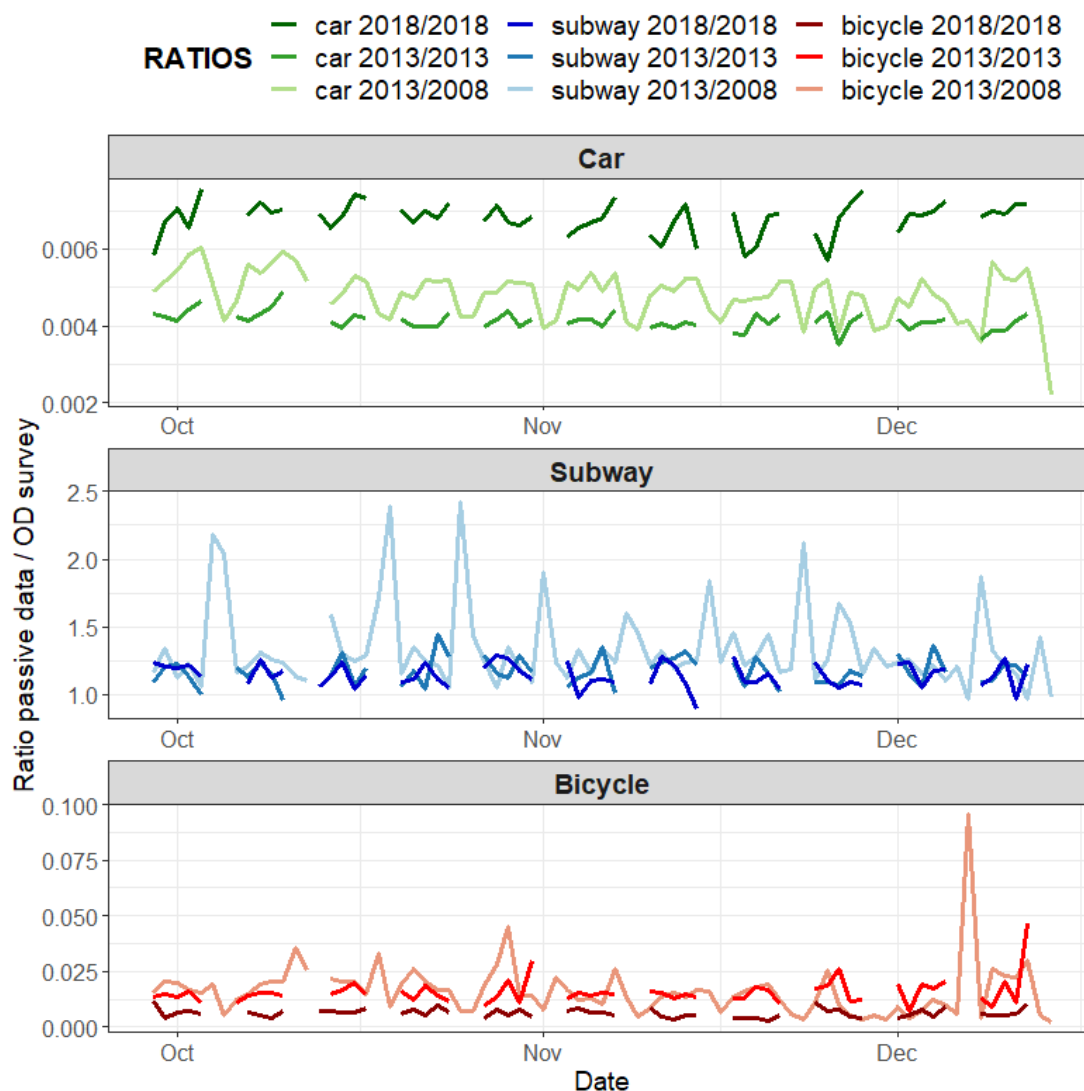


Figure 4.11 Ratios per mode between passive databases and Montreal Origin-Destination (OD) surveys

According to the results, the most regular ratios are obtained for the car: the estimated coefficients of variations are indeed low, below 9% for most of them. This means that even if the volumes are different from one database to another (ratios are different from 1), the two data sources approximately capture the same weekly seasonality: the same days of the week are similar in the two data types (according to the previous graphical comparisons) and the aggregated daily use of the mode for all day types (all Mondays, all Tuesdays, etc.) varies in the same way in the two data sources across the selected study period.

Table 4.3 Average ratios (passive data / OD surveys) and coefficient of variations, per mode and day type

MODES	RATIOS	WEEK						WEEK-END	
		Monday	Tuesday	Wednesday	Thursday	Friday	All days	Saturday	Sunday
CAR	2018 / 2018	6.58E-03	6.54E-03	6.80E-03	6.96E-03	7.10E-03	6.80E-03	-	-
		5.6%	7.2%	4.4%	3.8%	6.0%	6.2%	-	-
	2013 / 2013	4.04E-03	4.07E-03	4.07E-03	4.15E-03	4.35E-03	4.14E-03	-	-
		4.9%	4.2%	6.2%	4.3%	5.5%	5.6%	-	-
	2013 / 2008	4.72E-03	4.99E-03	5.02E-03	5.17E-03	5.28E-03	5.04E-03	4.46E-03	4.00E-03
		8.9%	7.9%	9.1%	6.6%	8.1%	8.7%	13.1%	17.4%
SUBWAY	2018 / 2018	1.17E+00	1.15E+00	1.18E+00	1.11E+00	1.10E+00	1.14E+00	-	-
		7.4%	8.1%	7.5%	6.1%	8.2%	7.7%	-	-
	2013 / 2013	1.16E+00	1.14E+00	1.18E+00	1.24E+00	1.11E+00	1.17E+00	-	-
		7.9%	4.5%	8.0%	9.6%	9.4%	8.6%	-	-
	2013 / 2008	1.28E+00	1.30E+00	1.27E+00	1.28E+00	1.15E+00	1.25E+00	1.63E+00	1.47E+00
		18.2%	10.1%	11.9%	8.9%	8.8%	12.4%	26.1%	33.3%
BICYCLE	2018 / 2018	6.87E-03	6.25E-03	5.80E-03	5.93E-03	6.48E-03	6.25E-03	-	-
		39.6%	25.0%	23.0%	32.9%	30.3%	30.5%	-	-
	2013 / 2013	1.42E-02	1.35E-02	1.78E-02	1.48E-02	1.81E-02	1.57E-02	-	-
		19.8%	23.0%	21.7%	18.7%	59.9%	36.7%	-	-
	2013 / 2008	1.28E-02	1.89E-02	1.93E-02	1.55E-02	1.63E-02	1.66E-02	1.40E-02	1.68E-02
		38.4%	39.8%	51.2%	32.8%	44.9%	44.2%	79.1%	161.4%

Conversely, the ratios obtained for bicycle are the most variable, but this may be explained by the irregularity of the weeks observed before in all databases (no regular weekly patterns emerge in either the survey or passive data). This variability around the mean (by day type) across the study period could also be due to the weather effect on the use of this mode. Weather should have the same impacts on both survey and passive data (for the same period), but commuting trips are typically more preponderant in survey data than in passive data and it is well known that weather has less effect on such trips than on discretionary activities. This contrast in weather impact could explain (in part at least) the differences observed between the two data sources. The authors tried to validate this assumption using open historical weather data for the 2013 and 2018 falls from the website of the Government of Canada: the effect of two weather variables (average daily temperature and daily total rainfalls) on the trips or counts (converted in proportions) on working weekdays only was estimated by training two multiple regression models (one with survey data, the other with passive data). The results confirm that weather has more impact on cycling counts

as measured by passive data than on typical cycling trips as measured by travel surveys (temperature is associated with a positive coefficient of 0.127 in survey data and 0.133 in passive data, while rainfalls are associated with a negative coefficient of -0.079 in survey data and -0.085 in passive data, all coefficients being significant at the 0.05 level and R^2 being equal to 76% and 70% for the models trained with survey and passive data respectively). However, the ANOVA test allowed us to conclude that the differences in weather impact between the two data sources are not so significant. Last but not least, the divergence between survey and passive data in the case of bicycle use could also be due to a lack of representativeness of the selected counters in passive data, which are mainly located in the center of the Montreal Island and therefore may not be representative enough of the entire Montreal metropolitan region. A lack of representativeness of the survey information is also possible as bicycle trips are known (by all Montreal transportation planning partners) to be under collected in the OD survey data.

In almost all cases (i.e., all modes and years), Thursday is the most similar day between the passive data and the OD surveys (lowest variations around the average ratio). This result may have implications for the development of fusion methods: the data selected on this day could be further explored to find commonalities to establish links between the two data types. Greater deviations are found on weekends (between 2013 or 2018 passive data and the 2008 OD survey). However, the coefficients of variations are always lower when the same years are compared (2013 survey vs 2013 passive data and 2018 survey vs 2018 passive data). This confirms, as expected, that two sources of data covering the same year are more similar than those describing different years. Hence, weekend information in the 2013 and 2018 travel surveys would probably have improved the comparison of the variations on Saturdays and Sundays between survey and passive data.

4.4.3.3 Periodicity

The curves (or time series) were also produced at a more disaggregated level to monitor the number of trips or validations or counts every hour of each day of the study period. Periodograms such as the one in Figure 4.12 (for the subway with the 2008 OD survey) are then plotted to examine the spectral density (based on the Fourier transform) of the time series. Typical frequencies are detected, then periods are deduced (as the inverse of the frequencies) and sorted according to their spectral strength. For instance, the six most important periods highlighted for the subway with the 2008 OD survey and the 2013 and 2018 smart card data are reported in Table 4.4 (the other OD

surveys were not considered because they do not contain information on weekends; therefore it was more difficult to find some periodicity, especially the weekly seasonality, in the data).

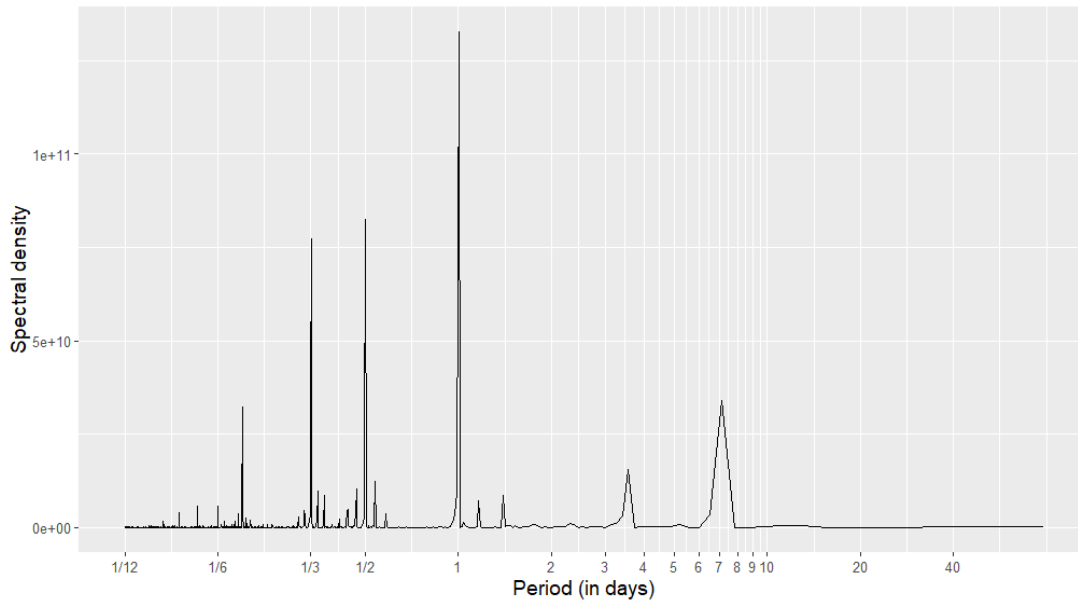


Figure 4.12 Periodogram of the hourly number of boardings in the subway (2008 OD survey)

Table 4.4 Periodicity analysis with spectral density: detected frequencies and periods for the subway from two data sources

Database	Order	Frequency	Spectral density	Period (hours)	Period (days)
2008 OD survey	1	0.042	1.33E+11	24.04	1.002
	2	0.083	8.25E+10	12.02	0.501
	3	0.125	7.72E+10	8.01	0.334
	4	0.006	3.41E+10	170.45	7.102
	5	0.125	3.32E+10	7.98	0.332
	6	0.209	3.23E+10	4.80	0.200
2013 passive data	1	0.042	2.57E+11	24.04	1.002
	2	0.125	5.96E+10	8.01	0.334
	3	0.083	5.20E+10	12.02	0.501
	4	0.006	4.49E+10	170.45	7.102
	5	0.125	2.80E+10	7.98	0.332
	6	0.209	2.57E+10	4.80	0.200
2018 passive data	1	0.042	3.09E+11	24.04	1.002
	2	0.125	7.69E+10	8.01	0.334
	3	0.083	6.98E+10	12.02	0.501
	4	0.006	5.44E+10	170.45	7.102
	5	0.125	3.72E+10	7.98	0.332
	6	0.209	3.17E+10	4.80	0.200

Exactly the same periods are found in the two data sources and their order is also almost identical: only the second and third most important periods are interchanged between the two types of data. Daily and weekly cycles (respectively the first and fourth most important periods) are captured in both passive and survey data. Typical periods of 12, 8 and 5 hours are also identified in each database. This means that the same repetitive patterns resulting from activity-travel dynamics are observed in the two data types. The results from the car and bicycle datasets are not presented in this paper but similar conclusions can be drawn: the same periods are detected in survey and passive data, but their order differs more between the databases than for the subway because the use of these modes is less regular.

4.5 Conclusion

This work is an example of benefiting from both survey and passive data to cross-analyze the variability of mobility behaviors. Such analyses are typically performed with either one of these data sources in the literature, but not with both. In this paper, the Origin-Destination household survey of Montreal and several passive data streams were processed to be compared throughout the 2008, 2013 and 2018 fall periods. First, a re-weighting method of the single-day samples of the regional survey was developed to extend the representativeness of the collected information to all days of the survey period (rather than on a “typical day”). The proposed method is a continuation of the method developed by Verreault et Morency (2011); we added a third original step to compensate the heterogeneity in the composition of the single-day samples (highlighted by the cited authors as a limitation of the previous method). Then, passive data were selected over the same period and time series were built from smart card validations (for subway) and count data (for car and bicycle), to be compared with the weighted numbers of trips aggregated by day for each mode in the survey. In this way, the proposed automated processes made the two data types comparable at the longitudinal level over several months in the fall. Moreover, average weeks were estimated using weeks of this time span. Both qualitative and quantitative comparisons were conducted between the two data types for the three modes. Indicators were calculated to cross-analyze the trend in all databases, as well as the variability by day type and the periodicity of the curves at the hourly level. Therefore, this paper provides tools to compare different kinds of temporal patterns between big data and traditional surveys, allowing the extraction of variations

regardless of the volume metric used in the dataset (trip, validation, count or other). They could be useful both in the field of transportation and in the field of time use and activity.

The results revealed similar linear trends over the fall (linked to seasonal variations), weekly variations, and periods in the two types of data, especially for the modes with more regular weekly patterns (i.e., car and subway). The comparison was less good for bicycle because the use of this mode is more variable and/or because of a lack of representativeness in both data sources, but a similar declining trend over the fall (explained by the less favorable weather conditions for cycling) was found in all databases. This outcome leads to several conclusions. First, the expansion method developed for single-day regional surveys is validated (at least at the systemic level). The corrected weighted person factors could then be used to study the evolution of different mobility indicators, such as the modal shares, throughout the fall. Secondly, the variability of travel behaviors over the study period is double validated (by the two types of data) and once again the concept of “typical weekday” is confirmed to be outdated and not representative of the complexity of mobility patterns. Even if the 24-hour period was found to be the most important period in the data, weekly rhythms were also highlighted and differences from the average working day were measured for all modes. This variability has policy implications and may bring recommendations on the data collection period for upcoming regional surveys. Mobility services may need to be adjusted (at least to the day of the week) for a better design. In this way, it could be better to carry out a one-week regional survey, but as the actual Montreal survey already contains large samples for all day types, the existing data may be used with corrected weighting factors (after further validations of the method). Only information on weekends may be required as it was observed in this paper (as well as in other works) that travel behaviors are very different from weekdays and more variable on weekends. Finally, the common variations reported in the two types of data demonstrate that links can be established between emerging and traditional data. This similarity was important to check before trying to combine them. Indeed, this validates the applicability of variations detected in passive data to survey data since there are at least common fluctuations (i.e., they capture similar patterns) over several months.

Nevertheless, there is still room to improve the comparison and validation of travel behavior variability from multiple data sources. The study period of this work is short: a longer time span would be preferable, especially to better compare the trends in the data (which, moreover, could be non-linear) and annual seasonality. To this end, longer surveys would be more appropriate, but

they are expensive and difficult to carry out. Moreover, week-end information was missing in the surveys used in this paper (except in the oldest) and therefore analogy with survey data was poorly measured. Last but not least, all the analyses of this paper were done at the aggregate level, while several authors demonstrated the small contribution of systemic day-to-day fluctuations compared to intra-/inter-individual variations (Chikaraishi et al., 2010; Pas, 1987; Raux et al., 2016). Aggregation was necessary to allow for longitudinal comparisons between the two data types because of their intrinsic differences, especially the different units of observation (validations, counts or trips) and the different time scales. Since the OD survey consists of one-day travel diaries, we could not trace each respondent over time, but the developed reweighting method allowed us to estimate an aggregated number of trips for each day of the survey period. In this way, we compared daily volumes, but not individual behaviors. Disaggregate longitudinal comparisons between survey and passive data would involve other databases. A multi-day or multi-week survey would be required to capture the variability of each respondent's behavior. In addition, some specific passive datasets, e.g., smart card data (without ticket validations), cellular data or other data with repeated information per identifier, would be more appropriate (than count data for instance), but this would imply several assumptions (such as one smart card = one person) that are not always true. Nevertheless, it would remain complicated to match a smart card or a phone with a survey respondent as there are no direct links or common attributes.

Further work will focus on leveraging the similarity between the Montreal OD survey and passive datasets for fusion purposes. Indeed, the two types of data should be used in complementarity as lots of (cross-sectional) information are gathered in the survey data while the passive data are available over longer periods. Data enrichment methods could be based on the detected common rhythms (instead of common variables): for instance, periodograms could be drawn with the survey data per travel purpose or destination, and this information may be conveyed to the smart card data using probabilities to observe certain periodicity for each travel purpose or destination. More directly again, these rhythms could also help match smart cards with survey respondents by identifying similar travel patterns (intra-individual variations) at a more disaggregated level, especially on Thursday as it was found the most similar day between the two data types. Furthermore, data fusion methods could be developed for long-term longitudinal monitoring of mobility behaviors: the variations detected in the longitudinal passive data (long-term trend, annual seasonality and weekly seasonality) could be applied to the average use of the modes in the surveys

to annualize and forecast modal shares beyond the fall. Therefore, the gap between two cross-sectional surveys could be filled thanks to operational data. In this way, the role of traditional surveys would be to collect average patterns and contextual information, while passive data would complement them by capturing variability over the year and between two surveys.

Acknowledgments

The authors wish to thank all the providers of the passive datasets used in this paper: *STM - Société de Transport de Montréal* (Montreal transit authority) for smart card data, *Ville de Montréal* (municipal organization of the city of Montreal) for cycling count data and *MTQ - Ministère des Transports du Québec* (Ministry of Transport of Quebec) for traffic count data, as well as the *ARTM - Autorité Régionale de Transport Métropolitain* (Metropolitan regional transportation authority) for giving access to the household surveys. They also acknowledge the financial support of the Canada Research Chairs (CRC), Quebec Research Funds (FRQNT) and Institute for Data Valorization (IVADO) programs.

CHAPITRE 5 ARTICLE 2 : COMBINING A REGIONAL HOUSEHOLD SURVEY AND PASSIVE DATA STREAMS FOR LONGITUDINAL MONITORING PURPOSES

Elodie Deschaintres, Catherine Morency, Martin Trépanier

Accepté pour publication dans *Transportation Research Procedia : 12th International
Conference on Transport Survey Methods*

Date de soumission initiale : 29 avril 2022

Abstract

More and more passive big data are available, but travel surveys are still widely used in transportation. This coexistence brings questions about the role of each data source and new data fusion challenges. Within this context, this paper aims to combine the Montreal household survey and passive data streams to provide longitudinal monitoring of aggregated travel behaviors. The proposed methodology, based on time series decomposition, consists in applying the trend and seasonal variations detected in passive data to the typical fall weekday observed in the survey. This allows traditional mobility indicators, such as modal shares, to be projected and annualized.

Keywords: cross-sectional survey; passive big data; data fusion; time series decomposition; typical day; longitudinal monitoring

5.1 Introduction

Large-scale household travel surveys have long been conducted to assist in transport and mobility planning. In addition, over the past few years, the rapid development of information and communication technologies have made available a wide variety of passive datasets, from GPS coordinates, sensor and count data to cell phone, social media, smart card or other travel transactional data (e.g., shared mobility services). However, despite this multiple-source data landscape, transport research is still suffering from a lack of both longitudinal and multimodal data. Indeed, there is no single database describing the continuous use of all modes of transport at the same time. On the one hand, trips made with all modes are gathered in traditional surveys, but multiple low-frequency and irregular modes are not sufficiently observed. Furthermore, the information is cross-sectional (i.e., collected at specific points in time, with gaps of several years between them), and the trip diaries are pooled to represent an average “typical” weekday. Therefore, such data does not account for the variability of mobility behaviors, and the compiled information quickly becomes dated and less representative of current conditions. On the other hand, passive data are longitudinal, hence available continuously over long periods, but they are collected independently and heterogeneously for each mode; thus, they make the analysis of interactions between modes difficult.

Data fusion is one solution put forward in the literature to take advantage of all data sources and compensate their respective limitations. In this line, this paper aims to enrich a regional household travel survey (the Montreal Origin-Destination survey, conducted every five years) with passive data streams to allow longitudinal monitoring of the use of several modes of transport. It is the continuation of a previous work that demonstrated there are large similarities between the Montreal survey and passive datasets regarding the patterns of use of three modes, expanded and aggregated over several months; the same fluctuations, especially the weekly rhythms, were revealed in the two data types (Deschaintres et al., 2022). This first analysis validated there is potential to combine the two data sources, and the next step is to apply variations detected in passive data to mobility indicators measured with traditional surveys to cover a longer period. To this end, this paper proposes an automated procedure based on time series decomposition. This method is applied to the 2013 Montreal survey and validated with the 2018 survey. The results allow annualizing the

daily modal shares of three modes and following their evolution over five years, between the two travel surveys, based on the seasonality and the long-term trend highlighted in passive data.

The remainder of this paper is organized as follows. First, some background elements regarding data fusion are provided. Then, the methodology is described, including all datasets used in this work and the processing steps. Finally, the results are exposed, and the paper is concluded by emphasizing the contributions and suggesting further investigations.

5.2 Background

Data fusion (or integration) is not new in transportation. Surveys have long been combined, for instance to merge revealed and stated preference data (Hensher, D. et al., 1998), or census and travel survey data (Venigalla, 2004), the latter case being also involved in the weighting of sampled data. However, the use of this concept has intensified in recent years. Three causes were advanced by Bayart et al. (2009). First, travel surveys have become less representative because of declining response rates and less complete/accurate answers. This trend can be explained by different factors, in particular the exposure to a growing number of survey requests, causing fatigue effects, but also an increasing anxiety about revealing personal information. In addition, difficulties are encountered in reaching certain population subgroups. Therefore, recent examples of survey combination include merging several survey modes, or integrating a special-purpose satellite survey (e.g., targeting students) with a central household travel survey (Verreault & Morency, 2018; Wang, Kaili et al., 2021). The second cause for the growing interest in data fusion according to Bayart et al. (2009) is the increasing awareness of the complexity of urban phenomena. This leads researchers to cross transportation databases with data from different fields, such as urban planning and health, or from economic, environmental, and social dimensions. Finally, the third cause (the one behind this paper) is the increasing availability of observation systems to collect passive data.

This emerging data is commonly seen as an opportunity to complement traditional data, especially surveys (Bonnell & Munizaga, 2018; Miller et al., 2018). The comparison of the two data types (passive data versus surveys) highlights the potential benefits of merging them. Surveys are cross-sectional and often based on single-day trip diaries, whereas emerging data is longitudinal and provides very precise spatial-temporal information. There are alternatives to cross-sectional surveys, e.g., multiday, panel and continuous surveys - see (Ampt, 2013; Ortúzar et al., 2011) for

definitions and a review -, but they are much more expensive and laborious than emerging data, collected passively and therefore without respondent burden. However, emerging data is unidimensional, focused on mobility without contextual information (due to confidentiality and privacy concerns), while traditional surveys contain socio-demographic attributes of the respondents, which are necessary to explain mobility behaviors (Bayart et al., 2009; Cherchi & Bhat, 2018). Consequently, emerging data and traditional surveys answer the questions “how?” and “why?” respectively (Callegaro & Yang, 2018). Other arguments for why emerging data cannot replace traditional surveys include concerns about data quality, completeness, and representativeness, as well as data access (Bonnell & Munizaga, 2018; Miller et al., 2018). In particular, few works benefit from enough data to study simultaneously different transport services (Morency et al., 2018). Furthermore, emerging data is generally massive (big data) and therefore difficult to handle, and data collection methods are still-evolving (Miller et al., 2018).

Therefore, both data types have pros and cons which can be exploited in a data fusion framework to maximize their value and overcome their limitations. The goal of the data fusion problem is to convey additional information from a donor database to a receptor database thanks to variables that are common across both databases. Data fusion procedures can be classified in two categories: exact matching, when two records (one from each database) are matched using a unique identifier, or statistical matching, based on explicit or implicit models (Bayart et al., 2009; Zhu et al., 2018). However, there are few applications of these methods in a transportation context. This lack may be due to several technical challenges, such as the heterogeneity of data formats and usage units (or travel indicators). In particular, as passive data was initially produced for functions other than mobility analysis and modeling, it does not contain trips as defined in travel surveys but transactions, counts, GPS points or other specific information (Bonnell & Munizaga, 2018). Beyond this semantic dissimilarity, there are also temporal and spatial incompatibilities between the data sources, which may have been collected with different resolution levels. In particular, the catchment areas and the corresponding reference population are different between travel surveys and passive data (residents of the whole region versus users, including residents and visitors, of a given system) (Miller et al., 2018). All these disparities lead to a lack of common variables and to the need of data aggregation. Therefore, many authors agree that new tools are required to allow the integration of various data sources, as well as to better define the role of both traditional and emerging data (Cherchi & Bhat, 2018).

The most common applications of data fusion with passive data can be found in road traffic engineering and intelligent transportation systems (El Faouzi & Klein, 2016). Several passive datasets were also used together to cross-validate or complete some information, such as origins and/or destinations (Giraud et al., 2016; Lovelace et al., 2016). However, for the reasons previously mentioned, examples of data integration between passive data and traditional surveys are much sparser. Nevertheless, some authors mined household surveys and smart card data for comparison or validation of spatial and temporal distributions (Munizaga et al., 2014; Spurr et al., 2015), and enrichment procedures (model-based or by implicit matching) were developed to add trip attributes, such as trip purpose (Kusakabe & Asakura, 2014), or sociodemographic information (Grapperon et al., 2016), from surveys to passive data. Therefore, these works capitalized on a travel survey (donor) to complete information in big data (receptor) or both data sources (both as donor) were valued for validation/comparison purposes, but none of them conveyed information from passive data (donor) to a traditional survey (receptor). Yet, the core-satellite data collection framework proposed in several papers (Miller et al., 2018; Wang, Kaili et al., 2021) consider the household travel survey as the core and passive datasets as ancillary. This is also the approach adopted in this work, which aims to apply the variations captured in longitudinal data to a cross-sectional survey.

5.3 Data and methodology

5.3.1 Data description and processing

This paper benefits from two types of data. First, the Origin-Destination (OD) regional travel survey of Montreal, Canada, is a telephone (and web since 2018) survey conducted every five years in the metropolitan area of Montreal to record the trips of about 4-5% of all resident households over one weekday in the fall. The last two surveys, namely the 2013 and 2018 ones (78,731 and 75,266 households respectively), are used in this work. Both were weighted by applying a margin calibration method (category totals being drawn from the Canadian census) to portray representative travel behaviors on a typical weekday. Therefore, they provide cross-sectional information; they allow calculating average daily mobility indicators, such as modal shares, for the 2013 and 2018 falls.

Secondly, several passive data streams are harnessed to describe the longitudinal variations in the use of three modes of transport from January 1, 2013 to December 31, 2018: validation data for the subway, and count data for bicycles and private cars. The subway network of Montreal is composed of 4 lines and 68 stations. STM (*Société de Transport de Montréal*), the transit authority, provided us with all tap-in validations (regardless of the support: smart card or ticket) collected by the automatic fare collection system, called the OPUS system, at the entrance of all subway stations over the study period. Cycling data comes from the Eco-counter system of the municipal organization of the city of Montreal (*Ville de Montréal*), while highway and bridge counters from the CIR system of the Quebec Ministry of Transportation (MTQ: *Ministère des Transports du Québec*) are handled for traffic data. The cycling counts include personally owned bicycles and (station-based) bikesharing, operated by BIXI in Montreal, while only private car counts (without trucks) are extracted from the classified traffic data. However, in both datasets, there are only a few counters with almost continuous information between 2013 and 2018. Two counters by mode are selected, based on the number of missing observations and on their location (counters on bridges are preferred for the car because they are typically less affected by congestion). The cycling counters are located near the city center, while the car counters are positioned on bridges connecting the island of Montreal to the suburbs. Therefore, by their location, these counters mainly capture commuting trips, which are typically well represented in the Montreal survey (Spurr et al., 2015).

Some data preprocessing steps are required to aggregate the use of each mode (in number of validations or counts) into daily times series over six years. Validations and counts are summed by 24 hours, from 4:00 a.m. to 28:00 p.m. (4:00 a.m. the next day): this is the definition of a day in the OD survey, where trips collected from 24:00 to 27:59 are trips that began the day after the one being surveyed and that typically allow for the conclusion of a travel chain. When validations are missing over several hours and for several stations, the daily sum is also considered missing in the subway data. The selected cycling and car counts are first summed by counter, i.e., without distinguishing the direction, then by mode. When a daily observation is missing for one counter, then the sum with the other counter is also considered missing. At the end, $365 \times 5 + 366 = 2191$ -day time series per mode are obtained, with 3, 8 and 122 missing values for the subway, cycling and private car respectively. Therefore, all datasets used in this work are incomplete, but also heterogeneous in nature (validations and counts in the passive data, versus trips in the OD

survey). Moreover, the cycling and car daily volumes are not representative of the whole region. However, the developed method allows for missing values and outliers, and only the variations in the passive data (neither the values nor the metric) are of interest in the proposed framework.

5.3.2 Method overview

The fusion procedure developed to combine the Montreal OD survey and passive data streams is presented in Figure 5.1.

For each mode of transport:

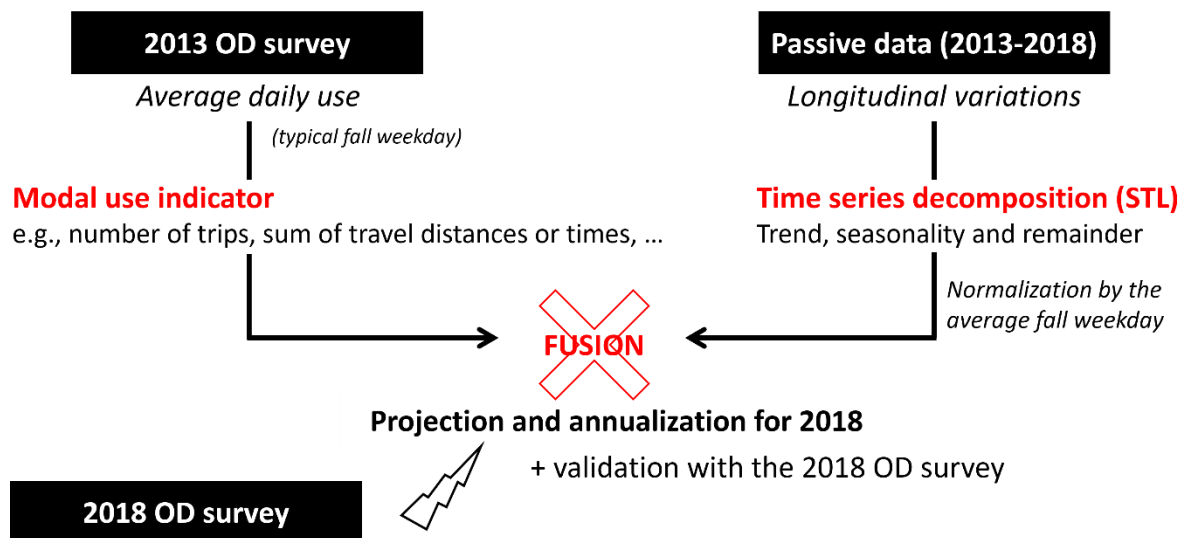


Figure 5.1 Data fusion procedure between the Montreal Origin-Destination (OD) survey and passive data

This consists in applying the longitudinal variations detected in the passive dataset (converted into a daily time series) to the typical daily use observed in the OD survey, e.g., the 2013 survey, for each mode of transport. The variations in mode use measured in the passive data, available from 2013 to 2018, are highlighted using a time series decomposition method. This method breaks down the data into three types of components (a long-term trend, seasonality, and remainder) linked by an additive or multiplicative equation, depending on whether the components are added or multiplied together to return to the original data. More specifically, the STL (Seasonal and Trend decomposition using Loess) algorithm is employed. This approach is detailed in the following section. The components are then normalized by the average fall weekday, leading to standardized coefficients which are applied to a modal use indicator, such as the number of trips made with a

given mode, calculated from the 2013 OD survey. This results in projecting and annualizing the typical use indicator observed in the 2013 fall over six complete years, including day-to-day fluctuations. The latest OD survey (2018) is finally used to validate the results by comparing the predicted average number of trips in the 2018 fall to the value estimated from the survey. In this way, the longitudinal patterns extracted from the passive data allow filling the 5-year gap between the two surveys, and accounting for the variability of mobility behaviors over time rather than a typical fall weekday.

5.3.3 Seasonal and Trend decomposition using Loess

There are several time series decomposition methods, but the STL method, developed by Cleveland et al. (1990), is chosen in this work because it has many advantages over the others. In particular, STL is more flexible and allows for missing values, which are present in our data. The goal of the method is to decompose a time series Y_t into a trend component T_t (i.e., low frequency variations in the data and long-term changes in level), one or several seasonal components S_t (i.e., variations in the data given a seasonal frequency) and a remainder component (i.e., residuals) R_t . The additive form is written as follows, for all t observations:

$$Y_t = T_t + S_t + R_t$$

STL does not provide facilities for multiplicative decomposition, but this can be reached by taking the logarithm of the data, such as:

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

To get this decomposition, STL applies a sequence of smoothing operations based on locally weighted regression, or Loess (i.e., local estimation of a polynomial of degree d for each value from its q nearest neighbors, with weights decreasing with the distance from the value). The full algorithm is detailed below. It encompasses an inner loop, made up of several steps to estimate the trend and the seasonal components, nested inside an outer loop, which calculates the remainder component and can reduce the influence of aberrant points (outliers); thus, it allows for robust estimation. If needed, robustness is achieved by iterated weighted least-squares fitting in the loess smoothing operations: the neighborhood weights are multiplied by robustness weights (the largest values in absolute of the remainder are given the lowest robustness weights).

STL - GENERAL ALGORITHM

Outer loop (n_o iterations)**Inner loop** (n_i iterations)

- Step 1: Detrending $\rightarrow Y_t - T_t$
- Step 2: Cycle-subseries (for seasonality) smoothing with $q = n_s$ and $d = 1 \rightarrow C_t$
- Step 3: Low-pass filtering of the smoothed cycles-subseries with $q = n_l$ and $d = 1 \rightarrow L_t$
- Step 4: Detrending of the smoothed cycles-subseries $\rightarrow S_t = C_t - L_t$
- Step 5: Deseasonalizing $\rightarrow Y_t - S_t$
- Step 6: Trend smoothing with $q = n_t$ and $d = 1 \rightarrow T_t$

Calculation of $R_t = Y_t - T_t - S_t$ and robustness weights for each observation (outlier detection)

The inner loop is composed of six interconnected steps which aim to prevent the trend and seasonal components to compete for the same variations in the data. First, the data is detrended, the trend being zero in the initial pass of the loop (step 1). The detrended series is then used in steps 2, 3 and 4 to assess the seasonal component. n_p , the number of observations in each period (or cycle) of the seasonal component, depends on the data properties. For instance, if the series is monthly with an annual periodicity, then $n_p = 12$. Cycle-subseries are the subseries of values at each position of the seasonal cycle, for instance the January values, the February values, etc. These cycle-subseries are first smoothed by loess and all smoothed values are gathered in the temporary series C_t (step 2). Next, a filter consisting of the application of several moving averages followed by a loess smoothing is applied (step 3), which leads to L_t . Finally, the smoothed cycle-subseries are detrended by subtracting L_t from C_t , resulting in the seasonal component S_t (step 4). The initial data are then deseasonalized (step 5), and the trend component is obtained by loess smoothing of the deseasonalized series (step 6). Therefore, three loess smoothing operations are included in the procedure (steps 2, 3 and 6). If the time series has two or more seasonal components, the procedure is repeated: the components are successively estimated (from the shortest-period component to the longest-period component), subtracted out, and the next component is determined from the residuals. The final trend component is then computed at the end of all iterations.

STL - DECOMPOSITION PARAMETERS

n_p : number of observations in each period or cycle of the seasonal component based on the data properties
n_o : number of iterations in the outer loop $n_o = 5$ or 10 if robustness is needed, $n_o = 0$ otherwise
n_i : number of iterations in the inner loop $n_i = 1$ if robustness is needed, $n_i = 2$ otherwise
n_s : smoothing parameter of the seasonal component odd integer greater than or equal to 7; based on diagnostic plots and/or prior knowledge
n_l : smoothing parameter of the low-pass filter $n_l = [n_p]_{odd}$
n_t : smoothing parameter of the trend component $n_t = [1.5n_p/(1 - 1.5/n_s)]_{odd}$
<u>Notation:</u> $[x]_{odd}$ is the smallest odd integer greater than or equal to x

Cleveland et al. (1990) prescribed values for almost all parameters (number of iterations and smoothing parameters) of the procedure, based on the mathematical properties of the decomposition. Note that all polynomial degrees involved in the smoothing operations are equal to 1; this means that the components are locally linear. Only the choice of n_s , the smoothing parameter of the seasonal component (i.e., the number of consecutive observations to be used when estimating C_t at step 2), which is intrinsically linked to n_t , the smoothing parameter of the trend component, is less straightforward. This parameter controls the amount of variation in the data that makes up the seasonal component. In other words, it determines how rapidly the component can change. The cycle-subseries can be plotted to decide how smoothly the values should evolve from one cycle to another (the higher n_s is, the smoother or lower the evolution). Therefore, the choice of n_s is generally based on diagnostic plots and a priori knowledge of the data.

See (Cleveland et al., 1990) for more methodological details and (Hyndman & Athanasopoulos, 2018) for practical applications using the R statistical software package *fpp2*, that is used in this work. Furthermore, Wang, Xiaozhe et al. (2006) proposed an indicator whose value is between 0 and 1 to measure the strength of each component:

$$F_T = \max_t \left(0; 1 - \frac{\text{Var}(R_t)}{\text{Var}(T_t + R_t)} \right) \text{ and } F_S = \max_t \left(0; 1 - \frac{\text{Var}(R_t)}{\text{Var}(S_t + R_t)} \right) \text{ where } \text{Var} \text{ is the variance}$$

The strength of the component is higher when F is closer to 1, i.e., when the deseasonalized data ($T_t + R_t$) or the detrended data ($S_t + R_t$) have much more variation than the remainder component (R_t).

5.3.4 Projection and annualization

In this paper, the components extracted from the passive data with the STL method are used to calculate projected and annualized curves (in daily number of trips) by mode. The projected curve, defined at all t observations (days) from 2013 to 2018, is equivalent to the trend component standardized by its average value over the 2013 OD survey period ($\bar{x}$ being the notation for the mean), and then applied to the typical number of trips calculated from the 2013 OD survey:

$$Projection_{t(2013 \rightarrow 2018)}^{mode} = \frac{T_{t(2013 \rightarrow 2018)}}{\bar{T}_{t(2013 \text{ OD survey period})}} * \#trips_{2013 \text{ OD survey}}$$

The predicted number of trips in the 2018 OD survey is computed by averaging the resulting values over the 2018 OD survey period. Therefore, this method allows estimating a projection that is not affected by seasonality (which is also the case in the OD survey), nor by data outliers.

Similarly, the annualization for a specific year (e.g., 2018) without the trend influence is made by applying the seasonal component (which can be a sum or a multiplication of several components), standardized by its average value in the 2018 OD survey period, to the typical number of trips calculated from the 2018 OD survey. The equation differs between the additive and multiplicative forms, the former being more complicated and requiring additional standardization to be able to add values with the same metric. For both forms, the equation is formulated so that the mean of the annualized curve over the 2018 OD survey period is equal to the typical number of trips observed in the 2018 OD survey.

$$\begin{aligned} Annualization [add]_{t(2018)}^{mode} &= \#trips_{2018 \text{ OD survey}} + (S_{t(2018)} - \bar{S}_{t(2018 \text{ OD survey period})}) * \frac{\#trips_{2018 \text{ OD survey}}}{\bar{Y}_{t(2018 \text{ OD survey period})}} \\ Annualization [mult]_{t(2018)}^{mode} &= \frac{S_{t(2018)}}{\bar{S}_{t(2018 \text{ OD survey period})}} * \#trips_{2018 \text{ OD survey}} \end{aligned}$$

5.4 Results

5.4.1 Passive data decompositions

The STL method is first applied to each passive dataset. Since the data was aggregated at the daily level, two seasonal components are estimated: one for the weekly seasonality and one for the annual

seasonality, with $n_p = 7$ and $n_p = (365 * 5 + 366)/6 = 365.17$ respectively. The calculation of the annual periodicity is based on the study period, composed of five years with 365 days and one year (2016) with 366 days. Note that it is not possible to delete February 29, 2016, without impacting the weekly seasonality. The crucial choice of n_s is supported by diagnostic plots (available upon request to the authors). The changes from one year to another or one week to another are assumed to be slow (or smooth); thus, n_s should be high. In addition, an iterative loop (successively testing different odd values of n_s from 7 to 99) is executed to find the optimal value in the validation of the procedure in Figure 5.1, i.e., giving the best projection for the 2018 fall average indicator (or the lowest error). The selected parameters are provided in Table 5.1 for each mode, decomposition form and seasonal component. All other parameters are determined using the prescribed values. Robust estimation ($n_o = 15$) is chosen because there may be outliers in the passive data, especially in the cycling and car counts.

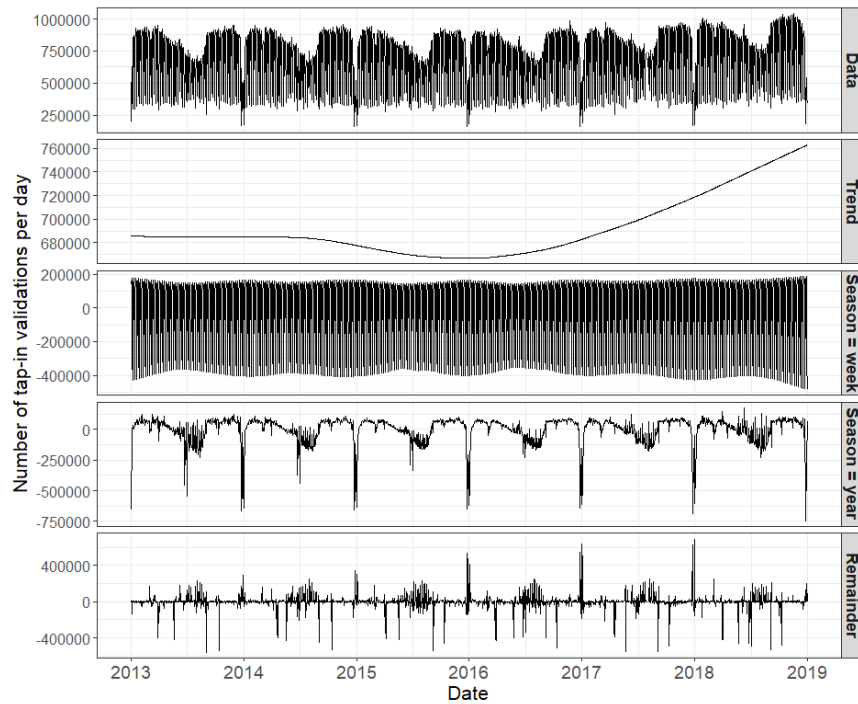
Table 5.1 Selected values of the smoothing parameters of the seasonal components for each mode and decomposition form

Decomposition form			ADDITIVE		MULTIPLICATIVE	
Mode	/	Season	Week	Year	Week	Year
SUBWAY			33	33	45	79
BICYCLE			7	7	47	79
CAR			13	7	19	53

The results of the application of the method to the subway tap-in validations are displayed in Figure 5.2, for both (additive and multiplicative) decomposition forms. The raw data (time series of the daily number of validations) is illustrated in the first row, while the following graphs represent the different components of the decomposition. Since the two decompositions are very similar, common comments are made below.

The trend component, plotted in the second position, indicates a slow growth in the total daily subway demand, except in 2015. This long-term curve can be compared to the annual ridership published by the STM in open official reports and similar profiles will be found. However, here the trend is described with one value per day, and not only one value per year. Note that the trend (and the seasonal components) is available at all days, even when data is missing; this is possible thanks to the loess smoothing.

(a)



(b)

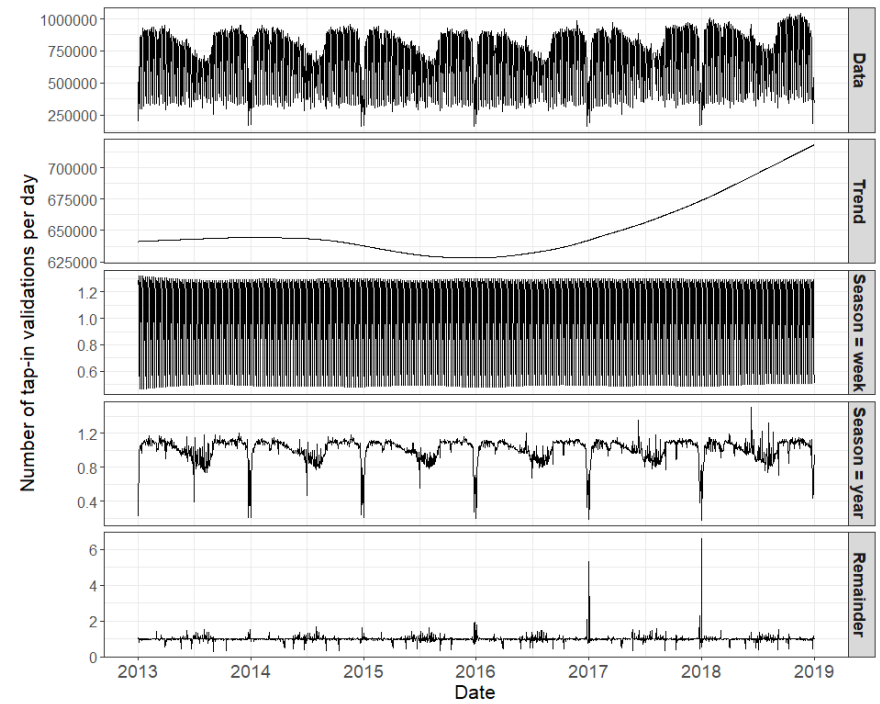


Figure 5.2 (a) Additive and (b) multiplicative STL decompositions of the subway tap-in validations

Unlike other decomposition methods, STL allows the seasonal components to change over time. Therefore, in this application, this makes it possible to capture variability from day to day, week to week and year to year. This flexibility allows us to see, for the subway, that the difference between weekdays and weekends is lower during the summer; the decrease in total ridership is rather considered in the annual seasonality. However, the corresponding seasonal component is quite regular from one year to another between 2013 and 2018. The most important dissimilarities are due to public holidays, which are not on the same day for all years.

Regarding the last component (remainder), we observe an increase in the residuals during the summer and holiday seasons. This could be explained by temporary factors, such as special events/activities or weather. As the estimation is robust, the impact of these specific cases on the calculation of the other components is reduced.

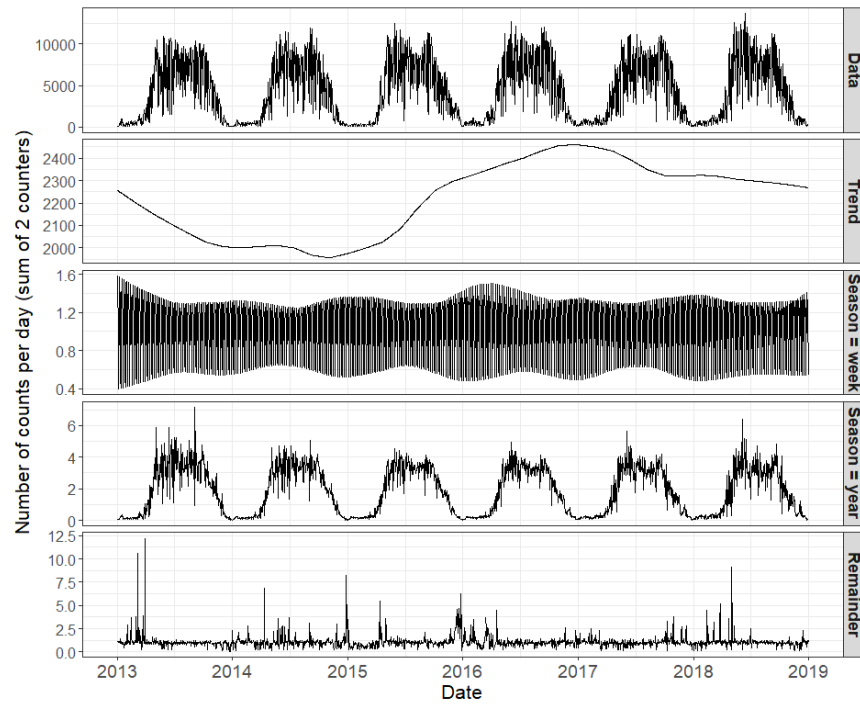
The decompositions obtained from cycling and car counts are exposed in Figure 5.3. Only the multiplicative form is presented because this form gives the best projection for 2018 (see the following section). For cars, the long-term trend is declining, with an acceleration in 2018. The seasonal components have points in common with those of the subway, but they are less regular. Indeed, the smaller n_s values in Table 5.1 indicate that changes are quicker. The usefulness of robust estimation is demonstrated since the unusual counts in 2015 have led to higher residuals. For bicycles, the trend is much less clear. The weekly seasonality varies over time, but the annual seasonality is very pronounced.

This is confirmed by the strength indicators in Table 5.2 (for the multiplicative decomposition only, but the results are very similar for the additive form), which reveal that the strongest annual seasonality is measured for cycling. The strength of the seasonal components is analogous between the subway and car, and important weekly rhythms are highlighted. However, the trend is much lower for all modes; this can be due to the length of the study period (6 years), which may not be sufficient to observe strong long-term changes.

Table 5.2 Strength of the trend and seasonal components (F indicator) in the multiplicative data decomposition of each mode

Component	SUBWAY	BICYCLE	CAR
Trend	0.05	0.05	0.07
Weekly seasonality	0.84	0.36	0.81
Annual seasonality	0.47	0.89	0.50

(a)



(b)

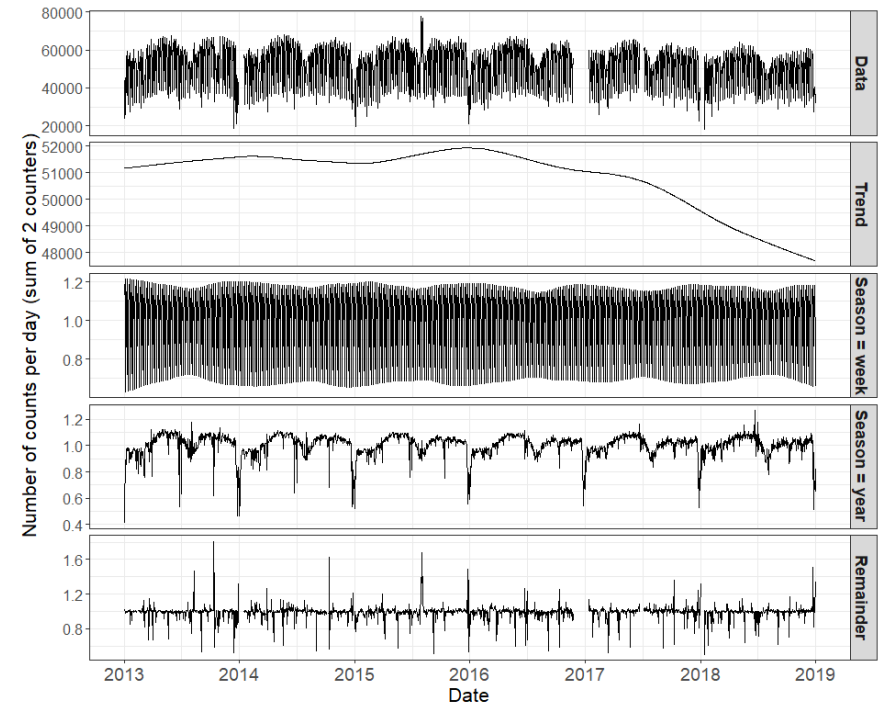


Figure 5.3 Multiplicative STL decomposition of the (a) cycling and (b) car counts

5.4.2 Projected and annualized modal shares

Using the trend of the previous decomposition for each mode, the projected average number of trips on the 2018 fall weekdays is evaluated and compared with the observed number in the 2018 OD survey. The resulting error rates between projection (with optimal choice of the parameter n_s) and observation are provided in Table 5.3 by mode and decomposition form. In addition, these percentages are compared with those obtained when applying the traditional (disaggregated) demand projection method of MTQ (Quebec Ministry of Transportation). This method projects the weighting factor of each individual trip according to the anticipated evolution of socio-demographic characteristics. A Furness adjustment is also made to modulate the spatial distribution of the work-related trips. The error reported in Table 5.3 is the comparison between the number of trips calculated with the 2016 projected factors (estimated from the 2013 OD survey) and the observed number in the 2018 OD survey (weighted with the 2016 census).

Table 5.3 Error rates between the projected and observed number of trips on the 2018 fall weekdays

Method	SUBWAY	BICYCLE	CAR
MTQ method (2016 projected factors)	1.8 %	4.5 %	8.6 %
Fusion procedure - Additive form	3.4e-04 %	11.1 %	6.5 %
Fusion procedure - Multiplicative form	1.8e-03 %	9.4e-04 %	6.4 %

In all cases (except one), the errors assessed with the proposed procedure are smaller than with the MTQ's traditional method. Therefore, our method works pretty well. This was not initially obvious because of all the semantic, temporal, and spatial disparities between passive data and traditional surveys which were described in the background and data section. For the subway, the difference is almost zero, regardless of the decomposition form, whereas the errors are larger for bicycles and cars, but the count datasets are of lower quality and these modes may be more difficult to forecast. Nevertheless, the multiplicative decomposition seems to perform better, that is why it is chosen for the following analyses. This better performance may be because the logarithm helps reduce the influence of extreme values, resulting in a more Gaussian distribution for local regression.

The projected and annualized curves derived from the multiplicative decomposition of the subway validations (by applying the trend or seasonal components to the typical number of trips observed in the OD survey) are shown in Figure 5.4. The black lines account for the numbers observed in the 2013 or 2018 OD survey.

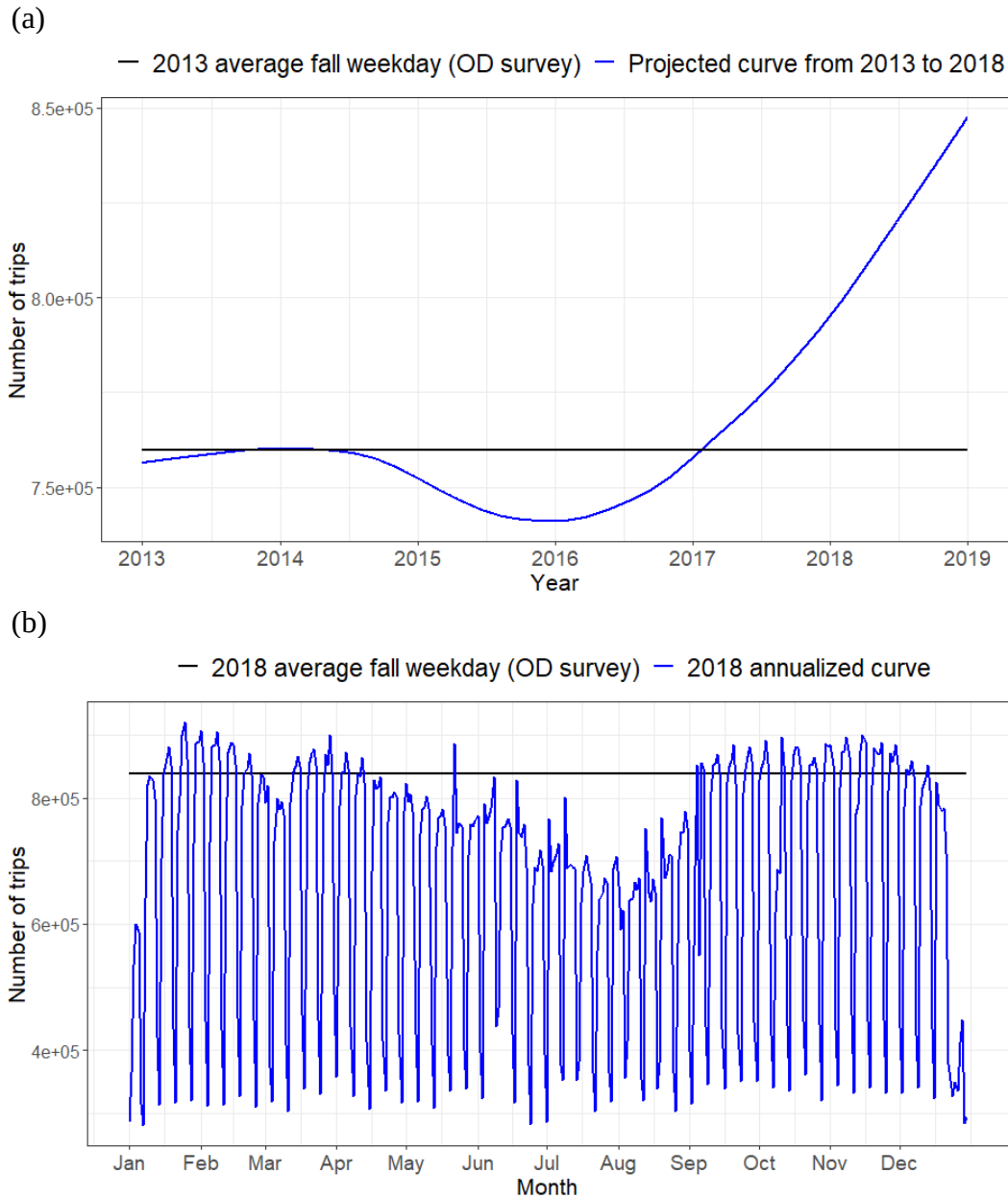


Figure 5.4 (a) Projected and (b) annualized curves of the total number of trips per day in the subway

The same curves are computed for each mode, then the corresponding modal shares (in proportions of total trips) are calculated. The projected modal shares from 2013 to 2018 are represented in Figure 5.5. Therefore, rather than having only one point in 2013 and one point in 2018 (i.e., the typical daily modal shares observed in the two surveys), the method allows for a longitudinal tracking of what happened during the 5-year gap. The variations are quite small, but this was expected because of the slow changes in aggregated mobility behaviors.

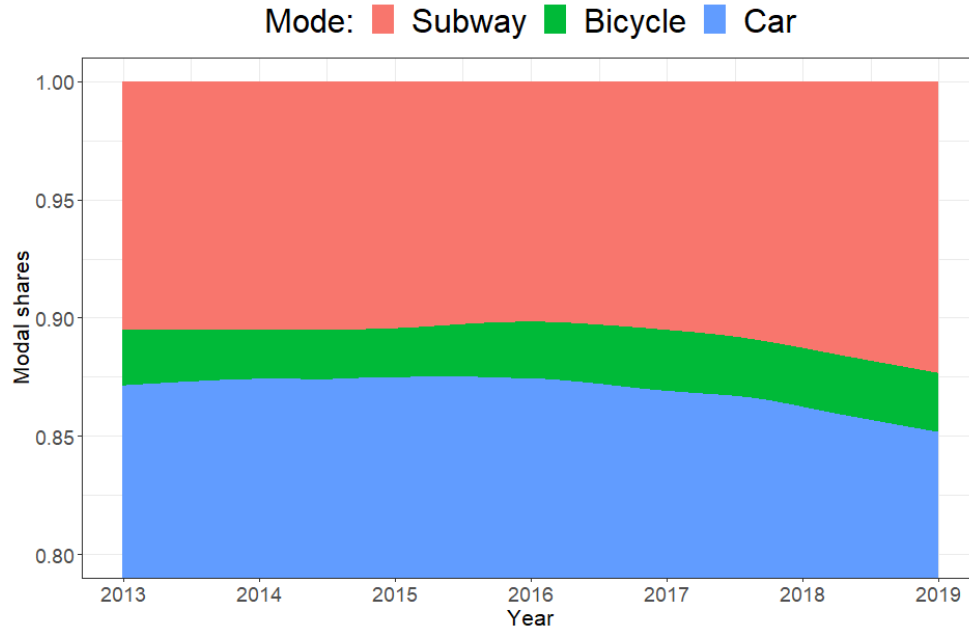


Figure 5.5 Daily modal shares of three modes projected from 2013 to 2018

The annualized (or seasonalized) modal shares for 2018, as well as the average week across the 2018 fall, are exhibited in Figure 5.6. Therefore, instead of an average number for the whole year, the method allows for a profile which varies by day of the week and day of the year. Looking at the variations by day type over an average week of the 2018 OD survey period plus weekends, we see that modal shares are not constant, even from Monday to Friday. This questions the representativeness of the “typical” weekday of the OD survey, already investigated by Verreault et Morency (2011).

In addition, the variations of the daily modal shares selected over the 2018 fall on weekdays only (hence without considering the variability between weekdays and weekends) are described in Table 5.4. The results prove there is a lot of variations around the mean, especially for cycling.

Table 5.4 Variations across the 2018 fall of the daily modal shares (on weekdays only)

MODE	Minimum	Mean	Maximum
SUBWAY	7.8 % (- 33 %)	11.5 %	13.0 % (+ 13 %)
BICYCLE	0.2 % (- 91 %)	2.3 %	6.3 % (+ 167 %)
CAR	82.6 % (- 4 %)	86.2 %	90.1 % (+ 5 %)

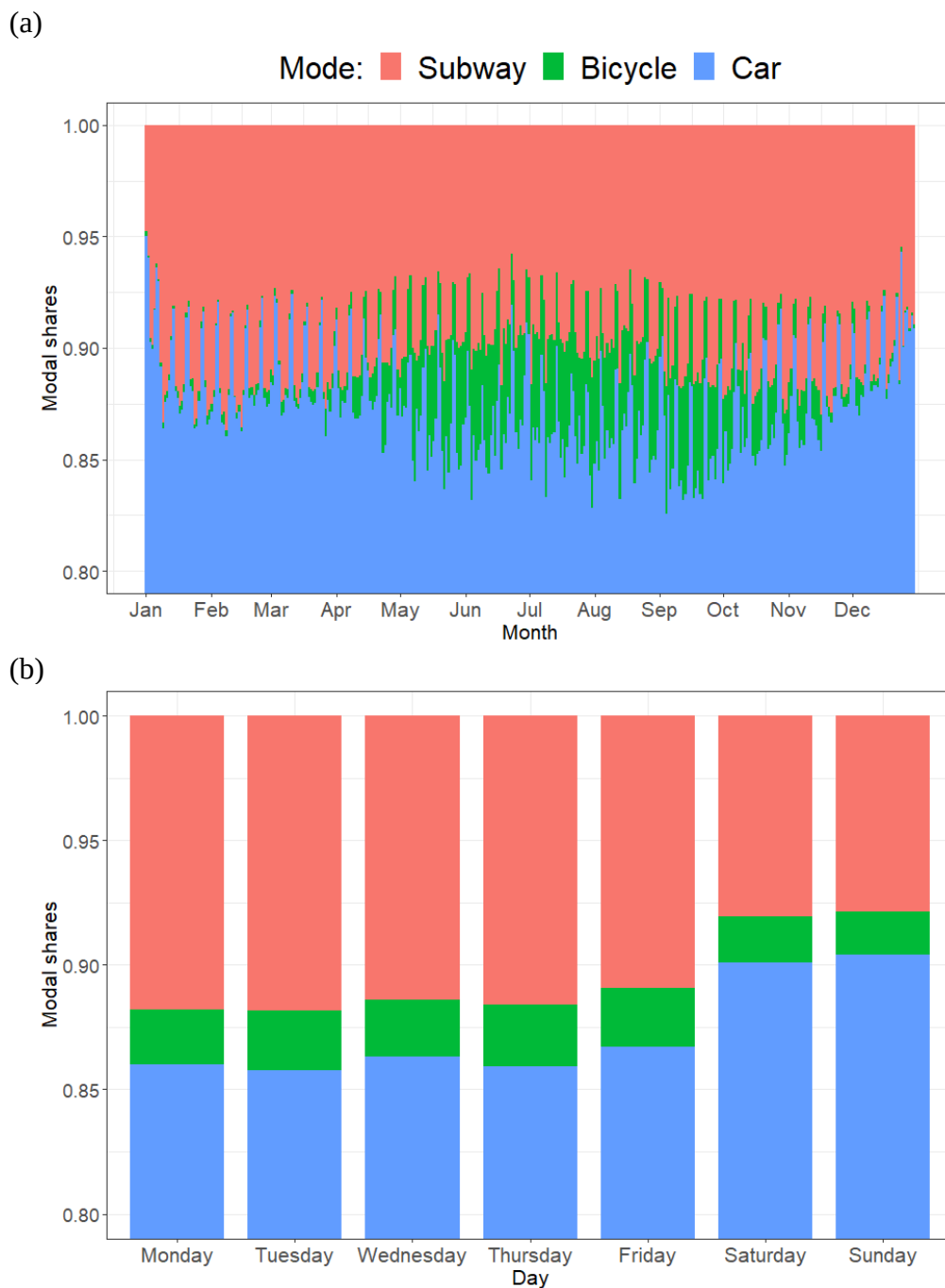


Figure 5.6 (a) Annualized daily modal shares of three modes in 2018 and (b) average modal shares on a 2018 fall week

5.5 Conclusion

This work proposes a framework to convey temporal information from passive data streams (donor) to a cross-sectional single-day household travel survey (receptor). Combining these two

types of data, especially in that direction (i.e., to enrich a survey), has been little explored in the literature because of a lack of common variables between the data. However, in this paper, aggregation at the daily level allowed us to create links between them. The developed method, based on STL decomposition of time series, consists in applying longitudinal fluctuations (locally-linear trend and seasonality) extracted from passive data to the typical daily use observed in the survey for a given mode of transport. This method has suitable properties to work with passive big data: it allows for missing values, and can be robust to outliers, thus compensating data incompleteness and quality weaknesses. In addition, the computation is easy and fast, which makes it possible to deal with massive datasets. In this paper, the procedure was applied to the Montreal OD survey, conducted every five years on a typical fall weekday, for three modes: subway, cycling and private car. While it was only possible to calculate average numbers of trips by mode for the fall periods of 2013 and 2018 from the surveys, the method allowed monitoring their evolution over the 5 year-gap and estimating a profile which accounts for annual and weekly variations. In this way, modal shares, which are typically available on a one-off basis, can become continuous, and therefore both multimodal and longitudinal information can be obtained thanks to the complementarity between traditional surveys and emerging data. From this perspective, the role of traditional surveys is to collect average patterns of use of the modes and contextual information, whereas passive data supplements them by capturing variability over the year and between two surveys. This may obviate the need for longitudinal surveys (at least at the aggregate level), which suffer from additional economic and technical challenges. However, as new mobility services are still underrepresented in travel surveys, more information must be collected on these modes to include both traditional and emerging modes.

Further research involves testing the proposed framework with other modes, such as bus or taxi, and carrying out more disaggregated analyses: per subway station (spatial component), per fare product or for specific times (e.g., peak hours). The variations assessed around the mean thanks to the passive data could also be leveraged to develop variability factors to partially correct the current weighting factors in the OD survey and consider day-to-day or weekly variations over the fall. This could allow generating a distribution of values rather than an average value from the same data. In addition, modeling the residuals of the time series decomposition or integrating the estimation of a regression component in the STL procedure (by iterative weighted fitting) would be interesting to include the effects of weather, holidays, and events. Instead of a regression component, machine

learning methods (e.g., random forests or neural networks) could also be investigated and combined with the decomposition procedure (Qin et al., 2019). This would allow for non-linearity and nonparametric specification, thus overcoming two potential limitations of the current method. Going even further, a longitudinal model could be developed to model the deseasonalized times series (rather than simply observe it in the data) and then forecast the future trend. In this paper, the trend was adjusted (interpolated) with two observations, one from the 2013 OD survey and one from the 2018 OD survey, but the ultimate objective would be to project this curve beyond 2018 using the passive data to forecast the number of trips that will be collected in the 2023 OD survey. Predicting the trend effect may depend on slow time-varying predictors, such as population evolution, but also on specific events such as supply improvements and changes in legislation which are implemented at a given date. The latter could be modeled using structural change point detection (Yeh & Lee, 2019). This forecasting exercise is even more relevant than ever in the context of the COVID-19 pandemic. Indeed, the changes which are currently impacting travel behaviors make it more difficult to predict the modal shares that will be observed in five years. However, a trend-based method may no longer be valid when such unexpected events occur. With more recent passive data, the effect of the health crisis on the trend and seasonality could also be analyzed. Teleworking, widely adopted in Montreal since March 2020, may have a strong impact on the weekly rhythms. Different scenarios for projecting mobility could then be proposed.

Acknowledgements

The authors wish to thank all the providers of the passive datasets used in this paper: STM - *Société de Transport de Montréal* (Montreal transit authority) for smart card data, *Ville de Montréal* (municipal organization of the city of Montreal) for cycling count data and MTQ - *Ministère des Transports du Québec* (Quebec Ministry of Transportation) for traffic count data, as well as the ARTM - *Autorité Régionale de Transport Métropolitain* (Metropolitan regional transportation authority) for giving access to the household surveys. They also acknowledge the financial support of the Canada Research Chairs (CRC), Quebec Research Funds (FRQNT) and Institute for Data Valorization (IVADO) programs.

CHAPITRE 6 ARTICLE 3 : MEASURING CHANGES IN MULTIMODAL TRAVEL BEHAVIOR RESULTING FROM TRANSPORT SUPPLY IMPROVEMENT

Elodie Deschaintres, Catherine Morency, Martin Trépanier

Publié le 15 avril 2021 dans *Transportation Research Record*

Volume 2675(9), Pages 533–546, DOI : 10.1177/03611981211003104

Abstract

Despite the desired transition toward sustainable and multimodal mobility, few tools have been developed either to quantify mode use diversity or to assess the effects of transportation system enhancements on multimodal travel behaviors. This paper attempts to fill this gap by proposing a methodology to appraise the causal impact of transport supply improvement on the evolution of multimodality levels between 2013 and 2018 in Montreal (Quebec, Canada). First, the participants of two household travel surveys were clustered into types of people (PeTys) to overcome the cross-sectional nature of the data. This allowed changes in travel behavior per type over a five-year period to be evaluated. A variant of the Dalton index was then applied on a series of aggregated (weighted) intensities of use of several modes to measure multimodality. Various sensitivity analyses were carried out to determine the parameters of this indicator (sensitivity to the least used modes, intensity metric, and mode independency). Finally, a difference-in-differences causal inference approach was explored to model the influence of the improvement of three alternative transport services (transit, bikesharing, and station-based carsharing) on the evolution of modal variability by type of people. The results revealed that, after controlling for different socio-demographic and spatial attributes, increasing transport supply had a significant and positive impact on multimodality. This outcome is therefore good news for the mobility of the future as alternative modes of transport emerge.

Keywords: travel behavior; multimodality; modal variability; indicator; transport supply; causal inference; difference-in-differences method

6.1 Introduction

Mobility as a Service (MaaS) and multimodality (or intermodality) are increasingly fashionable concepts in urban transport research. They are generally presented as essential in the move toward more sustainable mobility, even more now that alternatives to the exclusive use of a private car are emerging. Consequently, many projects focus on the development of technologies (such as mobile applications) to integrate several transportation modes. However, there is a lack of knowledge about multimodal behaviors among travelers, their associated attributes, and motivations. Few tools have been proposed to quantify and explain mode use variability. It, therefore, remains difficult for practitioners to measure the impacts of strategies implemented to promote a more diversified mobility.

This paper aims at assessing the influence transport supply may have on the multimodality level of travel behaviors over a five-year period. This required first defining an indicator to summarize the intensities of use of several modes of transport. This indicator was calculated by type of people (aggregated by socio-demographic and spatial attributes) using the last two (2013 and 2018) Montreal household surveys. The evolution by type of people between the two surveys was then cross-analyzed with changes in transport supply. To this end, a difference-in-differences approach that allows causal inference was applied. As a result, this paper provides a means to guide interventions to increase multimodality.

This paper is organized as follows. First, some former studies into analyzing and quantifying multimodal mobility, as well as factors affecting multimodal travel behavior, are reviewed. The household surveys used in this paper are then described, and some methodological elements are exposed. The results are reported in three parts: the typology of people is presented; a multimodality indicator is then applied by type; and the causal effect of transport supply on the evolution of multimodality between 2013 and 2018 is modeled. Finally, the paper is concluded and the discussion feeds into the programming of future work.

6.2 Background

Few authors have investigated multimodality in travel behavior because of the lack of data: it requires multiday data for several modes of transport. Little information was historically collected on mode use variability in travel surveys. Those have mainly focused on car and transit, and non-

motorized trips were often not even recorded. Moreover, they generally did not include questions on the sequence of modes used, let alone the trip legs to get to (access travel) or from (egress travel) transit (Clifton & Muhs, 2012). Many surveys are still based on the concept of the “main transportation mode,” each trip being represented by a single central mode (Kagerbauer et al., 2015). In addition, mobility behavior needs to be monitored over several days to support multimodality study at the individual level, that is, the use of different modes of transport over time (Kuhnimhof et al., 2006). The work of the few authors who have benefited from such data is described below.

6.2.1 Analyzing multimodal mobility

The concept of multimodal mobility can be applied at the trip level, when several modes are used to make the same journey (also called intermodality), and at the individual level, thus defining multimodality as the use of several modes of transport by one person over a given period of time (Diana & Pirra, 2016). At the trip level, research has mainly dealt with specific combinations of modes of transport or long-distance traffic, but few studies addressed intermodal practices in daily mobility using specific surveys (Kagerbauer et al., 2015; Oostendorp & Gebhardt, 2018). However, it is on the second definition of multimodality (at the individual level) that most research has focused to better understand travel behavior in a context where mobility options are diversifying.

Some works were based on German national surveys: the MOP, a seven-day panel survey, and the MiD, a single-day cross-sectional survey but containing retrospective questions on the frequency of use of several transport modes (Kuhnimhof et al., 2010; Kuhnimhof et al., 2006; Nobis, 2007). In these papers, multimodality of people was defined as the use of more than two modes (including car, transit, and bicycle) over a week. Researchers also examined changes in multimodality (measured by the percentage of multimodal people). Generational analyses (Kuhnimhof et al., 2011) allowed a reduction in car dependency among young people to be detected in Germany and Great Britain, partly as a result of an increasing modal variability. Today’s young people tend to use more alternative modes, and thus have more diversified travel behavior. In the U.S. context, similar trends were revealed with the NHTS national survey (Buehler & Hamre, 2015). Going beyond simple descriptive analyses, more complex methods such as association analysis (Diana, 2012) or segmentation (Molin et al., 2016) were also applied to figure out different patterns of multimodal behavior.

Furthermore, multimodal travel behavior can be analyzed at the systemic level using passive data streams and data mining methods (Morency et al., 2020; Morency et al., 2018). The cross-analyses of the cited authors highlighted correlations of use between several transport modes in space and time.

6.2.2 Quantifying multimodality level

The measurement of multimodality in travel behavior at the individual level has received little attention. Moreover, the limitations of the most commonly used indicators were pointed out (Heinen & Chatterjee, 2015). Indeed, multimodality is mainly appraised with simple statistics such as the percentage of travelers using a given number of modes of transport. The frequency of use of each mode is often not considered. However, some more advanced indicators were recently proposed to measure variability of the intensities of use of several modes of transport. Indices used in other disciplines such as economics to evaluate the inequality or heterogeneity of distributed values were adapted to this end: for instance, the Gini index (Xiao et al., 2020), the Herfindahl-Hirschman index (Heinen & Chatterjee, 2015) or the Shannon entropy (Diana & Mokhtarian, 2009).

Most of these indicators were gathered and compared in a conceptual framework (Diana & Pirra, 2016). The authors of this paper evaluated their applicability in a transportation context according to various properties. By analogy with three concepts derived from the study of diversity in biology (*species richness, evenness, and dominance*), the authors emphasized the need to consider the number of transportation modes, the relative distribution of their intensities of use and the value of the intensity of use of the dominant mode. Two major issues in the quantification of multimodality were also discussed in this work. First, the choice of the measurement unit of the intensities of use is not trivial. Considering either the number of trips, trip stages, travel times, or distances could lead to different results. The second issue deals with the mode classification (hypotheses of independence of the alternatives). For example, should we consider transit as a mode, or distinguish between bus, subway, and train? These issues have not yet been addressed in the literature and will require sensitivity analyses.

6.2.3 Factors affecting multimodal travel behavior

Using the indicators reported before as a dependent variable of regression models or forecasting the membership to multimodal groups, some authors have sought to identify predictors of modal variability in individual travel behavior (Buehler & Hamre, 2015; Heinen & Chatterjee, 2015; Nobis, 2007; Scheiner et al., 2016). Socio-demographic variables were mainly investigated. According to the overall findings, people at both ends of the age pyramid (youths and elderly) are more likely to be multimodal. Education levels and income are also positively correlated with mode use diversity. However, unimodal car use is more prevalent among full-time employees and individuals with young children. Car access also has a negative impact on multimodality, while owning a bicycle and having a transit pass have conversely positive impacts. Moreover, life course events are significantly associated with changes in multimodal behavior (Scheiner et al., 2016). For instance, entering professional life and a child moving out of the household increase multimodality. Going even further, some authors coined the concept of “modality style” (Heinen & Chatterjee, 2015; Vij et al., 2011), which implies the predisposition of an individual to choose a set of travel modes according to their habits and lifestyle. Multimodal travelers would be more likely to change their behavior and have a more positive attitude toward alternative modes of transport (Heinen, 2018; Molin et al., 2016).

Mode choice variability can also be explained by exogenous temporal factors such as seasonality, social activities, service disruptions, or weather conditions (Lepage & Morency, 2019; Morency et al., 2018). On the space dimension, multimodality tends to be higher in densely populated urban centers, especially where access to a subway system is provided (Buehler & Hamre, 2015; Diana, 2012; Heinen & Chatterjee, 2015). This interaction with the urban environment is related to the greater availability of transport services. Many authors have explored the effect of transport supply and emerging mobility services on the use of other specific modes, mostly car or transit (Becker et al., 2018; Campbell & Brakewood, 2017). However, none of them have studied their effect on multimodality levels.

6.3 Data and methodology

Two types of data were handled for this paper. On the one hand, information on mobility behavior comes from the Origin-Destination travel survey of Montreal (Quebec, Canada). This is a telephone

(and web since 2018) survey conducted every five years in the metropolitan area of Montreal to record the trips of about 4% to 5% of resident households over one fall weekday. The last two surveys, namely those from 2013 and 2018 (78,731 and 75,266 households respectively), were harnessed in this work. Both were weighted using census data so that the results were representative of the metropolitan population. These surveys encompass variables at the individual (for all household members), household, and trip level. In particular, the sequence of modes used is available for each trip, modes considered being car (driver and passenger), subway, bus, school bus, train, taxi, cycling, walking, paratransit, and other modes (including intercity modes). On the other hand, transport supply information was found in several data sources. For each mode, cross-sectional information was collected for two years (2013 and 2018 when it was possible) to measure changes over a five-year period. The road network was represented using the 2011 and 2016 Canadian open census files. Transit planned services were examined thanks to the GTFS files of the transportation organizations of the Montreal area for the fall periods of 2013 and 2018. Moreover, bikesharing and carsharing transactional data were provided respectively by BIXI and Communauto, the Montreal companies of these services, for the same periods.

Based on these data, the proposed methodology (illustrated in Figure 6.1) includes the following steps. First, participants of each survey were grouped by socio-demographic and spatial attributes because one-day data are not enough to measure multimodality at the individual level (longitudinal travel diaries would have been required to deal with individuals). This paper, therefore, assesses multimodality by type of individual rather than by individual using aggregated intensities of mode use. As panel data were not available, this aggregation was also a way to measure changes over time between the same two types of people observed in two cross-sectional surveys five years apart. In this way, an indicator of mode use diversity was calculated by type and its evolution between 2013 and 2018 was analyzed. Finally, the best predictors of the multimodality levels estimated by the chosen indicator were determined. More specifically, we modeled the causal impact of transport supply improvement using a difference-in-differences approach. More details on the methods applied are given below.

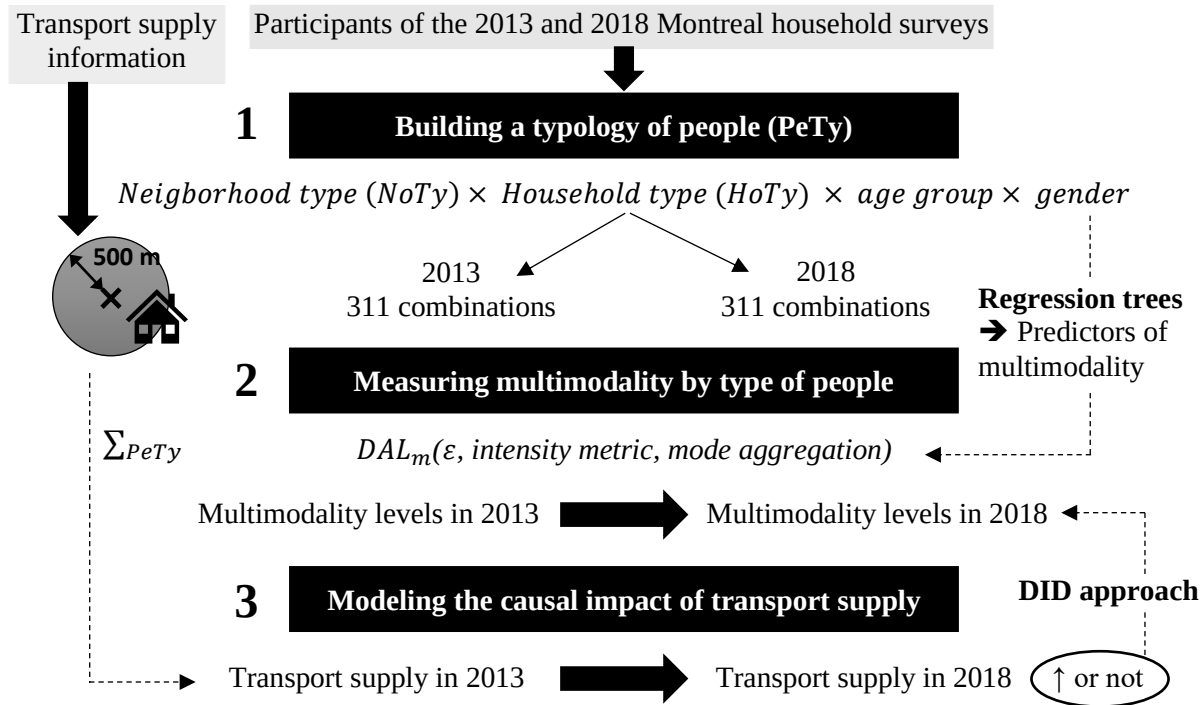


Figure 6.1 Methodological framework

Note: PeTy = type of people; NoTy = type of neighborhood; HoTy = type of household; DALm = variant of the Dalton index; DID = difference-in-differences.

6.3.1 Building a typology of people

The participants (respondents and other household members) of the two surveys mined in this paper were empirically clustered according to four attributes: neighborhood type of home location, household type, age group, and gender. Therefore, a neighborhood typology and a household typology (both based on variables defined at the household level) were first created. These two typologies are common to the 2013 and 2018 surveys for comparison purposes. For clarity, the following codes will be used for each typology: **PeTy** for type of people; **NoTy** for type of neighborhoods; and **HoTy** for type of households.

The segmentation method applied to build the neighborhood typology is inspired by a master's thesis (Martel-Poliquin, 2012). Eight variables were calculated within a 500-m radius around each household home location. The first three variables characterize land use, the five other variables describe transport supply for different modes:

- population density,
- employment density,
- commercial density,
- total length of the road network,
- density of intersections,
- total number of transit stop-times over 24 h,
- total number of docks in nearby bikesharing stations, and
- total number of cars in nearby carsharing stations.

The first three densities were evaluated with the survey data: the population density as the sum of the number of people living in nearby households; the employment density as the number of trip destinations for working activity; and the commercial density as the number of trip destinations for shopping activity. Road variables (total length and intersection density) were assessed from the Canadian census file, while transit stop-times were counted (regardless of transit mode) from GTFS files. As bikesharing and carsharing capacities (in number of docks and cars respectively) fluctuate over time, they were estimated as the maximum number of different bikes or cars available in stations over the study period. These variables were then normalized, and a K-means algorithm was applied. The number of groups was chosen thanks to a two-step method consisting of making 30 prior clusters with K-means, then drawing a dendrogram with the centers.

For the household typology, the method developed in a previous paper was used (Fabre & Morency, 2019). The households were segmented applying again a K-means algorithm based on the following attributes:

- average age of the household members,
- maximum age amplitude among the household members,
- number of household members,
- number of children 12 years old or less, and
- number of people between 76 and 84 years old.

Six NoTys and eight HoTys were established (see the results section). Moreover, five classes of ages ([0;15[, [15;25[, [25;40[, [40;65[, and 65 years and older) and two genders (male or female) were considered. Therefore, 480 combinations were possible for each survey (2013 and 2018 are treated separately to be compared afterward). However, combinations with fewer than 10 observations were removed, then only combinations common to both years were selected. This led to 311 different combinations (therefore PeTys) per year, which gather almost all participants in the two surveys (99.9% of the 282,583 individuals who made at least one trip during the observation day).

6.3.2 Defining an indicator of multimodality

The indicator chosen to measure multimodality is a variant of the Dalton index. In a previous paper (Diana & Pirra, 2016), several indices were theoretically compared and this one seemed to have interesting properties for assessing mode use diversity, especially because it is very sensitive to intensity differences. Additionally, its mathematical formula incorporates a parameter to adjust the sensitivity of the indicator to the lowest intensities: it makes it possible to better consider the least used modes. Therefore, this indicator is particularly adapted to situations in which some modes are used less than others, which is the case in Montreal where the car plays a predominant role. Note that other indicators (not presented in this paper) were tested by the authors, but this index was found to be the best.

6.3.2.1 Variant of the Dalton index

The Dalton index, which is very similar to the Atkinson index, is an inequality measure often used in economics to determine which end of the income distribution contributes most to the observed inequality. This indicator is transferable to a transportation context if applied to the distribution of intensities of use. Nevertheless, its variant DAL_m , proposed by Diana et Pirra (2016) and defined in Equation 1, may be even more useful since it considers the real number of modes used m rather than the number of modes studied n , with $f_1, \dots, f_n$ the corresponding intensities of use:

$$DAL_m = 1 - \frac{\frac{1}{n} \sum_{i=1}^m (f_i^{1-\varepsilon} - 1)}{\left(\frac{1}{m} \sum_{i=1}^m f_i\right)^{1-\varepsilon} - 1} \quad (1)$$

The indicator decreases with multimodality: the lower its value, therefore, the greater the level of multimodality. More specifically, it varies from 0 (equal use of all modes) to $(n - 1)/n$ (when only one mode is used). The parameter $\varepsilon < 1$, called the “inequality aversion parameter” in economics, represents the decreasing influence of the most used modes in determining the level of multimodality. This means that higher values of ε make the index more sensitive to the intensities of use of the least used modes. ε is determined empirically through sensitivity analyses.

6.3.2.2 Sensitivity analyses

DAL_m was applied by PeTy using aggregated (weighted) intensities f_i per mode, obtained by summing all the intensities of use of the individuals who belong to the same PeTy (each intensity being weighted by the corresponding individual weighting factor). We varied the parameter ε from 0.1 to 0.9 by increments of 0.1 and selected the value that gave the most homogeneous distribution of the indicator results to measure a greater variety of multimodality levels. Furthermore, in line with the two issues discussed before, we tested different measurement units of the intensities of use, as well as different mode aggregations. Four intensity metrics were investigated:

- number of trip segments (or travel stages),
- number of trips,
- travel distance (meters), and
- travel time (seconds).

The number of trips segments per mode was calculated from the declared mode sequences. Walking to get to or from another mode was not considered; this limitation will be discussed later. As for the number of trips, unimodal and multimodal trips were differentiated: unimodal trips were summed up per mode, but trips combining several modes were gathered in another independent category called “multimodal trips.” Moreover, the coordinates of the trip origins and destinations served to assign the trips on the network. Two routing tools were used to estimate the shortest-path itineraries: the Open Source Routing Machine (Luxen & Vetter, 2011) with customized routing profiles for trips made by car, walking, and cycling; and the trRouting transit routing algorithm (Bourbonnais, 2019). Travel network distance and time per trip were then estimated based on this assignment. This was possible only for unimodal trips (thus bimodal trips were neglected in the last two metrics). Additionally, the three following mode aggregations were defined:

- **AG1:** subway, bus, school bus, train, car driver, car passenger, taxi, walking, cycling, other mode;
- **AG2:** transit, car driver, car passenger, walking, cycling, other mode; and
- **AG3:** transit, car, active transportation, other mode.

The first option is the most disaggregated (almost all modes are separated). In the second option, transit modes were brought together, and taxi was placed in “other mode.” In the third option, car driver and car passenger were gathered in “car,” walking and cycling in “active transportation.” In this way, sensitivity analyses were carried out to define the most appropriate parameters (ε , measurement unit and mode aggregation), that is, in this paper, the parameters yielding the distribution closest to the normal law to allow the application of the model defined below.

6.3.3 Difference-in-differences approach for causal impact modeling

The indicator proposed above could be used as the dependent variable of a model, for instance a regression tree, to find significant predictors of multimodality. However, only correlations (not causality relationships) would be measured with this approach. Instead, we tried to explain the changes in multimodality levels between 2013 and 2018 by PeTy. More specifically, our goal was to understand the extent to which the evolution in travel behavior was a result of changes in transport supply. To this end, we used the difference-in-differences method. This approach of causal inference was recently applied in a transportation context to appraise the effect of emerging mobility services on traditional modes of transport (Becker et al., 2018; Campbell & Brakewood, 2017). In this paper, we analyze the effect of transport supply improvement rather than the appearance of a new service, on the multimodality of travel behavior rather than on the ridership of a specific mode of transport.

The difference-in-differences method allows us to assess the causal impact of a treatment, policy, or intervention on a variable of interest. It is used when observations or quasi-experimental data are available (i.e., when the treatment was not randomly assigned, which is the case for transportation interventions) and thus serves to reduce biases caused by the non-random nature of the experiment in estimating the treatment effect. The simplest case is when outcomes are observed for two groups (a treated and a control group) at two periods. The difference (in the average outcome) between the two periods for the control group is subtracted from the difference between

the two periods for the treated group. The difference of these two differences is the estimated effect of the treatment (δ). In this paper, the following effect is measured:

$$\delta = (E[I_{igt}|g = T, t = 2018] - E[I_{igt}|g = T, t = 2013]) - (E[I_{igt}|g = C, t = 2018] - E[I_{igt}|g = C, t = 2013]) \quad (2)$$

where I_{igt} is the indicator of multimodality of PeTy i that belongs to group g , measured at time t (2013 or 2018). Group membership is based on the evolution of transport supply between 2013 and 2018; an improvement was observed in the vicinity of the PeTys who belong to the intervention group (therefore the sum of a previously defined neighborhood variable, for instance transit stop-times, within 500 m of all home locations of each PeTy, increased), while transport supply stagnated or decreased in the comparison group. Equation 2 shows how the method removes biases resulting from permanent differences between the treated and control groups (such as inherent differences in the socio-demographic attributes), as well as biases caused by temporal trends having the same impact on both groups. However, in the absence of treatment, the variable of interest (here the multimodality level) is assumed to follow the same trend for both groups. This assumption will be discussed later.

The corresponding model, an ordinary least squares regression, is estimated as (Equation 3):

$$I_{it} = \alpha + \beta_0 Year_t + \beta_1 Group_i + \sum_k \gamma_k (Type_k)_i + \delta Group_i * Year_t * \Delta Supply_i + \varepsilon_{it} \quad (3)$$

The transport supply improvement effect ($Group_i * Year_t * \Delta Supply_i$) is the interaction between $Group_i$ (dummy variable for the group membership : $Group_i = 1$ if the aggregated transport supply variable has increased for PeTy i between 2013 and 2018, and $Group_i = 0$ otherwise) and $Year_t$ (dummy variable for the year: $Year_t = 1$ for 2018, $Year_t = 0$ for 2013) weighted by the augmentation of supply level $\Delta Supply_i$ (in number of transit stop-times, bikesharing docks or carsharing vehicles): $\Delta Supply_i = Supply_i(t = 2018) - Supply_i(t = 2013)$. Control variables ($Type_k$) were added to capture the effect of the socio-demographic and spatial features of the PeTy: NoTy, HoTy, age group, and gender.

6.4 Results

Results are reported in three parts. First, the typology of people is described. The multimodality measured by the chosen indicator is then analyzed by PeTy and correlated with the socio-demographic and spatial attributes of the PeTy. Finally, changes in multimodality between 2013 and 2018 are presented and the influence of transport supply improvement on the observed evolution is modeled for three variables: transit stop-times over 24 h, number of bikesharing docks, and number of carsharing vehicles.

6.4.1 Typology of people

All participants in the 2013 and 2018 Montreal regional surveys were segmented into combinations of the four following attributes: NoTy of home location, HoTy, age group, and gender. Six NoTys were inferred using the clustering algorithm previously described. Figure 6.2 spatializes households' home locations (from both 2013 and 2018 surveys) for each group. The proportions of households in each NoTy is also given in parentheses. Moreover, Figure 6.3 provides the centers of the groups and, therefore, the means of the normalized variables on which clustering was based.

NoTys 1, 2, and 3 gather households who live on the island of Montreal. They are characterized by high population, employment, and commercial densities within 500 m of the home locations. Households have access to all modes of transport, especially to the subway network and many bus lines (high number of stop-times over 24 h). Shared modes (bikesharing and station-based carsharing) are also reachable. Moreover, the high density of intersections makes these neighborhoods suitable for walking.

NoTys 4 and 5 are the largest (25% and 29% of the households). The corresponding neighborhoods are residential areas. The road network is particularly developed in NoTy 4. Most households are outside the catchment areas of the shared services, but many of them still have access to bus or train lines (the main axes stand out on the maps of Figure 6.2). Finally, the households in NoTy 6 live on the sidelines, with no (or little) access to alternatives to the car.

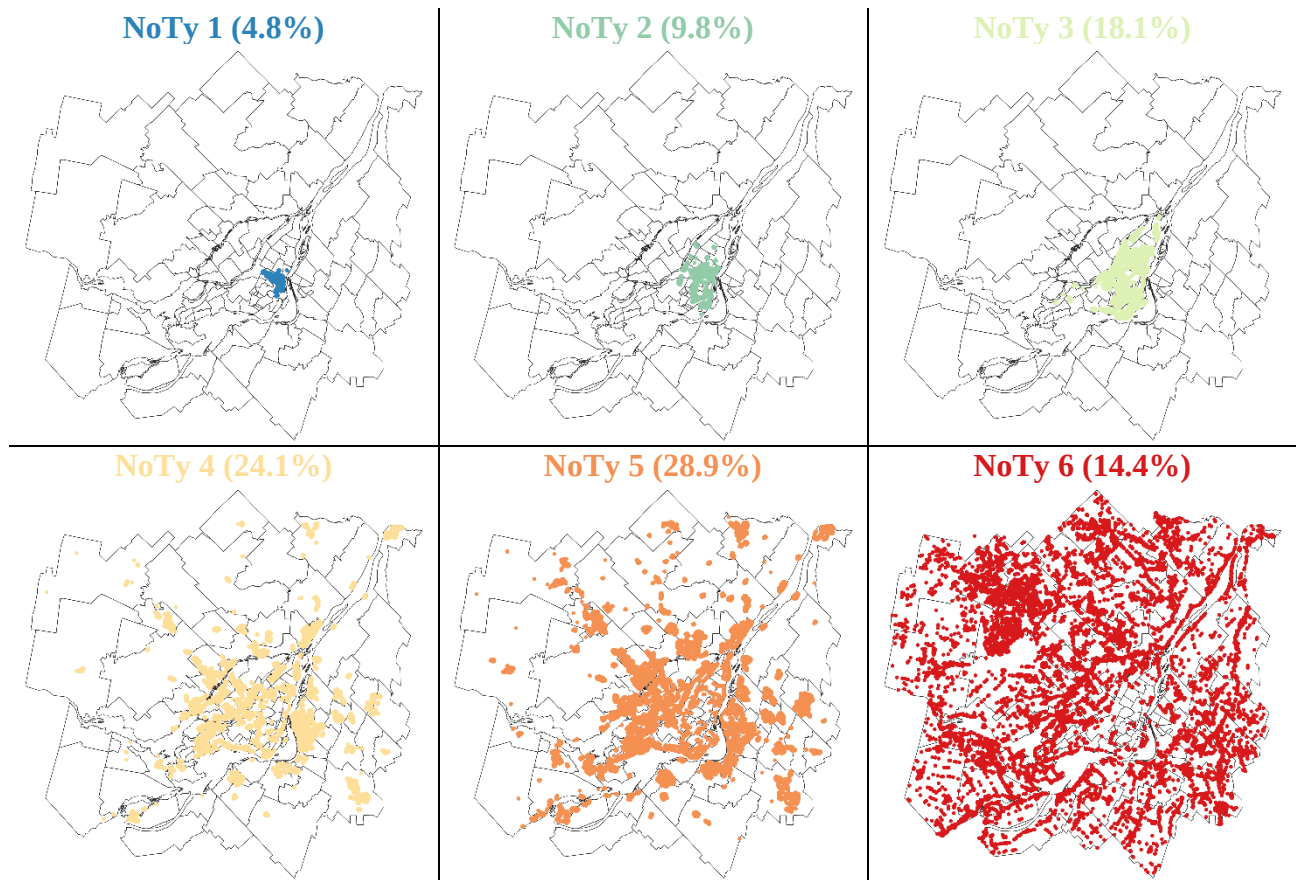


Figure 6.2 Typology of neighborhoods (proportion of households)

Note: NoTy = type of neighborhood.

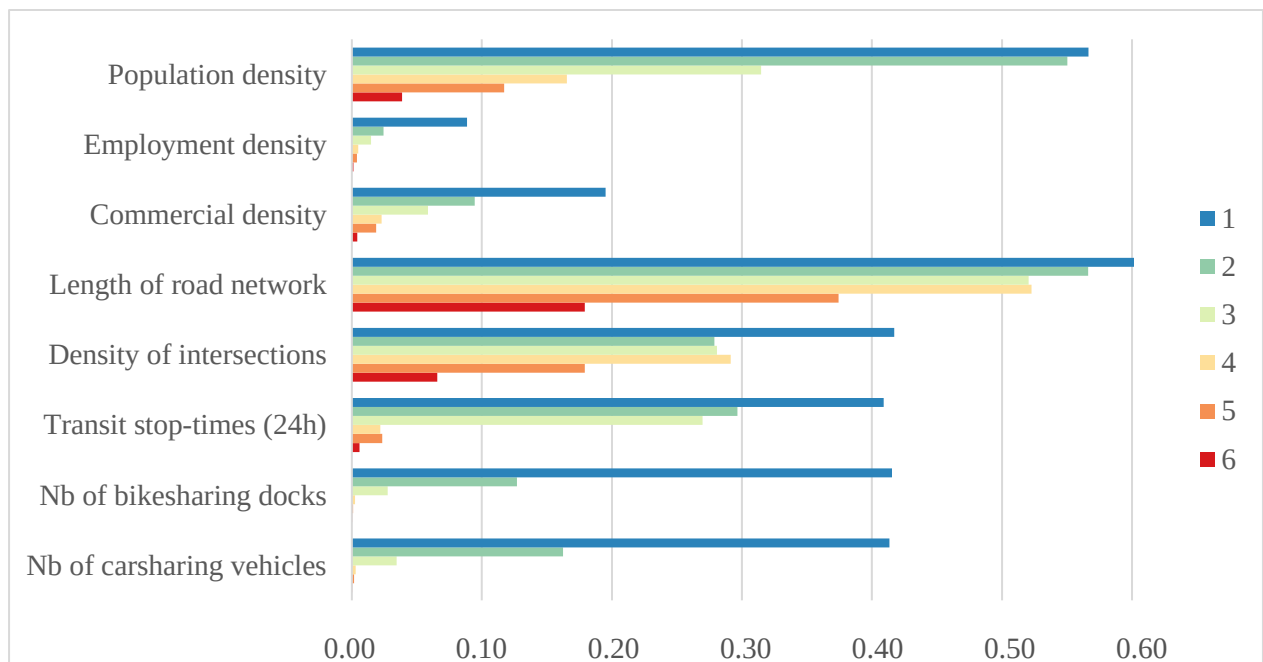


Figure 6.3 Typology of neighborhoods – Centers: normalized average variables by type

The results on the typology of households are reported in Figure 6.4 (representing the households according to the average age and the maximum age amplitude of their members, the color indicating the HoTy to which the household belongs) and Table 6.1 (centers, therefore, average household variables by HoTy). Each HoTy was given a meaningful name - the same as in a previous paper (Fabre & Morency, 2019). There are three HoTys of one-generation household (“young couples, roommates or singles”, “people in their fifties” and “retirees”); four types of two-generation household composed of parents and children (“very young families”, “young families” and “traditional families”) or “adults and seniors”; and one type of three-generation household (“multi-generational families”).

Based on these two typologies - two household (to which the individual belongs) attributes, and two individual attributes (age and gender) - the surveyed participants were manually placed into one of the 311 selected combinations per year.

6.4.2 Multimodality by type of people

The variant of the Dalton index defined in Equation 1 was calculated by PeTy using aggregated (weighted) intensities of use. Various measurement units and mode aggregations were tested, as well as different values for the parameter ε . The results of these sensitivity analyses are depicted in Figure 6.5 and Figure 6.6, in the form of histograms (in number of combinations or PeTys). The values obtained for the 622 combinations (311 combinations/year) are represented together.

Figure 6.5 shows that multimodality levels are more homogeneously distributed for the first option of mode aggregation (AG1), when transit modes are considered separately. The values tend to be divided into two groups for AG2, or they are concentrated on the left side for AG3. For the intensity metric, travel time (in green) leads to a wide range of multimodality levels, and the corresponding distribution is close to the normal law (assumption that will be required for the regression model developed afterward). Moreover, additional trial-and-error analyses (not presented in this paper) revealed that travel time better emphasizes the differences in multimodal travel behavior between 2013 and 2018. Therefore, the first mode aggregation and intensities in time were selected: DAL_m (time) AG1. Figure 6.6 gives the histograms produced by varying the parameter ε for this indicator (in fact, such analyses were carried out for each indicator of Figure 6.5 before drawing it). The values of ε from 0.5 to 0.7 returned the most homogeneous distributions; we, therefore, chose $\varepsilon =$

0.6. This fairly high value indicates that we needed to assign a higher weight to the least used modes to better highlight mode use diversity.

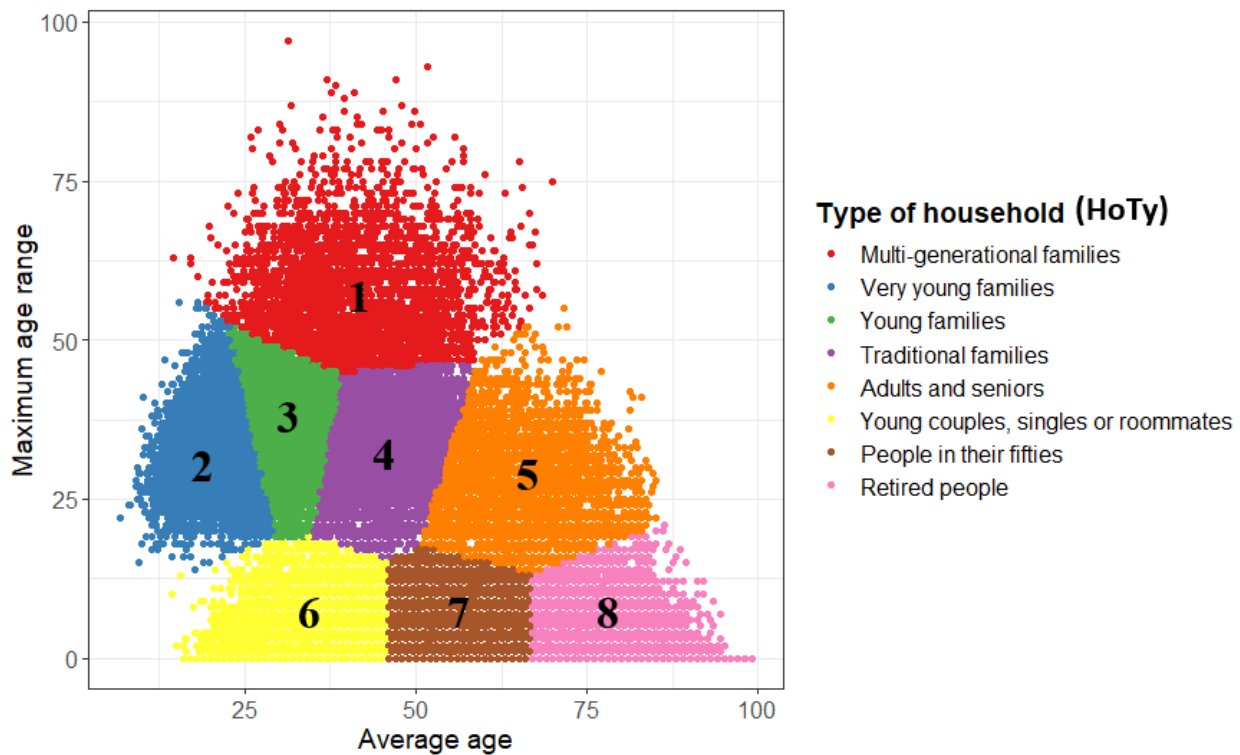


Figure 6.4 Typology of households (by maximum age range and average age)

Table 6.1 Typology of households – Centers: average variables by type

<i>HoTy</i>	<i>Size (in % of total households)</i>	<i># of people per household</i>	<i>Average age</i>	<i>Age range</i>	<i># of people < 12 years old</i>	<i># of people between 76 and 84</i>
1. Multi-generational families	2.2%	4.06	41.04	56.60	0.65	0.19
2. Very young families	13.1%	4.03	21.09	33.48	1.72	0.00
3. Young families	12.6%	3.67	31.31	35.74	0.41	0.00
4. Traditional families	8.3%	2.96	43.50	33.63	0.00	0.00
5. Adults and seniors	2.9%	2.49	64.85	27.37	0.00	0.40
6. Young couples, singles or roommates	10.9%	1.63	33.44	2.08	0.00	0.00
7. People in their fifties	28.2%	1.60	58.34	2.13	0.00	0.00
8. Retired people	21.9%	1.48	75.05	1.72	0.00	0.44

Note: HoTy = type of household. A different color is associated with each HoTy in the paper.

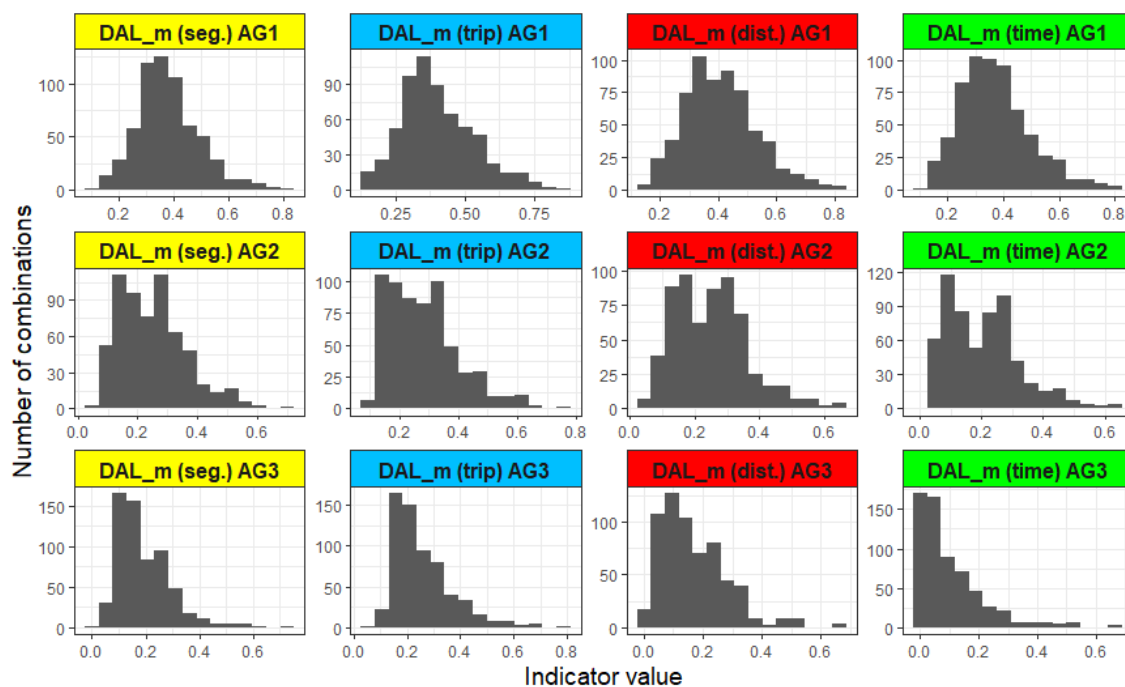


Figure 6.5 Histograms of the multimodality indicator calculated by type of people for different measurement units of the intensities of use (trip segments, trips, travel distance or time) and different mode aggregations (AG1, AG2 or AG3)

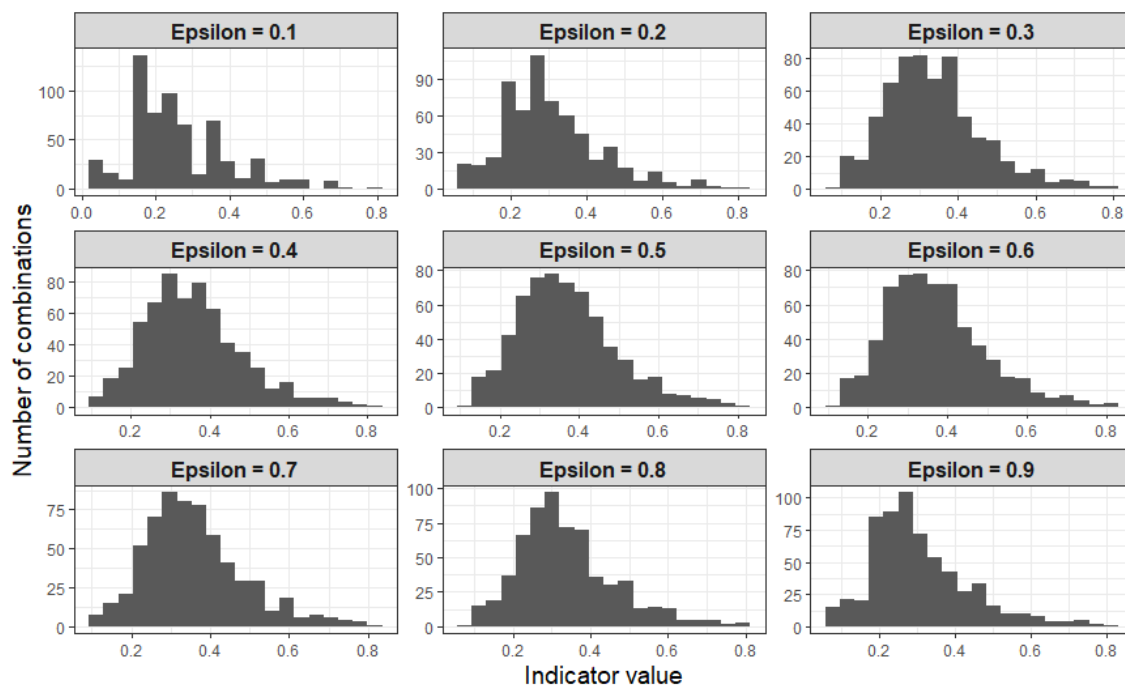
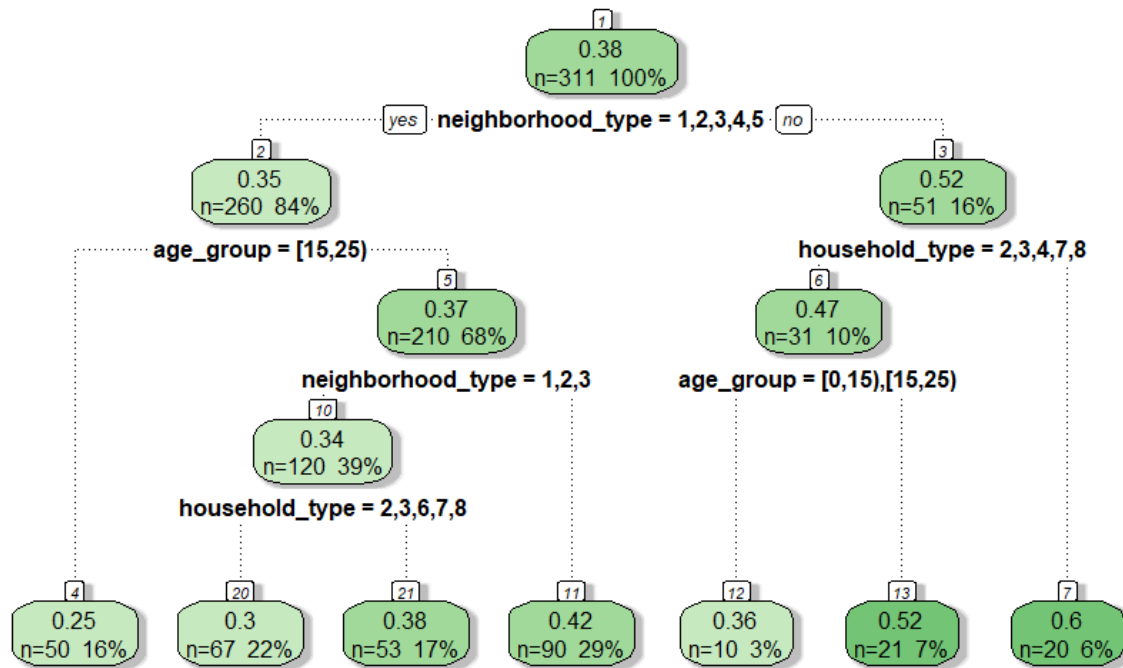


Figure 6.6 Histograms of the multimodality indicator calculated by type of people for different values of the parameter ε – Example for DAL_m (time) AG1

After having specified our multimodality indicator, the resulting values were correlated with the socio-demographic and spatial attributes of the PeTys. To this end, we developed two regression trees (one with the 311 combinations of 2013, the other with those of 2018) to predict the selected indicator, that is, DAL_m (time) AG1 with $\varepsilon = 0.6$. The predictor space of a regression tree is split into regions and the mean is applied to make a prediction in each region. However, in this paper we used regression trees to identify the best predictors of multimodality (selected in the splitting rules) rather than to make predictions. The potential predictors included in the model training were the four categorical features of the combinations, namely NoTy, HoTy, age group, and gender. The *rpart* package of the statistical software R was used, with the ANOVA method. The same stopping criterion (complexity parameter) was set for the two trees. Results are reported in Figure 6.7. The number and proportion of combinations as well as the average value of the indicator are provided at each node. The label of each split indicates the left-hand branch.

Similar trees were obtained for 2013 and 2018. NoTy is the most important predictor for both years (as revealed by the two main branches on top of the trees). PeTys in NoTy 6 demonstrate a higher average indicator value and, therefore, lower levels of multimodality. In 2013, the first three NoTys (where all alternative modes are available) stand out farther down the tree by even higher multimodality. Age group and HoTy are the second most important variables. Multimodality is more prevalent in very young, young families, or traditional families, especially for their youngest members. People between 15 and 24 years old have indeed a more diversified use of modes. Moreover, young couples, singles, and roommates, as well as people in their fifties and retirees, tend to be more multimodal. The multimodality of these three HoTys could be linked to a lack of constraints (families without children, work, or both). Such results are consistent with the literature. Furthermore, as NoTy is the most important predictor, we wanted to deepen the impact of transport supply on multimodal travel behavior.

(a) 311 types of people in 2013



(b) 311 types of people in 2018

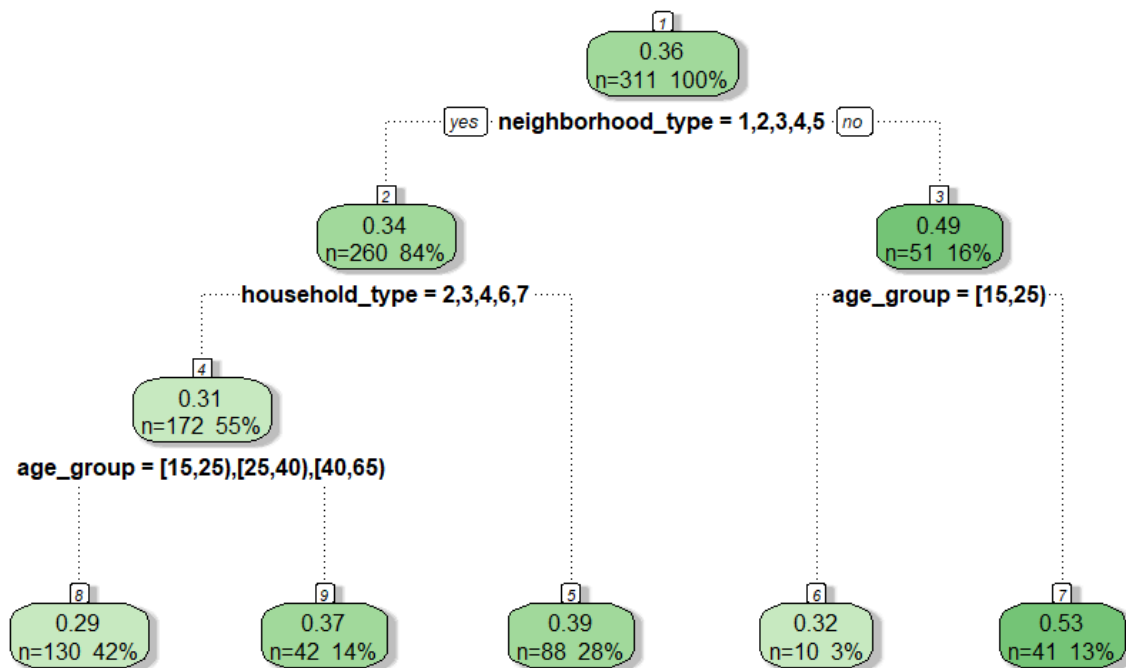


Figure 6.7 Regression trees for predicting the multimodality indicator DAL_m (time) AG1 $\varepsilon = 0.6$ calculated by type of people (a) in 2013, (b) in 2018

6.4.3 Multimodality evolution from 2013 to 2018: influence of transport supply

Comparing the same PeTys in 2013 and 2018, we found that, overall, multimodal travel behavior increased. The indicator value decreased (therefore mode use diversity increased) for 205 of the 311 combinations. More specifically, the average indicator was 0.38 per PeTy in 2013, against 0.36 in 2018, meaning an average reduction of 0.02. This difference between the two years is significant at 10% according to the Student's test. Moreover, this evolution in multimodality per PeTy was crossed with changes in transport supply aggregated variables (sum of all transport supply units in the vicinity of the individuals that belong to each PeTy): negative correlations were revealed. This means that when transport supply increased over the five-year period, the indicator value tended to decrease, thus multimodality increased.

The difference-in-differences approach allowed this relationship to be refined. As explained in the methodology, the causal impact of three alternative transport supply variables was investigated: transit stop-times over 24 h, number of bikesharing docks, and number of vehicles in carsharing stations. More specifically, we looked at whether these services had been improved over the five-year period, then we modeled the effect of this improvement on the chosen multimodality indicator by PeTy. This impact was weighted by the number of transport supply units that were added between 2013 and 2018. Moreover, the four categorical features of the PeTys were included as control variables (the category with the higher number of observations as the reference category). The three resulting regression models are detailed in Table 6.2. Note that all these models are pretty good: they explain about 50% of the variability of the multimodality indicator. The numbers of PeTys in the intervention and comparison groups are also provided at the bottom of the table.

The previous outcomes made with regression trees are confirmed with these models. Young people between 15 and 24 years old tend to have a more homogeneous use of the modes than those between 40 and 64 years old (reference category), while people over 65 are inversely more unimodal. Compared with traditional families, the travel behavior of very young, young families, and people in their fifties is more diversified, whereas multi-generational families as well as families of adults and seniors are less multimodal. The residents of all NoTys, except NoTy 6, are more multimodal than those living in NoTy 5. Gender does not seem to have a significant effect on mode use diversity.

Table 6.2 Results of difference-in-differences regression with treatment based on the evolution of transport supply between 2013 and 2018

	Transit stop-times (24h)	# of bikesharing docks	# of vehicles in carsharing stations
(Intercept)	4.00E-01 ***	4.26E-01 ***	4.15E-01 ***
Year = 2018	-1.14E-02	-9.31E-03	-1.33E-02 .
Group = augmentation of the transport supply	9.02E-03	-2.68E-02 **	-1.13E-02
Neighborhood type = 1	-2.15E-02	-1.28E-02	-2.34E-02 .
Neighborhood type = 2	-5.85E-02 ***	-5.77E-02 ***	-6.45E-02 ***
Neighborhood type = 3	-8.26E-02 ***	-8.15E-02 ***	-9.04E-02 ***
Neighborhood type = 4	-2.82E-02 .	-4.53E-02 ***	-3.40E-02 **
Neighborhood type = 6	1.22E-01 ***	1.15E-01 ***	1.15E-01 ***
Household type = Multi-generational families	3.77E-02 **	3.61E-02 **	3.63E-02 **
Household type = Very young families	-3.32E-02 *	-3.43E-02 **	-3.56E-02 **
Household type = Young families	-5.07E-02 ***	-4.40E-02 ***	-5.01E-02 ***
Household type = Adults and seniors	4.79E-02 **	4.24E-02 **	4.24E-02 **
Household type = Young couples, singles or roommates	-1.61E-02	-5.57E-03	-1.06E-02
Household type = People in their fifties	-4.97E-02 **	-4.67E-02 **	-5.50E-02 ***
Household type = Retired people	1.81E-02	2.05E-02	1.36E-02
Age group = [0;15[	1.92E-02	2.04E-02	2.11E-02
Age group = [15;25[	-1.01E-01 ***	-1.07E-01 ***	-1.02E-01 ***
Age group = [25;40[	4.77E-03	-2.09E-03	4.15E-03
Age group = 65 years and older	2.63E-02 *	2.62E-02 *	2.91E-02 *
Gender = Men	9.41E-03	9.00E-03	8.65E-03
Transport supply improvement effect = Year x Group x Δ Supply	-2.85E-07 **	-2.91E-06 ***	-4.29E-05 **
Adjusted R ²	0.502	0.511	0.501
Number of types of people in (Group = 1)	123	216	152
Number of types of people in (Group = 0)	188	95	159

Note: Significance codes: 0; '****' 0.001; '***' 0.01; '**' 0.05; '.' 0.1; ' ' 1. White and gray colors are alternated in the table to differentiate the variables. The effect of transport supply improvement (of most interest in this paper) is in blue.

However, the impact which is of most interest in this paper is the transport supply improvement effect (in blue in the table). It is found significant for the three modes. The estimated coefficients can be interpreted as follows: the addition of 10,000 transit stop-times in a 500-m radius around the people of the PeTy is associated with a reduction of -2.85e-03 of the multimodality indicator, which is equivalent to about 20% of the average decrease observed between 2013 and 2018. Similarly, the addition of 1,000 bikesharing docks (respectively 100 carsharing vehicles) could

explain the reduction of about 20% (respectively 25%) of the average decrease between 2013 and 2018. As these results refer to aggregated values, the numbers of transport supply units to add are not really meaningful, but the direction of the observed impacts is desirable: transport supply improvement led to a significant decrease in the indicator and, therefore, to a significant increase in multimodality. Moreover, carsharing is the mode that has the most important effect on multimodal travel behavior.

6.5 Conclusion

Based on two cross-sectional (2013 and 2018) regional surveys from Montreal, the proposed methodological set up allowed multimodality to be measured with daily survey diaries by gathering individuals into PeTys and summing their intensities of use of several modes across all observation days. Moreover, it was possible to assess changes in multimodal travel behavior and to make causal inference without panel data by comparing the same PeTys in two distinct years. This typology of people was built empirically from four attributes: NoTy (defined by land use and transport supply variables within a 500-m radius of the home location), HoTy (defined by the household composition), age group, and gender. This led to the analysis of 311 combinations. Multimodality was then appraised thanks to a variant of the Dalton index; it was applied to a series of aggregated (weighted) intensities of use by PeTy. This indicator was parametrized to better consider the least used modes, that is, modes other than car. Sensitivity of the indicator was also checked to determine the best intensity metric: travel time was selected. Moreover, several aggregations of modes were tested: transit modes were eventually distinguished. The multimodality levels were then correlated with the socio-demographic and spatial attributes of the PeTys through regression trees. Finally, the decrease between 2013 and 2018 in the modal variability indicator by PeTy was explained by the evolution of three alternative mobility services over the five-year period: transit, bikesharing, and station-based carsharing. More specifically, a difference-in-differences approach was used to estimate the causal impact of aggregated transport supply improvement in the vicinity of home locations on multimodality.

Therefore, in the context where multimodality analysis is of great importance, this paper provides tools to help transportation practitioners foster more diversified travel behavior. Indicators of mode use variability such as the Dalton index could be calculated to set targets in multimodal mobility strategies, and then to evaluate the extent to which the measures put in place have met these

objectives. Determining the predictors of multimodality is also useful for identifying travelers more receptive to using a variety of modes; they could be more open-minded to new services according to the literature about “modality styles.” In this line, our results are consistent with previous works: the most multimodal people are mainly young people, or people with fewer constraints, such as households without children or the autonomous elderly. However, the most important predictor is the accessibility to transportation services: most multimodal people live in well-covered areas. Furthermore, causal inference is particularly interesting for planners to isolate the causal effect of a strategy (implantation of a new service or service improvement) on travel behavior. It could allow the location of supply units, the quantity of supply units, or both, to be found, helping to achieve the followed objectives. According to our models, the addition of 1,000 transit stop-times (respectively bikesharing docks, carsharing vehicles) results in a significant decrease of $2.85\text{E-}04$ (respectively $2.91\text{E-}03$, $4.29\text{E-}02$) in the chosen indicator (decreasing with multimodality). Such results should be interpreted in the light of set targets. In this way, this paper could further assist policy makers, stakeholders, and planning partners in future transportation investment decisions.

Nevertheless, this work has some limitations. First, walking, when combined with other modes, is neglected in the calculation of the intensities of use. Travel time and distance on foot could be estimated for transit trips using a routing tool, but the information on access and egress travel in the surveys used is not enough for other modes. The proposed method depends on multiple parameters and other sensitivity analyses should be done, especially for the 500-m radius that delineates the neighborhood of each household. Moreover, the assumption of the difference-in-differences method is not totally true: the compositions of the intervention and comparison groups may not be constant over time, and some events could have different impacts on the two groups. Their trends may, therefore, differ. However, this could be compensated for by applying a method based on propensity score (Stuart et al., 2014). Future work will focus on overcoming the mentioned limitations. Moreover, a multiday survey (over at least one week) is envisaged to study multimodality at the individual level without having to create an empirical typology of people. Causal inference could also be analyzed at a more systemic level, by spatial zone rather than by PeTy, using passive data.

Acknowledgments

The authors wish to thank the Ministry of Transport of Quebec (MTQ) and the Regional Metropolitan Transport Authority (ARTM) for giving access to the Montreal household surveys, as well as the companies (BIXI, Communauto) who provided supply information. The authors also acknowledge the financial support of the Canada Research Chairs (CRC), Quebec Research Funds (FRQNT), and Institute for Data Valorization (IVADO) programs.

CHAPITRE 7 ARTICLE 4 : MODELING INTERACTIONS BETWEEN THE MONTREAL SUBWAY AND OTHER URBAN TRANSPORTATION MODES AT THE STATION LEVEL

Elodie Deschaintres, Catherine Morency, Martin Trépanier

Accepté pour publication dans *Tunnelling and Underground Space Technology*

Date de soumission initiale : 17 janvier 2022

Disponible en ligne depuis le 23 mars 2023

Volume 136, 105079, DOI : 10.1016/j.tust.2023.105079

Abstract

Subway systems are an example of how underground space can contribute to developing more sustainable cities. Their use can be encouraged by integrating other urban transportation modes to provide easy access to the stations, thus facilitating intermodal connectivity. However, few studies have addressed data and modeling issues with multiple modes. This gap, mainly due to the lack of appropriate methods, explains why relationships between traditional and emerging modes remain poorly understood; there is still debate about whether they are complementary or competitive. This paper investigates this question by examining interactions between three alternative modes (bikesharing, free-floating carsharing and taxi) and the Montreal (Quebec, Canada) subway over two months. To this end, several operational passive datasets are merged. A two-level growth model, including both fixed and random effects, is then developed based on time series of daily numbers of boardings at subway stations. This longitudinal model quantifies the effect of various variables (calendar, weather, socio-demographic and economic, network connectivity, land use and built environment attributes) on each station use. The correlations between subway boardings and alternative mode destinations are also measured to detect potential intermodal relations. Finally, a further categorical analysis is conducted by period and station type to specify these interactions in time and space. Results reveal that alternative modes tend to complement the subway in the network periphery (as possible access mode to the subway stations), but may compete with the subway in areas closer to the city center. Such findings could help policymakers determine the

location of the mobility services to enhance first-mile subway access, as well as the best mode integration around subway stations in planning strategies.

Keywords: subway systems; transit accessibility; intermodal relations; growth model; mixed-effects; passive data streams

7.1 Introduction

Urban underground space is exploited to accommodate several functions in cities, based on mixed-use (commercial, cultural, recreational, etc.) facilities, as well as utility and transportation infrastructure (Cui et al., 2021; Li, X. et al., 2016). Most common forms of underground space use for transportation purposes include underground car parking spaces, road tunnels, subway or other railway systems, freight transportation systems and pedestrian walkways (Cui & Nelson, 2019). The case study of this paper, namely the city of Montreal (Quebec, Canada), is particularly known for having a multilevel urban structure in the city center. The underground level, called the Indoor City or Underground City, is one the most extensive indoor urban system in the world, and its historical growth, as well as the factors which caused its rapid development, have been thoroughly explored (El-Geneidy et al., 2011; Sijpkens & Brown, 1997). The 32-kilometer-long pedestrian system of this hidden city connects several commercial and cultural buildings to seven stations of the subway network. This generates many underground walk trips that were studied to better understand individual itineraries in underground space (Zacharias, 2015). However, the large majority of the trips made to reach subway stations comes from outdoor, and walking can be combined with other transportation modes to access the subway. Therefore, interactions between the subway and other urban transportation services must be investigated to better evaluate the potential of urban underground space for transportation purposes.

North American cities have long been developed according to a car-oriented urban sprawl model. However, environmental problems such as traffic congestion, noise and air pollution gave rise to the desire for more sustainable urban living and mobility. In this line, underground spaces play an important role since they allow increasing density and developing more compact cities, as well as implementing subway networks (Broere, 2016; Cui & Lin, 2016; Qiao et al., 2019). The subway is often considered as the backbone of the urban transportation systems. Indeed, the high speed and capacity of this mode enable efficient movement of large population flows. Depending on the jobs-housing mismatch and on the spatial distribution of the other points of interest, many trips may rely on this mode to commute and for other activities, especially in Asian cities (Yang, X. et al., 2017). In the Montreal region, about 12% of the trips during morning peak hours are made using the subway (almost 70% of them for work purpose) according to the 2018 Origin-Destination household survey (Secr tariat   l'enqu te Origine-Destination, 2020). However, underground

development is costly and laborious due to the use of advanced excavation and dewatering techniques (Cao et al., 2018; Zeng et al., 2021), and the need for expert studies to monitor structure deformations (Liu et al., 2022; Ni et al., 2021) and ensure construction safety (Zhou, Z. et al., 2021). These difficulties often lead to a limited coverage of the subway network over the whole city and of the suburban periphery. Consequently, subway accessibility is an important issue, and is highly impacted by the station locations. Indeed, the use of stations (in magnitude and pattern) depend on neighborhood characteristics, including land use and built environment variables (An et al., 2019; Jun et al., 2015).

Recent research about urban underground space utilization emphasized that while densification could significantly improve subway accessibility in the long term, integrating other transportation modes to provide easy access to subway stations would be a good short-term strategy (Lin, D. et al., 2022). This is called intermodality (combination of several modes during a single trip) in the transportation jargon. Buses have long been employed to bridge the connectivity gaps of subway systems (Wu et al., 2019), but new urban modes, such as shared mobility services, have emerged in the past few years and could also be leveraged. The interactions between such new modes and the subway have received little attention. They could also depend on the station location (hence, analyses at the station level are required): specific facilities (e.g., bikesharing docks) must be present nearby the station, and the trip must be long enough to make a transfer between the mode and the subway worthwhile (i.e., interesting in terms of total travel time). Nevertheless, instead of improving accessibility to the stations, and thus completing the subway, the alternative modes may also be in competition with the subway (i.e., substitute transit trips) if they are more rapid or more suited to certain needs. Therefore, there is a need to better understand the conditions under which intermodal relations between the subway and other modes may exist to better integrate all mobility services in urban development.

In the same vein, the inclusion of all urban transportation modes in multimodal planning strategies was presented as a condition to tackle car dependency and thereby move towards a more sustainable mobility (Spickermann et al., 2014). However, historically, most of the transportation demand forecasting models and indicators have been fitted specifically for one or at best two transport mode(s) (Morency et al., 2018), and based on a typical weekday or even peak hours (Verreault & Morency, 2011). This led to a poor understanding of the interactions between modes over time, as well as of the transportation demand variability. Therefore, new tools have been

required to perform both multimodal (including several modes) and longitudinal (over continuous and long period of time) or at least multiday analyzes. Attempts to study travel demand across days exist, but there is still room for improvement. Such investigations imply the availability and fusion of multiple databases to encompass a wide variety of transport services. In this regard, the emerging data landscape opens new opportunities (Cherchi & Bhat, 2018).

Against this background, this paper aims to develop a longitudinal model to explain the daily ridership of the Montreal subway at the station level. The location and the local attributes (including land use and built environment variables) of each station are taken into account. Moreover, the model includes the use of three alternative modes of transport (bikesharing, free-floating carsharing and taxi) in order to detect usage correlations with the subway. To this end, two months of subway tap-in validations are mined and combined with other data sources based on a common spatiotemporal unit. A multilevel growth model is then developed, considering both temporal and spatial features of the station neighborhood. Overall, this paper presents a systemic overview of the interactions between a subway system and urban development at the station level (i.e., interactions between the subway use and the station neighborhood features, as well as interactions with other urban transportation modes). The outcome provides methodological and analytical insight to better understand variability (inter-stations and intra-stations) in subway ridership, factors which affect station use, as well as multimodal travel behaviors, especially intermodal relations between subway and alternative urban transportation modes.

This paper is organized as follows. Through a literature review, multimodal and longitudinal investigations in transportation research are documented and data issues are discussed. The multiple sources of data processed in this paper are then described and the proposed growth model is explained. Results are exposed in two parts: a general model is first presented to assess the overall daily subway use, then this use is appraised by time period and station type. Finally, this paper concludes with practical implications and research outlook.

7.2 Background

7.2.1 Multimodal analysis in transportation research

Over the past few years, the rapid development of information and communication technology led to a diversification of transport services. For instance, shared modes such as bikesharing, carsharing

and ridesourcing programs (e.g., Uber or Lyft) enhanced classic transport supply. This gave rise to investigations about the impacts of these emerging services on more traditional transport modes, such as car or transit. While many authors agree that the introduction of alternative modes usually reduces car use (Clewlow, 2016; Fishman et al., 2014) and car ownership (Becker et al., 2018), conclusions are more mixed regarding their influence on transit use: both increases and decreases are reported, even in the same study when including several cities (Hall et al., 2018; Shaheen et al., 2013). Therefore, in a context of declining transit ridership in some North American cities, questions about the complementarity (or competitiveness) of these modes are still raised (Feigon & Murphy, 2018; Graehler Jr et al., 2019).

The existing research about modal shift dynamics and substitution rates revealed that carsharing and bikesharing services may partially substitute transit (Fuller et al., 2013; Martin & Shaheen, 2011), but this trend depends on several factors. The type of carsharing service is one of them: free-floating carsharing is more likely an alternative to transit than station-based carsharing because of the flexibility of the service (Becker et al., 2017). However, carsharing could also be used as a complement to transit for trips that are less appropriate with transit (e.g., shopping, bad weather, etc.), at times or places where the transit network is less developed, and as a means to access/egress to/from transit nodes (Ceccato & Diana, 2021). Likewise, bikesharing may replace some short transit trips because cycling may be faster, especially in dense areas where the network is saturated, but it could also be used to connect to transit (Shaheen et al., 2013). In this way, bikesharing was found to increase rail use but reduce bus use (Graehler Jr et al., 2019): the authors assumed bikesharing plays a role in the first/last kilometer to reach the subway, but it competes with bus because of the shorter trips made with this mode. Moreover, the effect of bikesharing on transit ridership may vary in space: substitution seems more prevalent in the urban core whereas bikesharing complements transit in peripheral neighborhoods (Ma, T. & Knaap, 2019; Martin & Shaheen, 2014). In line with these previous studies, this paper aims to provide a deep dive into the role of free-floating carsharing and bikesharing as possible means of connecting to subway stations. Taxi will also be included in the analysis because it was observed that this mode may be used to connect to the subway too (Li, M. et al., 2017).

Overall, however, the interactions between several modes of transport have received little attention in the literature, especially at the systemic level. Previous research focused on analyzing multimodality at the individual level, defined as the use of different modes of transport over a given

period of time (Buehler & Hamre, 2015; Kuhnimhof et al., 2012; Nobis, 2007), or at the trip level, when several modes of transport are combined in the same trip (Kagerbauer et al., 2015; Oostendorp & Gebhardt, 2018). The latter is also called intermodality. Most of the identified papers were based on qualitative methods and survey data, hence they were subject to respondent bias, and the available data was often cross-sectional and limited to a short period while analyzing multimodality and variability in mode choice at the individual level requires multiweek (or at least multiday) information (Deschaintres et al., 2021; Schlich & Axhausen, 2003). Moreover, only classic modes of transport (i.e., personal car, transit, personal bike, and walking) were generally examined because emerging modes are not well represented in surveys. More importantly, while they provide useful contextual information, survey-based analyses lack spatial and temporal details, and they have sampling issues: the collected information may not be representative of the entire population and the frequency of each type of trip (substituting another mode or not) is often not recorded (or partially with retrospective stated frequencies). Therefore, they may not reflect the real impact on transportation demand at the systemic level. Studies at the systemic level are sparse because of data unavailability: few works benefit from enough operational data to analyze simultaneously different transport services. In this way, only two modes are generally considered in each paper. Furthermore, passive datasets are independent (each mode of transport is described by a different database because the systems are operated, and data collection is performed, by different companies). Therefore, it may be difficult to measure the relationships between modes, but correlations can be measured in time and space. A recent clustering analysis provided an example of combining several passive data streams to create typical and multimodal daily patterns of use, but the authors underlined the need to consider the spatial dimension (Morency et al., 2018). This was then done with the concept of MIAZ (mobility interaction analysis zones), but it is still necessary to model the interactions rather than to calculate descriptive statistics (Morency et al., 2020).

Going beyond these limitations, this paper deals with intermodal relations at the systemic level by examining how the destinations in three modes (bikesharing, free-floating carsharing and taxi) independently affect the subway entries. A conceptual framework will be proposed to model these interactions rather than simple qualitative analyses. Moreover, the impacts will be modulated by the station location, including neighborhood features, and for different periods, to examine the variability of the interactions over time and space. To this end, instead of handling cross-sectional

one-week/one-day travel diaries or retrospective stated frequencies as in previous studies, multimodality will be studied over two months using longitudinal passive data.

7.2.2 Longitudinal modeling of transport demand

In addition to multimodality, the longitudinal scope of mobility is another key point to cover when modeling transport demand. Indeed, the variability of travel behaviors has been amply demonstrated (Schlich & Axhausen, 2003; Verreault & Morency, 2011), especially for transit (Deschaintres et al., 2019). This highlights that the notion of “typical weekday” is outdated. For this reason, many studies addressed the need to model transit ridership while considering short and long-term variations. Some authors called for developing dynamic models of activity-travel demand across multiple time horizons (Rasouli & Timmermans, 2014). Furthermore, time series forecasting is an evolving field of research. Several types of models can be used to model longitudinal data (in the form of repeated measurements): from generalized regressions (possibly including an autoregressive error parameter to correct for temporal autocorrelation), which are largely the most commonly used models in transportation, to more advanced statistical methods such as autoregressive integrated moving average (ARIMA) models (Chiang et al., 2011; Lepage & Morency, 2021) or seasonal-trend loess-based (STL) decomposition (Qin et al., 2019). Moreover, the effectiveness of machine learning methods, for instance Random Forests (RF) or Long-Short Term Memory (LSTM) neural networks, to forecast transport passenger flows has been proved (Toque et al., 2017).

Multilevel models are best known for their use in the field of education, to compare test scores between students from different classes of several schools. However, they can also be applied for longitudinal analyses: they are then often called “growth models”. Such models are mainly used in sociology and psychology to study changes in individual behaviors (Singer & Willett, 2003). Their use in transportation is much less common, but relevant to consider the hierarchical structure of the services (for instance, the decomposition into different stops or stations, routes and then agencies for transit). Unlike usual linear models, multilevel models do not assume that the observations are independent: autocorrelation (within stations, routes, or agencies) in longitudinal data is considered. They also account for the differences in ridership between the stations/routes/agencies even if the variables that explain these differences are not all controlled for. Therefore, such an approach allows inference at the right level and, compared to ordinary least

squares regression, it offers greater flexibility and reduces the omitted variable bias (Huang, F. L., 2018). A few multilevel models were employed in a transportation context to determine factors affecting station-based carsharing (De Lorimier & El-Geneidy, 2013) or bikesharing (El-Assi et al., 2017) usage per station. To better understand the stagnation, if not the decline, of the transit ridership observed in North American cities, some authors also used multilevel models to explore the determinants of transit ridership over several years, at the city (Boisjoly et al., 2018) or route (Diab, DeWeese, et al., 2020) level.

However, these authors underlined the need to better assess the effects of the new mobility services with more detailed data. In this line, Campbell et Brakewood (2017) investigated the impact of the presence of bikesharing infrastructure at the bus-route level in New York City using a fixed-effect model. Our model will go a step further by quantifying the influence of the use (not only the presence) of three modes on transit ridership at the subway station level, and more flexibility will be given by adding random effects and explanatory variables, including land use information. Moreover, the daily scale will be investigated instead of the monthly or annual ones which were often used in the above-mentioned models. Regarding methodology, an original formulation of the growth model will also be proposed: variations will be studied between both spatial and temporal aggregations, and heteroscedasticity as well as correlation of the residuals will be considered in the model.

7.3 Data and methodology

As previously established, “multimodality” and “longitudinality” are two important concepts to account for to anticipate the evolution of mobility. However, they are rarely coupled: the previous longitudinal studies were carried out for one specific mode of transport while interactions between modes were often estimated from cross-sectional survey data. Therefore, the main originality of this paper lies in including the influence of several alternative modes on the longitudinal variations of subway ridership at the station level. As there is no single database describing the continuous use of all modes of transport at the same time, several passive data streams must be combined. Data fusion methods imply many challenges related, among other things, to the heterogeneity of data formats and usage units, and transportation-related examples are rare (Bayart et al., 2009; Cherchi & Bhat, 2018; Deschaintres et al., 2022). Nevertheless, this paper offers a method to allow the integration of various data sources in a single model.

7.3.1 Multiple-source data fusion

The case study of this paper is the subway network of Montreal (Quebec, Canada). This network, illustrated in Figure 7.1, is operated by STM (*Société de Transport de Montréal*), the transit authority, and is composed of 4 lines and 68 stations. However, only 67 stations will be modeled in the following: one of them (the “Jean-Drapeau” station, crossed out on the map) was removed because too difficult to predict due to its important particularities. This station is indeed located on a small uninhabited island (hence without socio-demographic features), within a park and its usage is highly dependent on events. The validations made in the subway network are collected by an automatic fare collection system called the OPUS system. This system gathers tap-in validations of smart cards and tickets (at boarding only). Bus validations are not considered because bus stops are not yet geolocated by the OPUS system. Therefore, the main dataset used in this paper is a sample of boarding validations in 67 subway stations. It spans over two months, from September 1st, 2016 to October 31st, 2016. The period is relatively short, but sufficient to analyze daily interactions; this paper is not intended to provide long-term forecasting of the subway ridership.

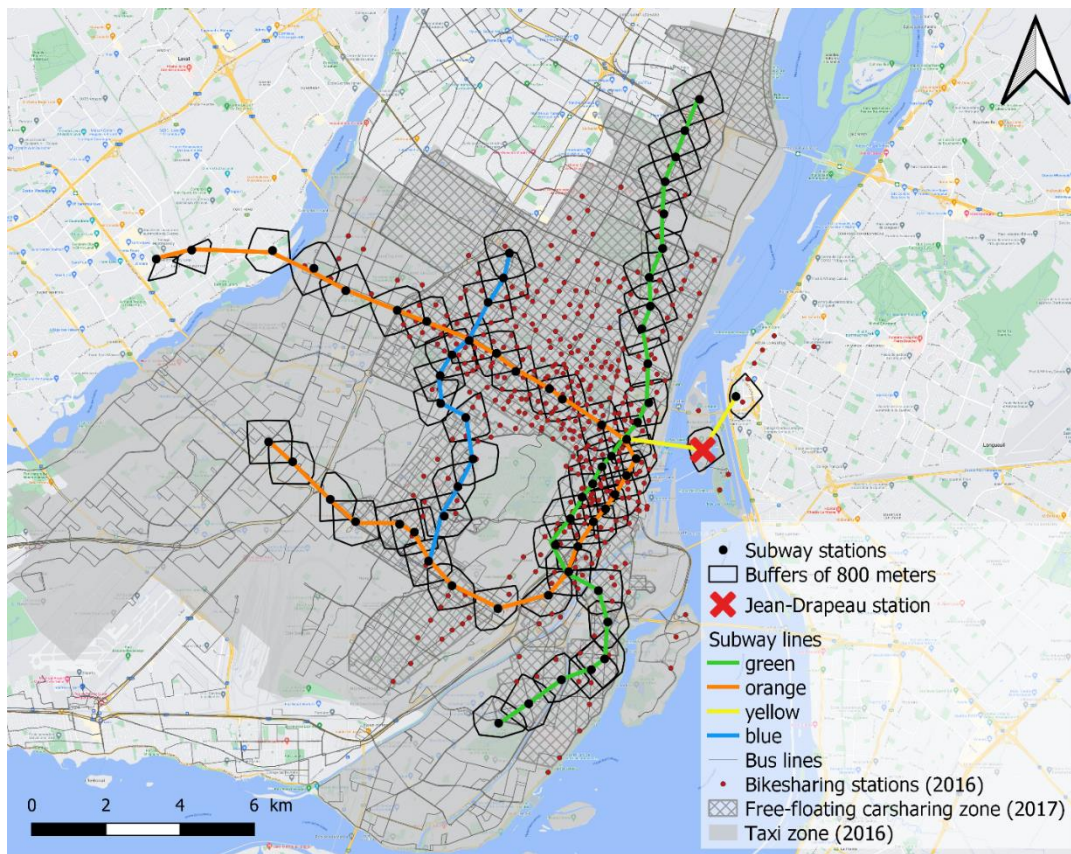


Figure 7.1 Island of Montreal with transit, bikesharing, taxi and free-floating carsharing services

Additional datasets were mined to have complementary information. Other passive data was processed to evaluate the use of bikesharing (BIXI), free-floating carsharing (combining usage of two different operators) and taxi (Taxi Diamond) within 800 meters (walking network distance) around each station. These data streams are transactional data which were collected by the private companies operating these modes. The samples used in this paper cover the same time period as the subway, i.e., from September 1st, 2016, to October 31st, 2016. The raw data contain the timestamp and coordinates of the origin and destination of each transaction. Moreover, the 2013 Montreal regional household survey and 2016 Canadian census were exploited to characterize the neighborhood of each subway station with socio-demographic features. GTFS files (bus stops) and information about bikesharing services (stations and docks) were also leveraged to describe the presence of feeder modes near the stations. Weather information was compiled from governmental open data (Environment Canada), whereas land use and built environment data were gathered from geo-referenced data of the Observatory of Montreal (CMM) and CanMap® GIS data (DMTI Spatial). In addition, STM provided an inventory of major social events that took place near each subway station between 2015 and 2017. They are mainly shows, exhibition festivals, concerts, and sporting events (hockey, soccer, football, tennis, or other).

Finally, all these datasets were combined based on a common spatial and temporal unit: the station neighborhood (defined by a walking buffer of 800 meters) on a specific day of the study period. This distance was chosen because commonly used in the literature to set the maximum access distance to subway by foot (El-Geneidy et al., 2010). Therefore, all datasets were aggregated (by summing or averaging the variables) per day and/or per subway station area, and all information was then linked thanks to the model definition.

7.3.2 Growth modeling

The growth model developed in this paper is part of the larger family of generalized multilevel (or linear mixed) models, based on data with a hierarchical structure. The chosen framework is composed of two levels, with repeated measurements nested in spatial units (subway station areas). Consequently, both space and time dimensions are considered in the model.

7.3.2.1 Two-level hierarchical structure

The hierarchical structure of our model is depicted in Figure 7.2. Time series of 61 observations (one observation per day) are nested in each of the 67 subway stations studied. In this way, all stations are considered at the same time in the level 2, unlike one model per station as we sometimes see in the literature (Lepage & Morency, 2021; Toque et al., 2017). This is more efficient and may avoid overfitting. The daily observations are in the level 1. Level-2 variables (Z_i) characterize the neighborhood of the stations and are constant over time, whereas level-1 variables (t_{ij}, X_{ij}) vary from day to day, with indices i (for station) and j (for day).

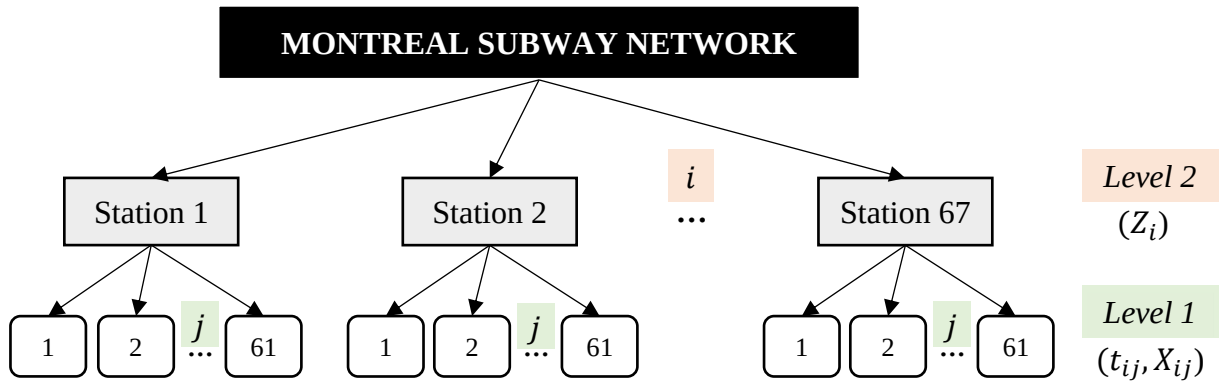


Figure 7.2 Hierarchical structure of the growth model developed in this paper

This two-level structure is expressed by two equations. The level-1 submodel (Equation 1) captures the daily changes in subway use at each station, hence within-station variations over time.

$$Y_{ij} = \pi_{0i} + \pi_{1i}t_{ij} + \pi_{2i}X_{ij} + \varepsilon_{ij} \quad (1)$$

where Y_{ij} is the subway ridership at the station i on the day j . t_{ij} takes values from 0 to 60 to account for elapsed time since September 1st, 2016 (in days). π_{0i} is the initial status of the station i (i.e., ridership at time 0), hence the intercept of its change trajectory (ridership evolution over time), while π_{1i} is the daily rate of change of the station i , hence the slope of its trajectory, controlling for independent variables. Level-1 predictors (for instance, X_{ij}) are added to evaluate the direct influence (quantified with the π_{2i} coefficient) of other time-varying phenomena on the boardings observed at the station.

The level-2 submodel (Equation 2, for random intercept and slope) measures between-station differences in daily subway use. The dependent variables of this submodel are the coefficients of the level-1 submodel. The between-station variations are explained by stationary independent variables (for instance, Z_i) describing the neighborhood of each station i .

$$\begin{aligned}\pi_{0i} &= \gamma_{00} + \gamma_{01}Z_i + \xi_{0i} \\ \pi_{1i} &= \gamma_{10} + \gamma_{11}Z_i + \xi_{1i} \\ \pi_{2i} &= \gamma_{20}\end{aligned}\tag{2}$$

where γ_{00} and γ_{10} represent respectively the average initial subway ridership and the average daily change of rate (across all stations) when all independent variables are zero. γ_{01} and γ_{11} account for the difference in average initial subway ridership and in average daily change of rate for the variation of one unit of the independent variable Z_i . π_{2i} can be assumed constant and equal to γ_{20} . Two types of residuals are present in the model: level-1 residuals (ε_{ij}) and level-2 random effects (ξ_{0i}, ξ_{1i}). Both are assumed to follow normal distributions (see Equation 3), with zero mean and within-station variance (σ_ε^2) or between-station variances (σ_0^2, σ_1^2) and covariance (σ_{01}). Note that normal distributions are considered (and not Poisson-family distributions) because the “zero” value is not possible in our main data (daily subway boardings per station); this violates one of the postulates of Poisson law. Moreover, the normality was obtained by transforming the independent variable of the model (see 7.3.2.2).

$$\begin{aligned}\varepsilon_{ij} &\sim N(0, \sigma_\varepsilon^2) \\ \begin{bmatrix} \xi_{0i} \\ \xi_{1i} \end{bmatrix} &\sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{bmatrix}\right)\end{aligned}\tag{3}$$

The combination of equations 1 and 2 gives the composite formulation of the growth model (Equation 4). This is the specification required by most multilevel statistical software programs. A structural and a stochastic part can be isolated in order to distinguish fixed and random effects, that is why the multilevel model is mixed-effects. For more details on the growth model, see the book of Singer et Willett (2003).

$$Y_{ij} = \underbrace{\gamma_{00} + \gamma_{10}t_{ij} + \gamma_{20}X_{ij} + \gamma_{01}Z_i + \gamma_{11}t_{ij}Z_i}_{\text{Structural part (fixed effects)}} + \underbrace{\xi_{0i} + \xi_{1i}t_{ij} + \varepsilon_{ij}}_{\text{Stochastic part (random effects)}}\tag{4}$$

7.3.2.2 Variables of the model

The daily number of boardings per subway station over two months (61 observations per station) was computed by summing daily validations at each station over the study period. To learn about the specification of our model, the evolution of this ridership indicator is drawn for 9 randomly selected stations in Figure 7.3. The global use of the stations (given by the blue regression line) tends to flat or slowly increase, but for all stations we observe cyclical variations around the linear trend according to the day of the week (from Monday to Sunday). Two public holidays (on September 5 and October 10) are also noticeable. Such calendar information falls within the independent variables of the model. To comply with the normality assumption of the multilevel model, the logarithm of the daily number of boardings per station is calculated and used as the dependent variable of our model. Therefore, the proposed model is a trip generation model for the subway.

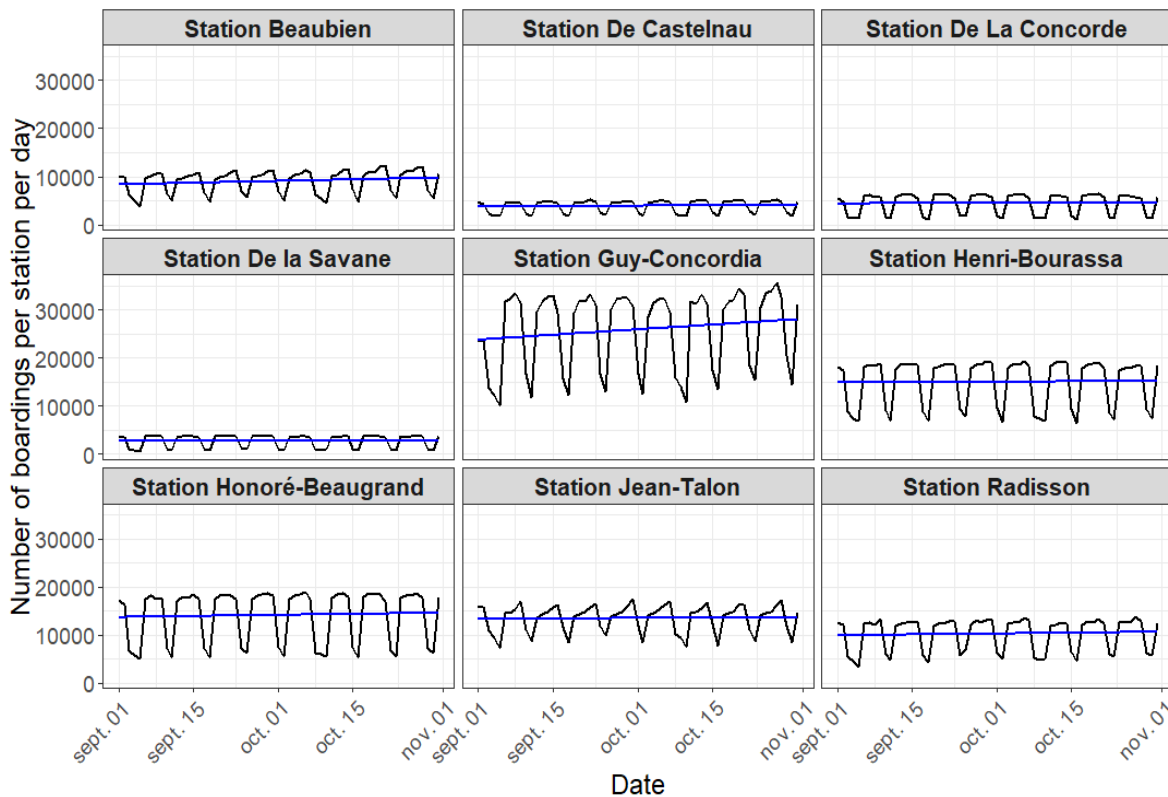


Figure 7.3 Trajectories (evolution of the daily number of boardings over time) for 9 randomly selected subway stations

All the independent variables estimated from the previously mentioned datasets are listed in Table 7.1. The role of each variable, its type as well as some descriptive statistics are also provided.

Level-1 (time-varying) variables include calendar (type of day, business day, social events) and weather (temperature, rainfalls) information. Saturday, Sunday, and public holidays are considered as non-business days. Calendar and weather predictors are typically used in longitudinal transportation modeling and may also interact (Zhang, Z. et al., 2022; Zhou, M. et al., 2017). The impacts of social events were less studied in the literature but, using the same dataset as in this paper, Lepage et Morency (2021) showed that they can particularly influence the Montreal subway ridership. However, in this paper, the different categories of events are not distinguished; only the holding of an event near a subway station on a given day is considered with a binary variable. Furthermore, the station-specific use of three alternative modes is assessed by the daily number of bikesharing, free-floating carsharing and taxi destinations within 800 meters around the subway station. Therefore, the influence of alternative modes is gauged in terms of use: we do not only consider the presence of the service with a binary variable as in other papers (Graehler Jr et al., 2019; Ma, T. & Knaap, 2019). This allows better quantifying the magnitude of the influence (accounting for its variability) and better reflecting the real impacts on ridership. Note that destinations (and not origins) are computed since only boardings are available for the subway and the aim is to detect potential complementarity between modes namely how other modes may be used to access subway stations (first-mile problem). In this way, the correlations (in use) between the three alternative mode destinations and subway boardings are considered as a proxy for potential intermodal relations. This assumption of potential intermodality is obviously simplistic (not all users of alternative modes who travel near the stations will necessarily enter the subway), but this is due to the complete independence of the passive datasets: we do not have trip data by individual, so we do not know the sequences of modes which are really combined to make trips. Such a definition can be seen as the upper bound of potential intermodality.

Level-2 predictors correspond to station neighborhood features (within the 800 meters of the station). They include typical variables which were found to affect transit ridership at the station level in previous studies and they are classified into three categories: socio-demographic and economic attributes, external and intermodal connectivity, as well as land use and built environment. The correlation of transit ridership with gender, age, income and car ownership is well known (Chiou et al., 2015; Chu, 2004), that is why all these features are considered in the tested variables. The car access rate is calculated to reflect car accessibility rather than car ownership. The availability of other modes is assessed in the following category: external and

intermodal connectivity. External connectivity refers to the location of the station in the network and is often measured by the distance (or time) to the city center (An et al., 2019; Zhao et al., 2013). Intermodal connectivity is quantified by indices which indicate the presence of infrastructures for feeder modes (bus stops, bike-sharing stations and docks, walkability) (Jun et al., 2015). Moreover, the walkability index, analyzed by Lefebvre-Ropars et Morency (2018) for Montreal, accounts for one of the three built environment dimensions ('3Ds') which typically influence travel demand, namely the pedestrian-oriented design (Cervero & Kockelman, 1997). The other two 'D' are examined in the last variable category with (population and employment) density and land-use diversity, using common indicators such as area proportion by land-use type (Jun et al., 2015; Li, S. et al., 2020) and number of points of interest (POIs) (An et al., 2019). Some other level-2 control variables (for instance, the education level) may also have an impact on the subway station use but were not included or not observed in the available data. However, the multilevel approach partially corrects for this omission and reduces the associated estimation biases by considering the clustered structure of the data and adding station-specific intercepts.

The number of different observations per variable (third column of Table 7.1) consists of three options: the level-2 independent variables involve 67 distinct values (one for each station) while 61 different observations (one for each day) are available for day-varying variables, or $61 \times 67 = 4087$ observations for variables varying in both dimensions (space and time). However, all these variables were gathered in a single dataset using a *station-period* format (4087 rows), hence each station has multiple records and spatial values are repeated. The produced database, which is the fusion of all information, is said *balanced* (same number of observations per station) and *time-structured* (same data collection schedules across all stations).

7.3.2.3 Correlation analysis

In addition, the correlation matrix between all pairs of variables is shown in Figure 7.4. The red color indicates a negative correlation while the blue color reveals a positive correlation (based on the Pearson coefficient). The strongest correlations with the dependent variable (the logarithm of the daily number of boardings per station) are observed for the business day binary variable, but also for the main independent variables of interest (destinations of the other three modes), and these correlations are positive.

Table 7.1 Description of the tested variables (selected variables in gray)

Information	Variable	Nb. obs.	Role	Type	Min	Max	Mean	St. Dev.
SUBWAY RIDERSHIP	Number of boardings per station per day	4087	DV	discrete	635.00	47947.00	10653.49	8238.63
	Log (number of boardings per station per day)	4087	DV	discrete	6.45	10.78	9.00	0.76
CALENDAR	Time	61	IV level 1	discrete	0.00	60.00	30.00	17.75
	Type of day (from Monday to Sunday)	61	IV level 1	categorical	-	-	-	-
	Business day	61	IV level 1	binary	0.00	1.00	0.67	0.47
	Event during a specific day near a specific station	4087	IV level 1	binary	0.00	1.00	0.01	0.12
WEATHER	Average daily temperature [°C]	61	IV level 1	continuous	1.30	23.50	13.53	5.75
	Total daily rainfalls [mm]	61	IV level 1	continuous	0.00	32.80	3.18	7.22
ALTERNATIVE MODES	Number of bikesharing destinations per day within 800 m around station	4087	IV level 1	discrete	0.00	2809.00	337.67	519.74
	Number of carsharing destinations per day within 800 m around station	4087	IV level 1	discrete	0.00	207.00	31.72	32.10
	Number of taxi destinations per day within 800 m around station	4087	IV level 1	discrete	0.00	1819.00	231.34	304.11
SOCIO-DEMOGRAPHIC AND ECONOMIC ATTRIBUTES (within 800 m around station)	Proportion of men	67	IV level 2	continuous	43.82	60.32	50.28	3.86
	Average age	67	IV level 2	continuous	33.04	52.81	39.16	3.22
	Proportion of people between 0 and 14 y.o.	67	IV level 2	continuous	4.18	22.05	12.06	4.26
	Proportion of people between 15 and 24 y.o.	67	IV level 2	continuous	5.44	35.27	13.57	5.40
	Proportion of people between 25 and 39 y.o.	67	IV level 2	continuous	14.83	43.27	29.77	6.58
	Proportion of people between 40 and 64 y.o.	67	IV level 2	continuous	20.40	40.67	30.12	4.57
	Proportion of people 65 y.o. and more	67	IV level 2	continuous	7.71	36.69	14.48	5.31
	Average median after-tax income in private households	67	IV level 2	continuous	13374.35	39853.85	26747.67	4879.46
	Average % of the population with after-tax income < \$ 20,000	67	IV level 2	continuous	15.71	48.21	35.69	6.61
	Average access rate to car (average number of cars/driving licenses in household)	67	IV level 2	continuous	0.30	0.81	0.50	0.10
EXTERNAL AND INTERMODAL CONNECTIVITY (within 800 m around station)	Distance to the city center (Euclidean distance) [m]	67	IV level 2	continuous	343.91	13499.69	4936.12	3109.05
	Travel time by subway to the city center (Network schedule + 5 min per transfer) [s]	67	IV level 2	continuous	0.00	1538.00	836.39	384.63
	Number of bus stops	67	IV level 2	discrete	2.00	99.00	68.60	19.10
	Number of bikesharing stations	67	IV level 2	discrete	0.00	39.00	11.48	11.37
	Number of bikesharing docks	67	IV level 2	discrete	0.00	7158.00	1522.82	1803.81
	Walkability index	67	IV level 2	continuous	5.76	38.87	17.63	9.47
LAND USE AND BUILT ENVIRONMENT (within 800 m around station)	Population density (number of people)	67	IV level 2	discrete	3307.63	36720.14	16193.43	7750.66
	Job density (number of destinations for work purpose)	67	IV level 2	discrete	1021.00	173724.00	25531.49	42358.61
	Proportion of area classified as "residential"	67	IV level 2	continuous	8.33	54.89	33.56	12.44
	Proportion of area classified as "commercial"	67	IV level 2	continuous	0.20	33.73	6.98	5.62
	Proportion of area classified as "office"	67	IV level 2	continuous	0.02	24.93	3.89	5.70
	Proportion of area classified as "industry"	67	IV level 2	continuous	0.00	23.58	2.61	4.50
	Proportion of area classified as "institutional"	67	IV level 2	continuous	1.52	57.91	11.45	9.51
	Proportion of area classified as "park"	67	IV level 2	continuous	0.16	41.27	4.92	5.27
	Proportion of area classified as "public utility"	67	IV level 2	continuous	14.29	49.36	33.12	7.23
	Proportion of area classified as "agricultural"	67	IV level 2	continuous	0.00	1.84	0.03	0.23
	Proportion of area classified as "vacant land"	67	IV level 2	continuous	0.09	9.65	1.90	1.92
	Proportion of area classified as "hydrography"	67	IV level 2	continuous	0.00	23.61	1.50	4.32
	Proportion of area classified as "golf"	67	IV level 2	continuous	0.00	2.41	0.04	0.29
	Number of points of interest	67	IV level 2	discrete	0.00	469.00	22.19	69.71
	Number of public schools	67	IV level 2	discrete	0.00	10.00	3.27	2.43
	University	67	IV level 2	binary	0.00	1.00	0.15	0.36
	Hospital	67	IV level 2	binary	0.00	1.00	0.27	0.45

Note: DV = Dependent variable, IV = Independent variable, y.o. = years old

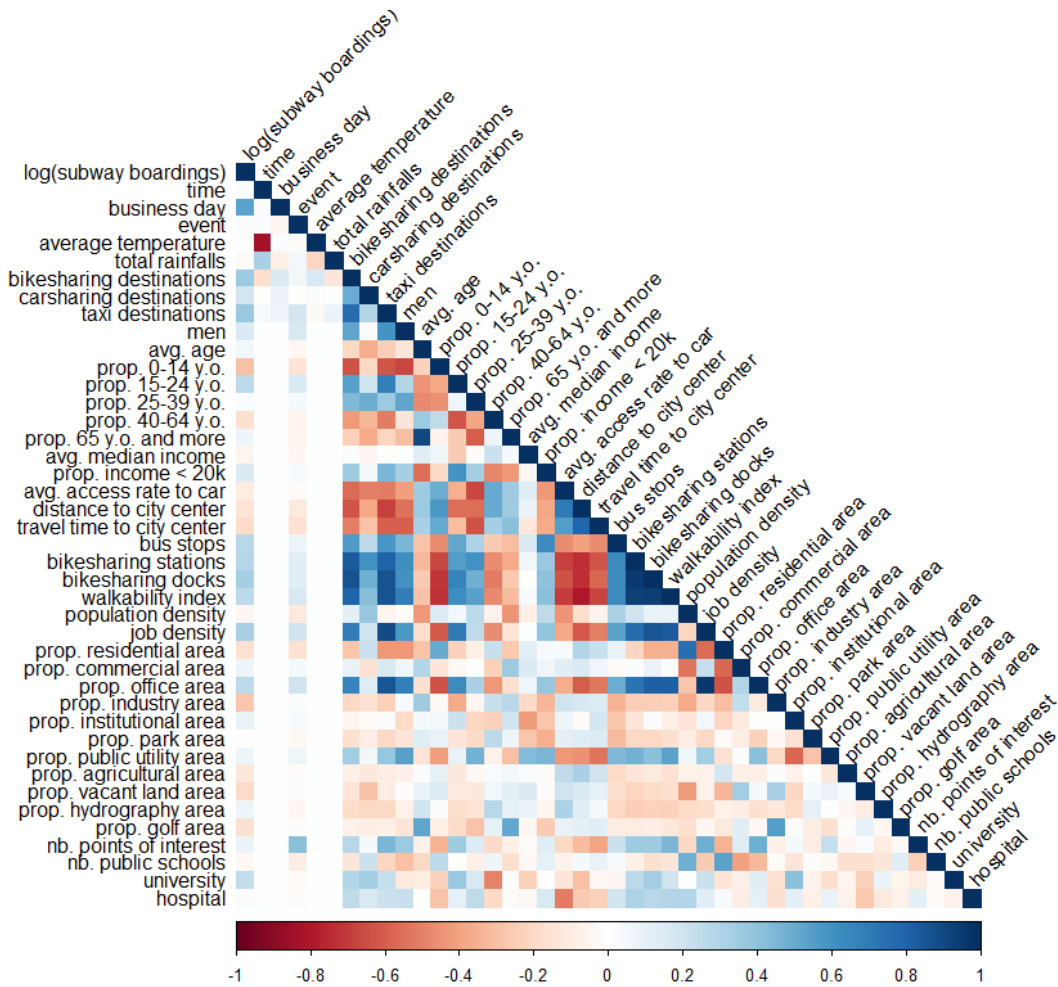


Figure 7.4 Pearson correlation coefficients between all pairs of variables

Regarding the level-1 (observation level) independent variables, there is a strong negative correlation between time and temperature since the climate cools down during the fall in Montreal. However, this variable is still kept because the corresponding variance inflation factor (VIF), which measures the amount of multicollinearity in the model, is lower than the usual threshold of 10. Moreover, we tried to remove the variable from the model, and this did not change the interpretation of the other effects. Therefore, the coefficient of this variable should be interpreted with caution, but it does not affect the other effects. The destinations of the three alternative modes are also positively correlated, but again the VIF values (all being inferior to 5) allow us to conclude that these correlations are not a major issue. Nevertheless, as they are the main effects of interest in this paper, precautions are taken by building three independent models (one for each mode) in the second part of the results (when modeling the subway use by time period and station group).

The level-2 (station level) independent variables are much more numerous, and correlations are inevitable. Among the socio-demographic attributes, the average age is eliminated in the following analysis since it is strongly correlated with the proportion of people of 65 years old and more. Distance to the city center and travel time by subway to the city center are obviously correlated, hence each variable is tested and the one with the most important effect is kept (namely, distance). Three intermodal connectivity indicators (bikesharing stations, docks and walkability) are strongly correlated with many variables, especially the destinations of the three studied alternative modes; therefore, they are finally removed and, as the highest density of transportation services is in the city center (negative correlation with the distance to the city center), the latter can be considered as a proxy for good bikesharing coverage and good walkability. Among the land use and built environment variables, the proportion of area classified as “office” is also discarded since it is strongly correlated with many other variables, especially job density. The proportion of residential area is strongly related to population density; thus, both are tried but only one of them is selected (namely, population density).

Finally, only the uncorrelated variables are included in the model formulation and construction. The VIF of all variables selected by the stepwise selection steps (see models D and E below) were checked and had low values (the only value greater than 5 is for the temperature: $VIF = 6.8$). Furthermore, as the level-2 variables are replaced by station group binary variables in a second step (see section 7.4.2 of the results), correlation becomes less problematic.

7.3.2.4 Model formulation and construction

Finally, the composite formulation of our model is given by Equation 5. The logarithmic number of boardings at station i on day j ($\log SUB_{ij}$) is written as a linear function of time ($TIME_{ij}$), modulated by the effect of the type of day with dummy factors from Monday to Sunday ($DAY_p, p = 1, \dots, 7$) and a binary variable for business versus non-business day ($BD_{ij} = 1$ or 0). The other level-1 time-varying predictors ($(X_m)_{ij}$) and level-2 independent variables ($(Z_k)_i$) of Table 7.1 previously selected in the correlation analysis are added (m and k are respectively the number of level-1 and level-2 variables included in the model).

$$\begin{aligned}
\log SUB_{ij} = & \gamma_{00} + \gamma_{10}TIME_{ij} + \gamma_{20}BD_{ij} + \sum_p \gamma_{p0}(DAY_p)_{ij} \\
& + \sum_m \gamma_{m0}(X_m)_{ij} + \sum_k \gamma_{0k}(Z_k)_i + \xi_{0i} + \xi_{2i}BD_{ij} + \varepsilon_{ij}
\end{aligned}$$

(5)

with

$$\begin{aligned}
\varepsilon_{ij} & \sim N(0, \sigma_\varepsilon^2) \\
\begin{bmatrix} \xi_{0i} \\ \xi_{2i} \end{bmatrix} & \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{02} \\ \sigma_{02} & \sigma_2^2 \end{bmatrix}\right)
\end{aligned}$$

Two random effects are considered: one for the intercept (ξ_{0i}) and the other for the impact of the “business day” binary variable (ξ_{2i}). Slope is not modeled as random because the study period is too short to detect significant variations between stations. This allows the trajectory of each specific station i to be scattered around the modeled trajectory (the average trajectory across all stations) in two ways: through the intercept value and the business day effect. Figure 7.5 shows a prototypical example (growth model without explanatory variables except day type) of such trajectories as well as the corresponding equations: ξ_{0i} and ξ_{2i} represent deviations between the coefficients of a station i (randomly selected) from the average parameters γ_{00} and γ_{20} . Note that, even if the predicted values (blue curve) underestimate the observed values (black curve), the model almost perfectly reproduces the shape of the weekly cycle. All station trajectories are normally distributed around the modeled average trajectory.

To achieve this model, we followed the traditional estimation method of a growth model, composed of various stages, hence a taxonomy of models was built (Singer & Willett, 2003). The different models of the taxonomy (from A to G) are described in Table 7.2. An unconditional means model (**model A**: $\log SUB_{ij} = \gamma_{00} + \xi_{0i} + \varepsilon_{ij}$), without independent variables, and an unconditional growth model (**model B**: $\log SUB_{ij} = \gamma_{00} + \gamma_{10}TIME_{ij} + \xi_{0i} + \varepsilon_{ij}$), including time effect, were first estimated, before considering the influence of the day type in a third step (**model C**). Then, level-2 and level-1 uncorrelated independent variables were introduced (respectively in **models D** and **E**) and a stepwise selection method was applied. This process keeps only the most relevant variables in the models (the less significant variables are removed) based on AIC (Akaike information criterion). This results in a more parsimonious model. The selected variables are in gray in Table 7.1. Moreover, the quantitative (continuous and discrete) variables were centered on their means to facilitate interpretation of the model coefficients (**model F**). Finally, to improve the precision of estimates of the fixed effects, the normal distribution of the residuals was modeled by

testing different error covariance structures (**model G**). The selected structure (based on goodness-of-fit indicators) is a heterogeneous autoregressive error covariance matrix, with standard deviation depending on the BD (business day) binary variable (Equation 6). This means that two observations closer in time are more likely to be correlated than two observations farther apart (ρ being the first order autoregressive parameter, hence autocorrelation between observations measured at time t and $t - 1$), and variance is different between business (α^2) and non-business days (β^2). In this way, we allow for correlated and heteroscedastic residuals at level 1, thus within station.

$$\varepsilon_{ij} \sim N(0, \Sigma)$$

with

$$\Sigma = \begin{pmatrix} \sigma_1^2 \mathbf{1} & \sigma_2 \sigma_1 \rho & \sigma_3 \sigma_1 \rho^2 & \dots & \sigma_T \sigma_1 \rho^{T-1} \\ \sigma_2 \sigma_1 \rho & \sigma_2^2 \mathbf{1} & \sigma_3 \sigma_2 \rho & \dots & \sigma_T \sigma_2 \rho^{T-2} \\ \sigma_3 \sigma_1 \rho^2 & \sigma_3 \sigma_2 \rho & \sigma_3^2 \mathbf{1} & \dots & \sigma_T \sigma_3 \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_T \sigma_1 \rho^{T-1} & \sigma_T \sigma_2 \rho^{T-2} & \sigma_T \sigma_3 \rho^{T-3} & \dots & \sigma_T^2 \mathbf{1} \end{pmatrix} \quad (6)$$

and

$$T = 60$$

$$\sigma_i = \begin{cases} \alpha & \text{if } BD = 1 \\ \beta & \text{if } BD = 0 \end{cases}$$

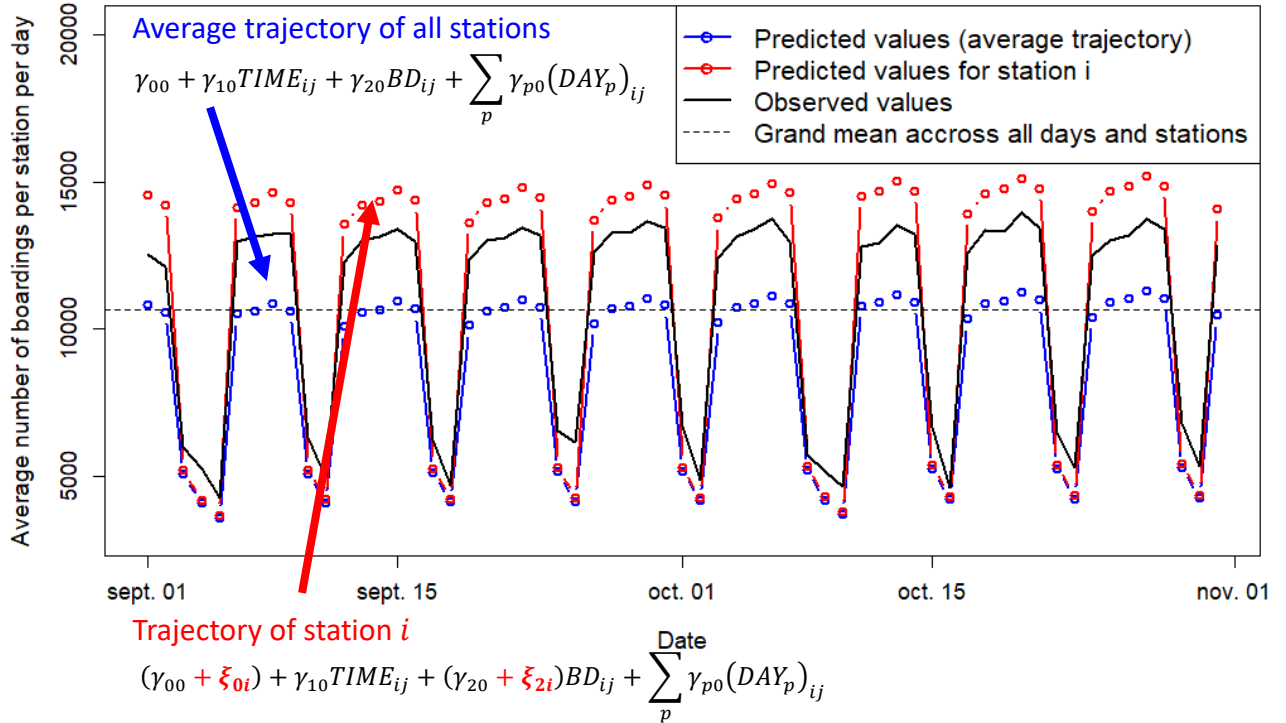


Figure 7.5 Growth model including day type effect (model C): Variations of the station-specific trajectories around the modeled average trajectory

Table 7.2 Definition of the taxonomy of growth models (from A to G)

NAME	MODEL	EQUATION
A	Unconditional means model	$\log SUB_{ij} = \gamma_{00} + \xi_{0i} + \varepsilon_{ij}$
B	Unconditional growth model	$\log SUB_{ij} = \gamma_{00} + \gamma_{10}TIME_{ij} + \xi_{0i} + \varepsilon_{ij}$
C	Growth model with effect of day type	$\log SUB_{ij}$ $= \gamma_{00} + \gamma_{10}TIME_{ij} + \gamma_{20}BD_{ij} + \sum_p \gamma_{p0}(DAY_p)_{ij}$ $+ \xi_{0i} + \xi_{2i}BD_{ij} + \varepsilon_{ij}$
D	Growth model with effects of level-2 independent variables → Stepwise selection with AIC	$\log SUB_{ij}$ $= \gamma_{00} + \gamma_{10}TIME_{ij} + \gamma_{20}BD_{ij} + \sum_p \gamma_{p0}(DAY_p)_{ij}$ $+ \sum_k \gamma_{0k}(Z_k)_i$ $+ \xi_{0i} + \xi_{2i}BD_{ij} + \varepsilon_{ij}$
E	Growth model with effects of level-1 independent variables → Stepwise selection with AIC	$\log SUB_{ij}$ $= \gamma_{00} + \gamma_{10}TIME_{ij} + \gamma_{20}BD_{ij} + \sum_p \gamma_{p0}(DAY_p)_{ij}$ $+ \sum_m \gamma_{m0}(X_m)_{ij} + \sum_k \gamma_{0k}(Z_k)_i$ $+ \xi_{0i} + \xi_{2i}BD_{ij} + \varepsilon_{ij}$
F	Model E with variables centered on their mean → To facilitate interpretation	$\log SUB_{ij}$ $= \gamma_{00} + \gamma_{10}TIME_{ij} + \gamma_{20}BD_{ij} + \sum_p \gamma_{p0}(DAY_p)_{ij}$ $+ \sum_m \gamma_{m0}(X_m^{CENT})_{ij} + \sum_k \gamma_{0k}(Z_k^{CENT})_i$ $+ \xi_{0i} + \xi_{2i}BD_{ij} + \varepsilon_{ij}$
G	Model F + modeling residuals → To improve the precision of estimates of the fixed effects	Same equation as F with $\varepsilon_{ij} \sim N(0, \Sigma)$

7.4 Results

7.4.1 Growth model of the daily subway use

The resulting taxonomy of models (from model A to G) is presented in Table 7.3. All these models were fit with the "nlme" package in the R software using the maximum log-likelihood method. All goodness-of-fit indicators were improved (pseudo R^2 increased while deviance, AIC and BIC decreased) from model A to model G. The pseudo R^2 for our final model G (53.5%) is quite good. The accuracy of this model was also validated by a forecasting exercise (with a test dataset): we

applied our model to forecast the first fifteen days of November (before the end of the bikesharing services in Montreal). The errors were compared with those obtained from three other modeling approaches: a global linear model (one model for all stations), a fixed-effects linear model (without level-2 random effects but with dummy variables to account for the station-specific intercepts), and individual linear models (one model per station). These three models were trained using the same variables selected in model G and then applied to the test dataset. The mean error rate and the root-mean-square error (RMSE) of each model are provided in Table 7.4. On average, model G is mistaken for 567 boardings per station per day, corresponding to an error rate of 5%, which is at least twice lower than the error rates from all other tested models. This demonstrates that the multilevel approach, controlling for the clustered structure and offering greater flexibility in residual modelling, better fits the data. However, forecasting is not the primary goal of this paper. Interpreting the effects of various urban interactions with the subway is the main interest.

7.4.1.1 Interpretation of the fixed effects

In the first part of Table 7.3 are reported the estimates of the fixed effects for both level-1 and level-2 independent variables. As the dependent variable of the model is a logarithm, the coefficients must be transformed using the function $x \mapsto 100(e^x - 1)$ before interpreting them. The result, given in the last column of Table 7.3 for the final model G, is equivalent to the percentage change in the initial dependent variable (the number of boardings per station per day) per unit difference in the independent variable. For instance, this means that, on average, subway ridership increases by about 0.1% [$100(e^{9.54E-04} - 1) = 0.0955$] per station per day. Regarding the effect of the day type, the results reveal significant differences between days. This confirms that the average day is not representative of the whole week. As expected, ridership is higher (by 175% on average) during business days. Compared to Mondays, all other weekdays are busier (positive coefficients), especially Thursdays. During non-business days, more boardings are counted on Saturdays than on Sundays, but Sundays are also busier than non-business Mondays (reference case). Moreover, the daily number of boardings tends to increase by 14% when an event occurs near the station.

Table 7.3 Results of the taxonomy of growth models to forecast the daily use of subway

		MODEL A	MODEL B	MODEL C	MODEL D	MODEL E	MODEL F	MODEL G	
Fixed effects									100($e^x - 1$)
Intercept and slope	Intercept	9.00E+00 ***	8.96E+00 ***	8.18E+00 ***	7.68E+00 ***	7.48E+00 ***	8.26E+00 ***	8.23E+00 ***	3.76E+05
	Time		1.18E-03 **	7.84E-04 ***	7.84E-04 ***	3.33E-04 *	3.33E-04 *	9.54E-04 ***	9.55E-02
Day type (Reference = non-business Monday)	Tuesday (0 or 1)			4.65E-02 ***	4.65E-02 ***	3.64E-02 ***	3.64E-02 ***	3.79E-02 ***	3.86E+00
	Wednesday (0 or 1)			5.62E-02 ***	5.62E-02 ***	4.13E-02 ***	4.13E-02 ***	4.09E-02 ***	4.17E+00
	Thursday (0 or 1)			7.98E-02 ***	7.98E-02 ***	4.55E-02 ***	4.55E-02 ***	5.57E-02 ***	5.72E+00
	Friday (0 or 1)			5.61E-02 ***	5.61E-02 ***	4.85E-03	4.85E-03	1.43E-02 ***	1.44E+00
	Saturday (0 or 1)			3.45E-01 ***	3.45E-01 ***	3.26E-01 ***	3.26E-01 ***	3.20E-01 ***	3.77E+01
	Sunday (0 or 1)			1.28E-01 ***	1.28E-01 ***	1.26E-01 ***	1.26E-01 ***	1.16E-01 ***	1.23E+01
	BD - Business day (0 or 1)			1.03E+00 ***	1.03E+00 ***	9.24E-01 ***	1.02E+00 ***	1.01E+00 ***	1.75E+02
Station features within 800 m	Job density				5.00E-06 ***	7.00E-06 ***	7.00E-06 ***	7.00E-06 ***	6.81E-04
	Prop. Industrial area				-4.16E-02 **	-4.01E-02 **	-4.01E-02 **	-4.02E-02 **	-3.94E+00
	Prop. Age > 65				2.90E-02 *	2.65E-02 *	2.65E-02 *	2.62E-02 *	2.66E+00
	University (0 or 1)				3.42E-01 .	3.89E-01 *	3.89E-01 *	3.94E-01 *	4.82E+01
	Distance to the city center					4.70E-05 .	8.60E-05 **	7.90E-05 **	7.94E-03
Social events	Event near station (0 or 1)					1.36E-01 ***	1.36E-01 ***	1.28E-01 ***	1.37E+01
Weather	Total rainfalls					-4.49E-03 ***	-4.49E-03 ***	-4.77E-03 ***	-4.76E-01
	Average temperature					-6.37E-03 ***	-6.37E-03 ***	-5.46E-03 ***	-5.45E-01
	BD x Total rainfalls					4.95E-03 ***	4.95E-03 ***	4.16E-03 ***	4.17E-01
	BD x Average temperature					5.80E-03 ***	5.80E-03 ***	6.49E-03 ***	6.51E-01
Modes of transport	Bikesharing destinations					-6.00E-06	1.48E-04 ***	5.70E-05 *	5.72E-03
	Carsharing destinations					-3.68E-04 .	3.89E-04 **	5.05E-04 ***	5.05E-02
	Taxi destinations					1.20E-04 ***	6.21E-04 ***	6.13E-04 ***	6.13E-02
	Dist. to CC x Bikesharing desti.					3.12E-08 ***	3.12E-08 ***	1.73E-08 **	1.73E-06
	Dist. to CC x Carsharing desti.					1.53E-07 **	1.53E-07 **	7.78E-08 .	7.78E-06
	Dist. to CC x Taxi desti.					1.02E-07 ***	1.02E-07 ***	1.05E-07 ***	1.05E-05
Variance components									
Level 1	Within-stations σ_ε^2	0.197	0.197	0.007	0.007	0.006	0.006	-	$\rho = 0.420$ $\alpha^2 = 0.003$ $\beta^2 = 0.006$
	Between-stations (intercept) σ_0^2	0.380	0.380	0.419	0.335	0.329	0.329	0.327	
Level 2	Between-stations (BD) σ_2^2			0.076	0.076	0.075	0.075	0.068	
	Covariance σ_{02}			-0.293	-0.483	-0.503	-0.503	-0.500	
Pseudo R^2 and Goodness-of-fit									
	Pseudo R^2		0.001	0.295	0.513	0.530	0.530	0.535	
	Pseudo R_ε^2		0.002	0.965	0.965	0.970	0.970	0.987	
	Pseudo R_0^2			-0.103	0.119	0.134	0.134	0.139	
	Pseudo R_2^2				0.000	0.014	0.014	0.101	
	Deviance	5282.2	5273.3	-7852.6	-7879.4	-8554.9	-8554.9	-10266.9	
	AIC	5288.2	5281.3	-7826.6	-7845.4	-8496.9	-8496.9	-10204.9	
	BIC	5307.2	5306.6	-7744.5	-7738.0	-8313.8	-8313.8	-10009.1	

Significance levels: 0 '****' 0.001 '***' 0.01 '**' 0.05 '.' 0.1 '.' 1

Table 7.4 Forecast for the first 15 days of November - comparison of model G with other models

Model	Average error rate (%)	RMSE
Model G (mixed-effects linear model)	5.3	0.080
Global linear model	45.2	0.547
Fixed-effects linear model	41.9	0.538
Individual linear models	10.4	0.190

The presence of universities near the subway station leads to an increase in ridership, a relation also observed for bikesharing (El-Assi et al., 2017). Inversely, the number of boardings is lower when there are industries in the station neighborhood. A higher job density is beneficial for subway use, probably due to commuting trips. A higher proportion of elderly people is also positively correlated to subway ridership. This could be related to their lower ownership of driver's license, making them more dependent on public transport (Di Ciommo & Shiftan, 2017). The distance to the city center has a contradictory effect: the coefficient is positive, which means the station ridership increases with the distance to the city center. However, this could be explained by the higher density of stations in the city center, and/or the appeal of the city center may be implicitly contained in the other predictors. Nevertheless, we still keep the variable in the model because of the interaction terms with alternative modes (principle of marginality).

Regarding the weather parameters, rain has a negative effect on the subway ridership, but the impact is lower on business days. This result, consistent with the literature (Zhou, M. et al., 2017), was expected: it is indeed easier to cancel (or not plan) weekend activities than to not go to work under the pretext of bad weather. Temperature has a negative impact during non-business days but a positive impact during business days: this could be due to modal shifts towards active transport modes during non-business days only, for unconstrained activities. Finally, the number of bikesharing, free-floating carsharing and taxi destinations around the station is significantly and positively correlated with the number of boardings at the subway station, especially for stations located further from the city center (the interaction with the distance to the city center is positive). This suggests potential intermodality between subway and alternative modes in the urban periphery. This is consistent with previous research on interactions between bikesharing and transit (Ma, T. & Knaap, 2019; Martin & Shaheen, 2014).

7.4.1.2 Interpretation of the random effects

A few comments can also be made on the variance components. First, an intraclass correlation coefficient can be calculated by applying Equation 7 to model A.

$$r = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_\varepsilon^2} = 0.658 \quad (7)$$

The result allows us to conclude that 66% of the total variation of the model lies between stations (against 44% in the within-station variability over time). The between-station variances (σ_0^2 , σ_2^2) remain higher than the within-station variance (σ_ε^2) in all other models of the taxonomy. Additional level-2 variables should be added to better explain between-station heterogeneity. Moreover, the negative sign of the covariance (σ_{02}) indicates that stations with higher intercept tend to have fewer differences between business days and non-business days. That means that the ridership of the busiest stations tends to be more stable throughout the week. Furthermore, the estimated temporal autocorrelation measured in model G is important ($\rho = 0.42$) and variance is found twice higher for non-business days ($\beta^2 = 0.006$) than for business days ($\alpha^2 = 0.003$). This means that the residuals are strongly correlated and heteroscedastic, but this is considered in the model.

7.4.2 Growth model by time period and station group

Instead of modeling continuous effects per day and station, this section provides new results with categorical variables to disaggregate time and space (by time period and station group). This will allow us to analyze the impacts of the time of day and of the station location on the subway use, but also to specify the interactions observed previously between modes in time and space.

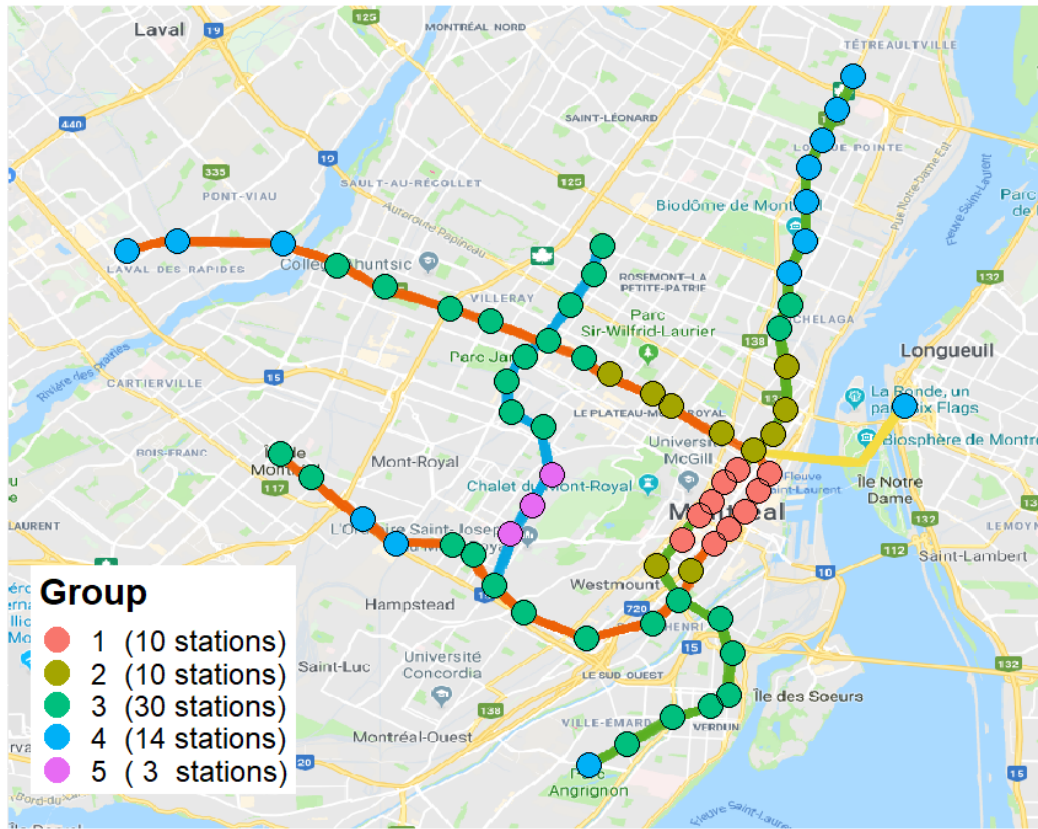
7.4.2.1 Station typology and time decomposition

The residuals of the previous model show that some subway stations are more difficult to predict than others, maybe because our analyses are too case specific. Moreover, groups of stations share similar levels of use. This is a categorical issue that we investigated by creating a station typology based on their neighborhood features, hence the level-2 independent variables of Table 7.1 (i.e., the socio-demographic and economic, connectivity, land use and built environment variables). The 5 groups obtained using a hierarchical clustering algorithm are illustrated in Figure 7.6 a). The group-1 stations are typically located in the city center, while group 5 accounts for the university

campus. The area corresponding to group 2 is dynamic too (high population and job densities) and characterized by a lower car ownership as well as easy access to alternative modes. Most of the stations which belong to group 4 are at the end of the subway lines, and the intermediate stations are gathered in group 3. Moreover, many variables may have a different effect depending on the time of the day, especially differences between peak and off-peak hours. This is why each day was decomposed into 4 periods of 5 hours: morning (AM), midday (MI), afternoon (PM) and evening (EV). Figure 7.6 b) displays these periods on the average daily pattern of boarding across all stations. An interaction between these station groups and time periods will be assumed in the following because validations are concentrated in the network periphery in the morning but in the city center in the afternoon: on an average business day, 62% of the AM validations are made at stations of group 3 and 4 (against 32% at stations of group 1 and 2, and 6% at group 5 stations), whereas 58% of the PM validations are made at stations of group 1 and 2 (against 21% at stations of group 3 and 4, and 21% at group 5 stations). This is characteristic of commuting patterns.

To quantify this phenomenon, we developed another growth model where the dependent variable is the logarithm of the number of boardings aggregated per day-period per station. Consequently, $67 \text{ stations} \times 61 \text{ days} \times 4 \text{ periods} = 16348$ observations were used to fit this model. Dummy variables were introduced to consider the effect of the group membership and time period. We also tried more complex models such as three-level and semi-parametric models but the results with dummy variables were better and simpler. Group 4 and AM period are the reference labels (group 4 was chosen because it is in these stations that the influence of alternative modes is most visible). Interaction terms were added between group and period, mode and group, and mode and period. Three models were calculated, one for each alternative mode to avoid correlation. The results are given in Table 7.5 and discussed below. The impacts of time, day type, social events and weather on subway use being comparable to those in model G, we only analyze here the interaction effects between time period, group membership and alternative modes of transport.

a)



b)

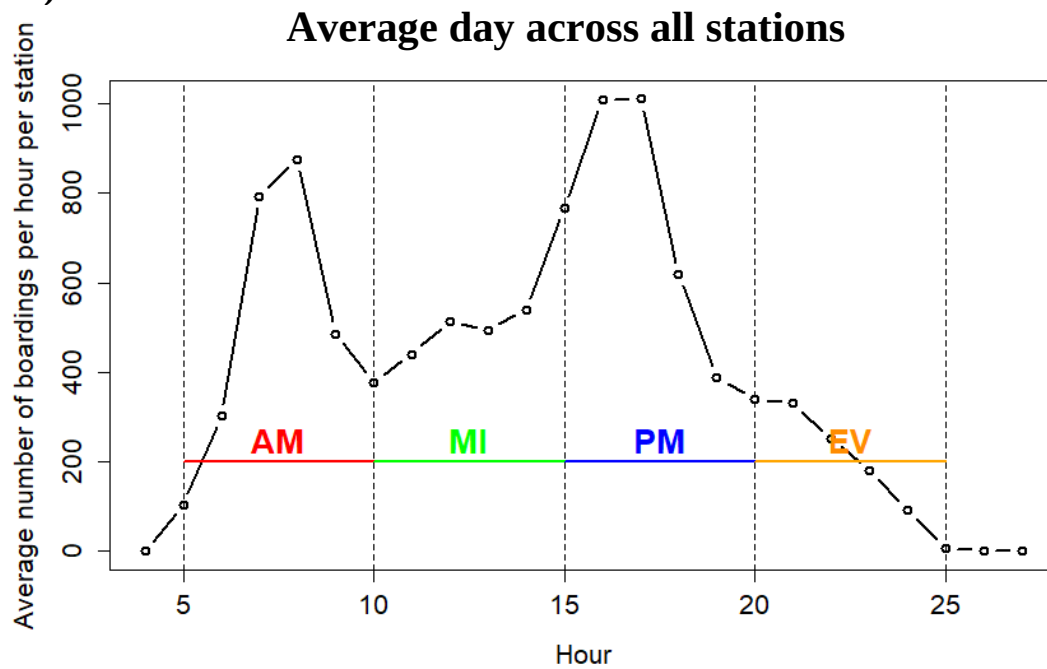


Figure 7.6 Decomposition of space (a) and time (b)

Table 7.5 Model results to forecast subway use by station group and time period

		MODEL 1		MODEL 2		MODEL 3	
		Mode = bikesharing		Mode = carsharing		Mode = taxi	
Fixed effects		Coefficient	100($e^x - 1$)	Coefficient	100($e^x - 1$)	Coefficient	100($e^x - 1$)
Intercept and slope	Intercept	7.49E+00	1.80E+05 ***	7.33E+00	1.53E+05 ***	7.01E+00	1.10E+05 ***
	Time	9.17E-04	9.18E-02 ***	5.97E-04	5.97E-02 **	4.56E-04	4.57E-02 *
Day type and time period (Reference = non-business Monday AM)	Tuesday (0 or 1)	3.83E-02	3.91E+00 ***	3.79E-02	3.86E+00 ***	3.17E-02	3.22E+00 ***
	Wednesday (0 or 1)	5.29E-02	5.43E+00 ***	5.22E-02	5.35E+00 ***	4.35E-02	4.44E+00 ***
	Thursday (0 or 1)	8.67E-02	9.06E+00 ***	8.44E-02	8.80E+00 ***	5.99E-02	6.18E+00 ***
	Friday (0 or 1)	8.45E-02	8.81E+00 ***	8.61E-02	8.99E+00 ***	5.58E-02	5.74E+00 ***
	Saturday (0 or 1)	3.23E-01	3.82E+01 ***	3.24E-01	3.82E+01 ***	2.79E-01	3.22E+01 ***
	Sunday (0 or 1)	1.09E-01	1.15E+01 ***	1.08E-01	1.14E+01 ***	9.48E-02	9.94E+00 ***
	BD - Business day (0 or 1)	9.05E-01	1.47E+02 ***	9.01E-01	1.46E+02 ***	8.84E-01	1.42E+02 ***
	MI period (0 or 1)	-4.88E-01	-3.86E+01 ***	-5.27E-01	-4.10E+01 ***	-5.17E-01	-4.04E+01 ***
	PM period (0 or 1)	-3.28E-01	-2.79E+01 ***	-3.88E-01	-3.22E+01 ***	-2.21E-01	-1.98E+01 ***
	EV period (0 or 1)	-1.65E+00	-8.08E+01 ***	-1.77E+00	-8.30E+01 ***	-1.69E+00	-8.16E+01 ***
Station group (Reference = group 4)	Group 1 (0 or 1)	-1.23E+00	-7.08E+01 ***	-1.15E+00	-6.84E+01 ***	-7.14E-01	-5.10E+01 **
	Group 2 (0 or 1)	-6.98E-01	-5.02E+01 **	-5.76E-01	-4.38E+01 *	-1.97E-01	-1.79E+01
	Group 3 (0 or 1)	-4.54E-01	-3.65E+01 *	-2.73E-01	-2.39E+01	2.15E-01	2.40E+01
	Group 5 (0 or 1)	-1.57E+00	-7.92E+01 ***	-1.78E+00	-8.32E+01 ***	-1.34E+00	-7.39E+01 ***
	Group 1 x MI period	1.31E+00	2.69E+02 ***	1.49E+00	3.44E+02 ***	1.50E+00	3.50E+02 ***
	Group 1 x PM period	2.06E+00	6.83E+02 ***	2.30E+00	8.98E+02 ***	1.62E+00	4.04E+02 ***
	Group 1 x EV period	2.19E+00	7.95E+02 ***	2.46E+00	1.07E+03 ***	2.04E+00	6.71E+02 ***
	Group 2 x MI period	6.72E-01	9.59E+01 ***	7.56E-01	1.13E+02 ***	7.16E-01	1.05E+02 ***
	Group 2 x PM period	9.44E-01	1.57E+02 ***	1.04E+00	1.82E+02 ***	7.77E-01	1.18E+02 ***
	Group 2 x EV period	1.08E+00	1.94E+02 ***	1.30E+00	2.68E+02 ***	1.14E+00	2.11E+02 ***
	Group 3 x MI period	2.06E-01	2.28E+01 ***	2.26E-01	2.53E+01 ***	1.72E-01	1.87E+01 ***
	Group 3 x PM period	2.68E-01	3.08E+01 ***	3.01E-01	3.51E+01 ***	1.90E-01	2.10E+01 ***
	Group 3 x EV period	2.55E-01	2.90E+01 ***	2.77E-01	3.19E+01 ***	2.39E-01	2.70E+01 ***
	Group 5 x MI period	1.73E+00	4.66E+02 ***	1.80E+00	5.04E+02 ***	1.75E+00	4.78E+02 ***
	Group 5 x PM period	2.21E+00	8.13E+02 ***	2.30E+00	8.98E+02 ***	2.18E+00	7.85E+02 ***
	Group 5 x EV period	2.42E+00	1.03E+03 ***	2.37E+00	9.75E+02 ***	2.35E+00	9.52E+02 ***
Social events	Event near station (0 or 1)	1.04E-01	1.10E+01 ***	1.05E-01	1.11E+01 ***	9.03E-02	9.45E+00 ***
Weather	Total rainfalls	-4.01E-03	-4.00E-01 ***	-4.41E-03	-4.40E-01 ***	-4.98E-03	-4.97E-01 ***
	Average temperature	-5.22E-03	-5.21E-01 ***	-4.73E-03	-4.72E-01 ***	-4.47E-03	-4.46E-01 ***
	BD x Total rainfalls	4.63E-03	4.64E-01 ***	4.55E-03	4.56E-01 ***	4.31E-03	4.32E-01 ***
	BD x Average temperature	5.17E-03	5.18E-01 ***	5.12E-03	5.14E-01 ***	4.91E-03	4.92E-01 ***
Mode of transport	Mode destinations	3.18E-03	3.18E-01 ***	1.68E-02	1.70E+00 ***	-6.45E-03	-6.43E-01 ***
	Mode destinations x Group 1	-3.45E-03	-3.44E-01 ***	-1.74E-02	-1.73E+00 ***	6.01E-03	6.03E-01 ***
	Mode destinations x Group 2	-3.77E-03	-3.76E-01 ***	-1.12E-02	-1.11E+00 ***	6.34E-03	6.36E-01 ***
	Mode destinations x Group 3	-3.22E-03	-3.22E-01 ***	-9.48E-03	-9.44E-01 ***	1.04E-02	1.04E+00 ***
	Mode destinations x Group 5	2.58E-03	2.58E-01 .	-8.92E-03	-8.88E-01 *	7.56E-03	7.59E-01 ***
	Mode destinations x MI period	3.89E-04	3.89E-02 ***	-5.54E-03	-5.52E-01 ***	-1.88E-04	-1.88E-02
	Mode destinations x PM period	4.87E-04	4.88E-02 ***	-5.63E-03	-5.61E-01 ***	2.94E-03	2.95E-01 ***
	Mode destinations x EV period	1.78E-03	1.78E-01 ***	5.30E-04	5.30E-02	2.52E-03	2.52E-01 ***
Variance components							
Level 1	Within-stations σ_ϵ^2	0.095		0.099		0.091	
	Between-stations (intercept) σ_0^2	0.355		0.358		0.352	
Level 2	Between-stations (BD) σ_2^2	0.057		0.060		0.061	
	Covariance σ_{02}	-0.247		-0.255		-0.284	
Pseudo R^2 and Goodness-of-fit							
	Pseudo R^2	0.555		0.549		0.563	
	Pseudo R_ϵ^2	0.848		0.841		0.854	
	Pseudo R_0^2	0.036		0.027		0.043	
	Pseudo R_2^2	-0.099		-0.161		-0.189	
	Deviance	11721.1		11846.5		11286.9	
	AIC	11815.1		11940.5		11380.9	
	BIC	12177.1		12302.4		11742.9	

Significance levels: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘.’ 1

7.4.2.2 Interaction effects of time period, group membership and alternative modes of transport

First, the influence of time period and group membership is examined using the coefficients of model 1 (but they are similar to those of models 2 and 3). The intercept of this model (7.494) is the logarithm of the average number of boardings at time 0, on a non-business Monday in the AM period, for a group 4 station, after controlling for the other independent variables. This same number is calculated for a working Monday ($BD = 1$) by $7.494 + 0.905$, then we can apply the exponential function: $\exp(7.494 + 0.905) = 4441$. To quantify the impact of other day periods and the effect of belonging to another station group, the appropriate coefficients must be added to this reference value: this is what was done to produce Table 7.6 (neglecting the effect of time, assumed to be 0). For instance, the average number of boardings for a group 1 station in the PM period is obtained by adding the coefficients of the variables *Group 1*, *PM period* and *Group 1 x PM period*, hence $\exp(8.399 - 1.230 - 0.328 + 2.058) = 7328$. The same principle is applied for all the other cells in Table 7.6. The results demonstrate that the model succeeds in capturing ridership variations between station groups and time periods. As observed before, stations near the city center (groups 1 and 2) have a higher number of boardings in the afternoon, whereas stations in the network periphery (groups 3 and 4) are busier in the morning. The boarding pattern of group-5 stations (in the university campus) looks like that of group-1 stations (in the city center).

Table 7.6 Average number of boardings per 5-hour period per station for a working Monday at time 0 (controlling for other variables), according to model 1

	Group 1	Group 2	Group 3	Group 4	Group 5
AM period	1298	2211	2820	4441*	924
MI period	2941	2658	2127	2726	3208
PM period	7328	4096	2657	3200	6080
EV period	2238	1250	701	855	2007

* Reference

Furthermore, the three models of Table 7.5 reveal that the effects of alternative modes on the Montreal subway use depend on the station group and day period. This conclusion was also drawn regarding the impact of Capital Bikeshare on Metrorail ridership in Washington (Ma, T. & Knaap, 2019). In the AM period, an increase in the number of bikesharing destinations is significantly correlated with an increase in the number of boardings at the subway stations of group 4 (reference

case of model 1). This trend is reversed for the stations of group 1, 2 and 3 (negative coefficients), but is slightly more pronounced for group-5 stations (positive coefficient). This potential complementarity increases throughout the day and reaches its maximum in the evening period, when transit services are less frequent. A strong and positive interaction is also detected between free-floating carsharing and group-4 stations (model 2). This interaction decreases for stations of groups 2, 3, 5 and becomes negative for group-1 stations. The highest correlations between free-floating carsharing and subway are measured for the AM and EV periods. The positive effect of taxi on subway ridership (model 3) is mostly detected in the PM and EV periods, for stations of groups 3 and 5. However, this impact is negative for stations of groups 1, 2 and 4. Therefore, the influence of bikesharing, free-floating carsharing and taxi use on the subway ridership is, overall, positive for stations further away from the city center (groups 3, 4 and 5) while it tends to be negative in the city center (groups 1 and 2). This confirms the previous result that the positive influence of these modes (and the potentiality of intermodal relations) increases with distance to the city center.

7.5 Conclusion

This paper provides a methodological framework for growth modeling in a transportation context and interesting findings are put forward regarding the interactions between the subway, the surrounding environment and other urban transportation modes. More specifically, a two-level growth model was developed to forecast subway ridership at the station level over two months. Built on the combination of (smart card and ticket) validation data and several contextual data sources, the proposed model makes it possible to cross two important dimensions of mobility: time (level 1 - longitudinal evolution of the daily number of boardings at each station) and space (level 2 - effect of neighborhood characteristics on the station use). Moreover, the influence of the use of three alternative modes (destinations in bikesharing, free-floating carsharing and taxi) on the daily number of boardings in subway stations was investigated using transactional data streams. These correlations were then assessed by time period and station group. In the end, the model explains the effect of various variables on subway ridership and highlights potential systemic intermodal relations at the station level.

Therefore, this paper fills the gap in the literature by combining both multimodal and longitudinal (or at least multiweek) analyses based on passive data. Unlike surveys, this type of data provides

spatial-temporal details and quantitative information for the entire transportation systems under study (not only the sampled respondents or sampled trips). This allowed us to specify the interactions between modes in time and space (first using continuous variables and then categorical variables) at the systemic level. Interactions are evaluated by the correlation between destinations and origins of trip legs as an indicator of intermodality (combination of the modes in the same trips). In this way, potential complementarity between modes (with alternative modes serving to connect to the subway network) is assessed despite the independent datasets. In addition, this is one of the only works that encompass more than two modes of transport, including more recent modes such as bikesharing and carsharing, and the interactions were measured in use (volumes) rather than only the presence/absence (binary variable) of mode infrastructure. Furthermore, a longitudinal modeling framework was proposed instead of simple descriptive statistics. The multilevel approach is appealing since it has not yet become the standard in transportation research looking at the impact of the surrounding environment on mode usage, but it has several advantages over traditional regression models. In particular, it offers greater flexibility in modeling residuals and can account for several important biases.

According to the results, the station daily use (in number of boardings) highly depends on location and neighborhood features, but also on temporal context. Ridership, higher during business days especially on Thursdays, is positively impacted by social events, while the influence of weather depends on the day type (business or non-business). The level of use of the station is higher when there are universities nearby, but lower when there are industries. Higher job density and proportion of elderly people also have a positive impact on the station boardings. More generally, different daily patterns of use (through AM, MI, PM and EV periods) were revealed depending on the station location: this accounts for commuting dynamics. Furthermore, subway ridership was found to be temporally autocorrelated, as well as heteroscedastic (more variability is measured during non-business days). The consideration of these biases in the modeling of the residuals improved the fit of the model to the data and allowed it to outperform other traditional linear models. Regarding interactions between modes, potential complementarity (intermodality) was detected between the subway and alternative modes in the urban periphery, whereas more competition may be apparent in the city center. These relationships, observed in Montreal, were also reported (mainly for bikesharing) in other cities according to the literature review. Overall, these findings can assist policymakers and planning partners in designing subway systems (e.g., determining the station

locations of a new subway network or of a subway line extension according to the neighborhood characteristics), but also in strategically positioning shared modes services to better integrate them with the subway for complementarity and accessibility purposes.

Indeed, the outcome can have several policy implications. First, this suggests that greater collaboration between the transit authority and the alternative mode operators in the network periphery could be a win-win opportunity to increase the ridership of all modes. This could be encouraged by the development of bikesharing, free-floating carsharing and taxi services close to the stations at the end of the subway lines. Moreover, payment integration, fare incentives and intermodal facilities around the subway stations could help facilitate transfers and foster multimodal trips (Ma, T. & Knaap, 2019). Such initiatives could be part of TOD (Transit Oriented-Development) development projects. In this way, alternative modes could solve the first-mile/last-mile access problem of transit, and subway systems would be better integrated in urban developments. In addition, the fact that there is no (or less) complementarity in the downtown area may warn the transit authority against the possible saturation of the subway network in the central stations: this saturation could explain the growing use of alternative modes in the city center, which become potential substitutes because of capacity issues. But this relation can also be considered as complementary since alternative modes then allow reducing transit overload.

However, such conclusions must be qualified considering the methodological limitations of this paper. First, other level-2 independent predictors should be introduced in the model to decrease the high between-station variance. Indeed, even if the multilevel model is less affected by the omission of high-level variables than the ordinary least square regression model, the estimates can still be slightly biased (Huang, F. L., 2018). The sensitivity of the “neighborhood” definition around subway stations (including the distance of 800 meters) should also be checked. Moreover, the detected relationships between modes of transport should be analyzed in causality (rather than in correlation). To this end, causal inference methods such as propensity score weighting could be mixed with the multilevel model (Schuler et al., 2016) to reduce confounding effects and generate unbiased estimates of the causal impact of the other modes on subway demand. In addition, the proposed model accounts for temporal autocorrelation but not for spatial autocorrelation. Further investigation could test a hybrid version between the multilevel model and the geographically weighted regression model (Li, S. et al., 2020). Linear models are typically used in transportation modeling because they are simpler to interpret, but as nonlinear relationships are likely to be

observed among the variables, nonparametric machine learning algorithms (for instance, gradient boosting methods) could also be examined.

Finally, the interactions between the subway and alternative modes should be further explored. Only one direction of the intermodal relation (first-mile) was examined in this paper. However, data preprocessing (inference of subway alightings) could allow expanding the method to intermodal relations between subway exits and alternative mode origins (last-mile rather than first-mile problem). Furthermore, the capacity of the different systems should be considered to control for possible distortions due to their saturation (for instance, when there are no more bikes in the nearby bikesharing stations to make the connection with the subway). More importantly, only potential intermodal connections were evaluated in this paper. The longitudinal collection of complete trips (and not only trip legs), for instance using a mobile phone application, would be required to capture real intermodal relations. Such data would also allow more disaggregate analyses, for instance at the trip level, to better specify the interactions. Indeed, aggregation based on a common spatial-temporal unit (station neighborhood and day or day-period) was needed in this paper to combine all the independent datasets, and this limited us to measuring correlations between the modes. Not only intermodal relations, but other complementarity situations should also be investigated: alternative modes can be used to fill mobility gaps when/where transit services are not available (complementarity in time or space), or when the transit trip is not competitive (waiting time or travel distance or travel time too long). Moreover, the analyses should be performed for different time scales (short and long term) to account for dynamic interactions. All these further investigations could help anticipate future challenges in increasing sustainable mobility and developing integrated planning strategies including all modes of transport.

Acknowledgments

The authors wish to thank all the companies (STM, BIXI, Communauto, Taxi Diamond) who provided data used in this paper. They also acknowledge the help of Jean-Simon Bourdeau and Éric Lacourse, as well as the financial support of the Canada Research Chairs (CRC), Quebec Research Funds (FRQNT) and Institute for Data Valorization (IVADO) programs.

CHAPITRE 8 ARTICLE 5 : ASSESSING THE IMPACTS OF THE COVID-19 PANDEMIC ON SUBWAY RIDERSHIP AND ON THE INTERACTIONS WITH OTHER TRANSPORTATION MODES

Elodie Deschaintres, Catherine Morency, Martin Trépanier

Soumis pour publication dans *Transportation*

Date de soumission initiale : 26 novembre 2022

Abstract

The COVID-19 pandemic has led governments to implement restrictive policies which have caused unprecedented effects on transportation systems. This paper assesses which measures had more impacts on subway daily ridership in Montreal (Quebec, Canada) and on the interactions between modes using time series approaches. Change point detection methods, based on regression structure and Bayesian posterior probabilities, are first applied to automatic fare collection (AFC) data available from January 2019 to December 2021. Nine breakpoints (or ten phases) are found and linked to the COVID-19 timeline of the city. The impacts are then quantified by phase, and their variability is analyzed by day type and period. The evolution of the daily and weekly patterns in subway usage is also examined using time-frequency wavelet analysis. Finally, changes in correlations between the subway ridership and the use of three other transportation modes (cycling, private car, and carsharing) are modeled using interrupted time series models with autoregressive errors. The results reveal that lockdown implementations had a negative, immediate but decreasing impact on subway use, while release measures combined with transit-specific policies (such as free distribution of masks) led to a gradual recovery. The impacts varied in time, but traditional 5-day-a-week peak hour travel declined the most. The use of the other modes, as well as their system-level interactions with the subway, were also affected. Potential modal shifts were highlighted. Such findings provide practitioners and planners with useful insights into the COVID-19 pandemic impacts on mobility.

Keywords: COVID-19 impacts; subway ridership; transportation interactions; change point detection; wavelet analysis; interrupted time series

8.1 Introduction

The global COVID-19 pandemic has caused important impacts on transportation systems all over the world. In many cities, transit has experienced the largest declines in ridership, and the recovery is currently slow: 2019 levels have still not been reached even after the lifting of most measures. Assessing the effects of the various measures implemented to prevent the spread of the virus could help policymakers and transit agencies be more resilient in case of new outbreaks or crises. Modeling the impacts could also be useful for forecasting future travel demand and proposing different scenarios since traditional models may not be as suited to such level of variability. There is a need to identify the changes in transit use behavior which could better explain the “new normal”. To this end, the longitudinal perspective should be prioritized since the COVID-19 pandemic was composed of different waves which may have led to different traveler responses, with short- or long-term consequences. Moreover, modal shifts from transit to individual modes, especially active transportation and private cars, may have occurred. Therefore, the multimodal perspective must be considered, for example by cross-analyzing changes in the use of several modes.

In line with this context, this paper aims to document the impacts of the COVID-19 pandemic on the Montreal (Quebec, Canada) subway ridership and on the interactions with three other transportation modes (cycling, private car, and carsharing). Longitudinal analyses over three complete years (from 2019 to 2021) are carried out by handling multiple daily time-series data, including the number of hospitalizations, subway boardings (tap-in validations), bike and car counts, as well as carsharing reservations. Change point detection methods are first used to find breakpoints in the subway ridership. The COVID-19-related measures that may have caused them are identified and the impacts by phase (between two breakpoints) are quantified. The variations of these systemic effects are also analyzed for different times and frequencies, in accordance with usual rhythms observed in mobility patterns. Finally, changes in correlations between subway, cycling, car and carsharing use (system-level interactions) are modeled applying an interrupted time series approach. The findings provide a better understanding of the changes in travel behavior due to the COVID-19 pandemic. Therefore, this paper can assist practitioners in better monitoring the recovery as well as for future waves of the current pandemic or other similar outbreaks.

This paper is organized as follows. First, the literature about COVID-19 impacts on mobility is summarized and the specific case of Montreal is introduced. The datasets used in this paper are then described, and the methodological framework is presented. The results are reported in three parts: the breakpoints detected in the subway ridership, as well as the magnitude of the impacts during the resulting phases, are commented, then time and frequency analyses by phase are performed, and correlations with other transportation modes are modeled. Finally, the outcomes are discussed, while the conclusion gives insights on policy and planning implications and proposes future work.

8.2 Background

The coronavirus disease 2019 (COVID-19) was recognized as a global pandemic on March 11th, 2020, by the World Health Organization (WHO) (Cucinotta & Vanelli, 2020). This resulted in many governments implementing lockdown and other public health measures, such as social distancing and mask wearing, that affected all modes of transportation around the world (Shortall et al., 2021). The impacts on travel behaviors and transportation systems have received a lot of attention in recent research, which is reviewed below. Moreover, the COVID-19-related events timeline of Montreal, as the case study of this paper, is presented.

8.2.1 Literature review

Overall, the COVID-19 pandemic triggered a significant drop in travel demand. The main goal of the first measures implemented by the governments in mid-March 2020 was indeed to restrict movements, limited to essential travel only during the lockdown. Several works justified such a strategy since strong correlation was found between mobility and disease spreading (Badr et al., 2020; Heiler et al., 2020; Santamaria et al., 2020). As a result, the trips for all purposes have fallen, especially commuting (Beck & Hensher, 2020; de Haas et al., 2020; Pawar et al., 2020). This is due to remote working and learning, requested or advised in most of the stay-at-home guidelines. The growing popularity of online services, especially during the lockdown in Montreal (Pirie et al., 2021), also had impacts on health and shopping trips (Mouratidis & Papagiannakis, 2021). Nevertheless, in proportion of total trips, the share of each travel purpose has remained constant in several cases (Beck & Hensher, 2020; de Haas et al., 2020). Instead of descriptive statistics at the trip level, geometric and graph-based methods were also used to demonstrate a reduction in human

mobility and activity space (Heiler et al., 2020; Iacus et al., 2021). The decrease in these indicators, especially the shrinking size of highly interconnected geographic zones, as well as the decline in traveled distance (de Haas et al., 2020), suggest that local travel was less affected. Attitudes, descriptive norms, protective behaviors toward COVID-19, as well as experience of information and communication technology, were found to contribute to activity-travel behavior changes, i.e., reducing out-of-home activities and travel frequency (Irawan et al., 2022). Such changes in mobility habits were observable quickly after the first COVID-19-related policies, but some of them could persist for some time. As teleworking (work-from-home) is expected to be integrated to the “new normal”, some authors underlined the importance to understand its determinants and to model its impact on travel behavior (Beck et al., 2020; Nguyen, 2021). In Montreal, it was demonstrated that 1% of teleworkers could result in a decrease of 0.36% of the total number of trips (Morency & Verreault, 2021). More flexible working conditions are also seen as an opportunity to flatten the peak hours, and then reduce road congestion and transit crowding (Beck & Hensher, 2020; Hensher, D. A., 2020).

Transit modes have experienced ridership declines of over 60% compared to the 2019 levels: see (Gkiotsalitis & Cats, 2021) for a review of several study cases. Transit agencies had different responses to the crisis. Some limited their service span, canceled certain services, and closed selected stations. Others, including Montreal, maintained high levels of service to transport essential workers and avoid overcrowding. Several capacity management methods were investigated to ensure social distancing (Hörcher et al., 2022; Tirachini & Cats, 2020). However, even when services were preserved, passengers have been reluctant to use transit because of hygiene concern (Beck & Hensher, 2020), except when no alternative is accessible (Pawar et al., 2020). Such risk perceptions also explain the slow recovery observed in the post-lockdown period. For these reasons, collective/shared modes, such as transit and carpool, may have been abandoned in favor of individual/unshared modes, such as private cars, motorize two-wheelers, taxi cabs and active travel modes (Das et al., 2021; Thombre & Agarwal, 2021). Cycling has tended to increase during the pandemic, especially on weekends and during the afternoon and early evening, for recreational activities. Much smaller increase (or decrease) was reported during weekday mornings at peak hours, in part due to the overall decline in commuting; see (Buehler & Pucher, 2021) for a complete review. Nevertheless, even when a reduction in bicycle trips was recorded, its modal share has often increased (Beck & Hensher, 2020; Bucsky, 2020; Molloy et al., 2021). Similar

trends were exhibited for bikesharing (Li, H. et al., 2021; Wang, H. & Noland, 2021). In addition, the impacts were found to depend on the bikesharing station position: the use of the stations of the London bikesharing service varied differently if they were near rail stations, hospitals, and parks (Li, H. et al., 2021). The increasing use of cycling and bikesharing could be attributable to different causes. First, in many countries, including Canada, walking and cycling have been permitted for physical activity as exceptions to travel restrictions. Consequently, round trips by walking and cycling gained in popularity (de Haas et al., 2020). Policies were also rolled out to promote cycling and ensure social distancing – see (Buehler & Pucher, 2021) for a review of all types of measures implemented. In Montreal, “public health corridors” were put in place, which increased the amount of street space allocated to active transportation and led to more fair allocation (Lefebvre-Ropars et al., 2021). Furthermore, modal shifts from transit to cycling and bikesharing may have been made (Das et al., 2021; Teixeira & Lopes, 2020). Modal shifts between transit and private cars were also revealed (Das et al., 2021). At the beginning of the pandemic, car use declined in several cities, especially car use as passengers (de Haas et al., 2020), but the car remained dominant in modal share (Bucsky, 2020) and car trips quickly increased again (rebound effect), until road congestion was back to pre-COVID levels (Molloy et al., 2021). According to survey results, people feel more comfortable to travel by car than by transit because the car is perceived to be safer (Beck & Hensher, 2020; de Haas et al., 2020). This modal shift worries researchers and practitioners, hoping that the private car will not be the big winner of this crisis. Regarding shared mobility, perceived physical risk also had a negative impact on carsharing usage intention (Garaus & Garaus, 2021). Therefore, some authors question what role mobility as a service (MaaS) may have post-COVID-19 and stress the need to study several scenarios for the short, medium and long terms (Hensher, D. A., 2020; Shaheen & Wong, 2020).

Based on our review, the COVID-19 impacts on mobility have been mainly diagnosed using cross-sectional surveys (Beck & Hensher, 2020; Das et al., 2021; Mouratidis & Papagiannakis, 2021; Pawar et al., 2020), hence at a specific point in time, or at best panel data (de Haas et al., 2020; Shakibaei et al., 2021), with descriptive analyses. This does not allow following the longitudinal effects of the pandemic on travel demand. Passive data such as GPS (Molloy et al., 2021) or mobile phone data (Badr et al., 2020; Heiler et al., 2020; Santamaria et al., 2020) and trip or count data (Bian et al., 2021; Li, H. et al., 2021; Wang, H. & Noland, 2021) were also mined, but only over a short period (mostly one wave of the pandemic), without the following waves and the recovery

phases between the waves. However, it is important to assess the variations during the different steps of this crisis to better prepare for future outbreaks and support practitioners in the recovery process. Based on longitudinal data, several methods can be used to isolate and quantify the impacts of COVID-19-related measures. The difference-in-differences (DID) approach, often used to appraise casual impact, cannot be applied in this context because the nationwide policies make it impossible to determine a control group (Li, H. et al., 2021). Some authors benefited from an exhaustive database containing information on the type, timing and intensity of the confinement measures applied in several European countries and, using ordinary regression models, they found that a large proportion of the change in mobility patterns can be explained by the COVID-19 confinement measures (Santamaria et al., 2020). Prais-Winsten models, controlling for serial correlation of type AR(1), were also built to model bikeshare and subway ridership changes during the COVID-19 pandemic in New York City (Wang, H. & Noland, 2021), while interrupted time series (ITS), based on segmented regression, and Bayesian structural time series models were used to evaluate the impacts of the lockdown and the first lockdown ease on the usage of public bicycle share in London (Li, H. et al., 2021). Change point detection (CPD) methods were investigated to study the time lag effects of COVID-19 policies on several transportation systems based on three months of passive data (Bian et al., 2021). Indeed, several works showed that behavioral mobility changes were already observable before public health directives were implemented (lead effect), indicating that people adjusted their behavior before the restrictions (due to self-awareness and pandemic risk perception) (Badr et al., 2020; Pawar et al., 2020). This paper also applies CPD methods, but the aim is to divide the time series data into different phases and to provide a better understanding of the changes in travel behaviors occurring at each phase. Correlations with other transportation modes are also modeled using ITS approach. The main originality lies in handling a longer period of longitudinal data and in cross-analyzing the COVID-19 impacts for multiple modes.

8.2.2 COVID-19 timeline in Montreal

The study case of this paper is the city of Montreal, in the province of Quebec, Canada. All events and measures related to the COVID-19 pandemic in Quebec have been listed online by the national institute of public health (INSPQ, 2020). Their large number shows that the Quebec government's policies have been rather restrictive, due to the limited hospital capacity. Five main lockdowns

were implemented, during which most non-essential businesses and services were closed. The reopening phases were also very gradual. The measures specific to the region of Montreal are summarized in Table 8.1. They are gathered into three warning periods (in red) and three release periods (in green). The first lockdown took place at the beginning of the first warning period (on March 13th, 2020), while the two following lockdowns were established during the second warning period: the second one (on October 1st, 2020) was not as restrictive as in March and based on a 4-color code (not all the regions were impacted similarly), but it was then hardened and expanded to all Quebec with the third lockdown (on December 23rd, 2020), which was announced for the Christmas holidays and extended until February 2021. A fourth lockdown was declared in several regions of Quebec in April 2021 but not in Montreal. The fifth and last lockdown (on December 22, 2022) was enforced for the Christmas holidays because of an increase in cases due to the Omicron variant (third warning period). In addition to the events listed in Table 8.1, access to several regions in Quebec was restricted (with roadside checkpoints) during the first warning period, and the Canadian border stayed closed during almost the whole timeline. Travel to Canada has been very limited (mandatory quarantine, 3-day stay at the hotel, compelling reason only, etc.) until fall 2021, when non-essential travel has been authorized for those who are fully vaccinated and tested negative for COVID-19. This led to a big drop in the number of tourists and newcomers in Montreal, known as an international city (Tourisme Montréal, 2021). Daycare services remained open most of the time (except for some weeks during the first lockdown), whereas universities and colleges have been closed until September 2021. Primary and secondary schools were closed between March 2020 and September 2020 (with online classrooms), but they have been reopened since then (with specific guidelines). Teleworking has been largely adopted in Montreal and, when allowed, the return to the offices was very partial over the entire study period. Furthermore, measures were implemented for transit by the government and the Montreal transit organizations (written in blue in Table 8.1). They are drawn from a timeline created by the Metropolitan regional transportation authority of Montreal (ARTM, 2020).

8.3 Data and methodology

In this section, the different passive datasets used in this paper are presented. The methodological framework of our analyses, based on time series approaches, is also detailed.

Table 8.1 COVID-19 Timeline in Montreal

Date	Event
2020-02-27	First coronavirus case in Quebec.
2020-03-13	Health emergency declared in Quebec. First lockdown. Closure of several public places, including daycare and schools. Implementation of social distancing measures. Closure of the Canadian border.
2020-03-17	Suspension of fare validation on buses.
2020-03-21	Prohibition of all (indoor and outdoor) gatherings.
2020-03-24	Closure of all nonessential businesses.
2020-03-27	Health emergency declared in Montreal. Travel to/from the city is not recommended.
2020-03-30	Closure of all essential businesses on Sundays.
2020-04-01	Suspension of transit passes (but each user can reactivate their pass at any time).
2020-05-11	Full reopening of the construction industry and partial reopening of manufacturing.
2020-05-22	Authorization for outdoor gatherings (max 10 people from 3 households).
2020-05-24	Reopening of essential businesses on Sundays.
2020-05-25	Reopening of stores with exterior access.
2020-05-25	Distribution of free reusable face covers begins.
2020-06-01	Reopening of daycare services.
2020-06-19	Reopening of shopping malls.
2020-06-22	Authorization for on-site restaurant services and indoor gatherings (max 10 people from 3 households).
2020-06-26	Return of fare validation on buses.
2020-06-30	Announcement that masks will become mandatory on public transportation.
2020-07-13	Wearing a mask is mandatory on public transportation.
2020-07-18	Wearing a mask is mandatory everywhere in Quebec in closed public places. Gradual return of the staff in government buildings and private companies (maximum occupancy rate of 25% of workplace capacity).
2020-08-03	The maximum number of people allowed at indoor and outdoor events is increased from 50 to 250 people.
2020-09-01	Creation of contactless bus tickets (new fare available on mobile phone applications as an alternative to cash payment on board buses).
2020-09-08	Establishment of a regional alert system based on a 4-color code: vigilance (green), early warning (yellow), moderate warning (orange) and maximum warning (red).
2020-09-20	Orange level in Montreal. Gathering restrictions.
2020-09-28	Red level in Montreal.
2020-10-01	Second lockdown. Additional measures specific to the red level: prohibited private gatherings, closure of restaurants and bars, etc.
2020-10-08	Reinforcement of sanitary measures in red zones for education, higher education, and sports networks.
2020-10-15	Availability of information on crowding in transit vehicles.
2020-12-14	Vaccination starts in Quebec.
2020-12-17	Update of Holiday's instructions: private gathering (indoor and outdoor) prohibited, but a single person (and their children) may join those residing at only one other address, public outdoor activities allowed (max 8 people).
2020-12-25	Third lockdown. Closure of all nonessential businesses.
2021-01-09	Implementation of a curfew.
2021-02-08	Reopening of nonessential businesses.

Table 8.1 COVID-19 Timeline in Montreal (continued and end)

Date	Event
2021-02-26	Reopening of cinemas, swimming pools, arenas.
2021-05-28	Relaxation begins everywhere in Quebec: end of the curfew, private outside gatherings allowed (max 8 people), opening of outdoor restaurant patios.
2021-06-07	Orange level in Montreal. Reopening of the restaurants' dining rooms.
2021-06-11	Reopening of outdoor bar patios.
2021-06-14	Yellow level in Montreal. Reopening of the bars' indoor rooms.
2021-06-25	Fully vaccinated people can meet in private residences, without masks and without distancing.
2021-06-28	Green level in Montreal.
2021-07-12	Additional relaxation: social distancing lowered from 2 meters to 1 meter, lifting of capacity restrictions in retail stores.
2021-08-23	Back to school: return to class in person for all students, with instructions on wearing masks.
2021-08-27	End to the state of emergency in Montreal.
2021-09-01	Implementation of the vaccine passport allowing access or participation in nonessential activities.
2021-10-08	Relaxation of sanitary measures for theaters, sports and cultural venues and organized public gatherings.
2021-11-01	Relaxation of sanitary measures for restaurants and bars.
2021-11-29	Detection of the Omicron variant in Quebec.
2021-12-20	Fifth lockdown. Tightening of health measures: telecommuting is mandatory, 10 people maximum for private indoor gatherings, reduced capacity to 50% in businesses and malls, closure of elementary and secondary schools, sports activities, cinemas, theaters, bars, etc. Restaurants remain open, with dining rooms limited to 5 a.m. to 10 p.m.
2021-12-26	Lowering the limit to six people (or the occupants of two residences) for private indoor gatherings.
2021-12-31	Private gatherings are prohibited. Implementation of a new curfew. Closure of businesses on Sundays. Closure of restaurant dining rooms. School start date postponed to January 17.

8.3.1 Data description

Several passive data streams are harnessed to depict the use of various modes of transport: tap-in validation data for the subway, count data for cycling and private cars, and reservation data for (free-floating and station-based) carsharing. The respective transportation systems or devices used to collect information on the modes are illustrated in Figure 8.1. Each dataset is described below and spans over three complete years, from January 1st, 2019, to December 31st, 2021. Data were cleaned and aggregated into daily time series, represented in Figure 8.2. COVID-19 information is also displayed to draw a parallel between the use of each mode and the number of hospitalizations. In addition, Table 8.2 compares 2019 average daily use with 2020 and 2021 average daily use of each mode, distinguishing between business and non-business days. Average use is preferred over

total annual use because it accounts for calendar effects (the number of total days, business days and non-business days is not the same between the two years being compared).

8.3.1.1 Tap-in validation data for the subway

Subway data were provided by the Montreal transit authority (STM). They were collected by the OPUS automatic fare collection (AFC) system, which gathers tap-in validations (at boarding only) from smart cards and tickets. The data sample used in this work includes all validations (regardless of the fare support) made in the subway network, composed of 4 lines and 68 stations (see Figure 8.1). During this period, subway ridership dropped significantly due to the COVID-19 pandemic: the average daily number of entries decreased by about 60% compared to 2019. This declined was uniformly distributed between business and non-business days in 2020 (-59.3% for weekdays and -59.8% for weekends and holidays), but non-business days (-54.8%) were ultimately less affected than business days (-60%) in 2021.

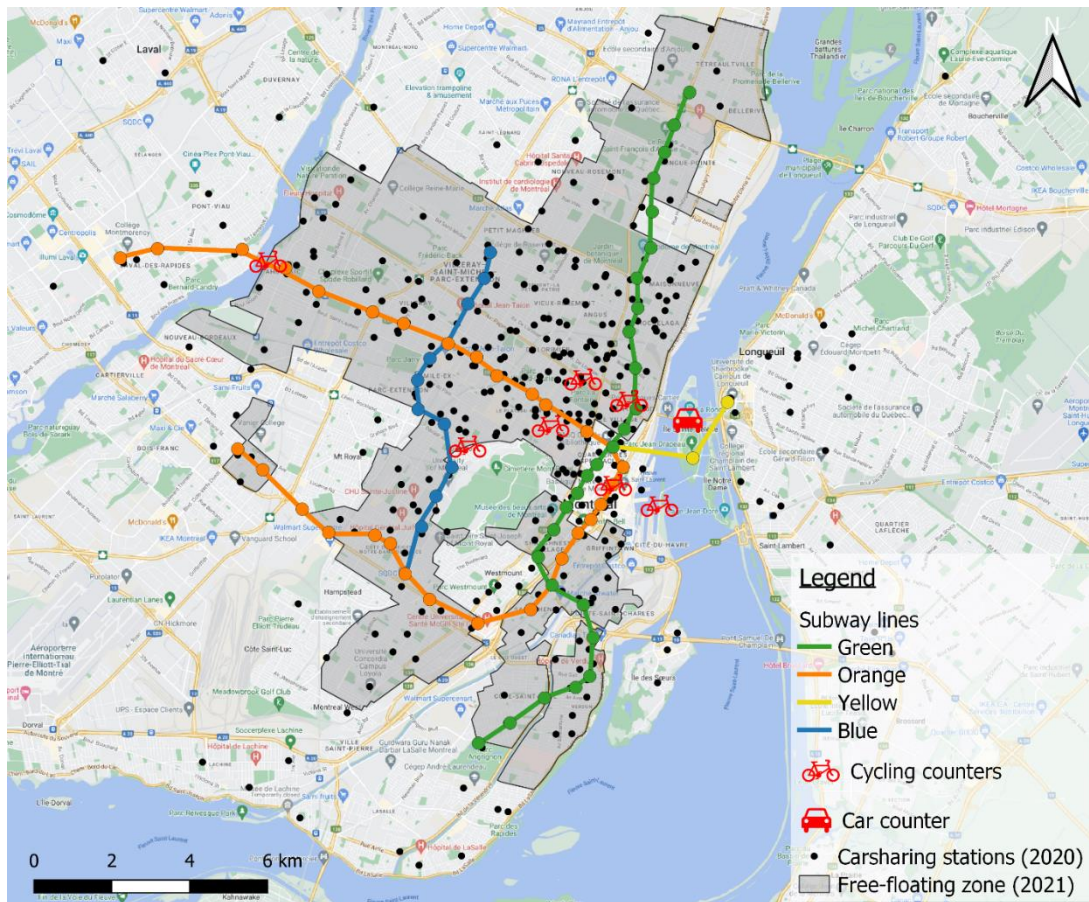


Figure 8.1 Montreal subway network, selected cycling and car counters, free-floating carsharing zone and carsharing stations

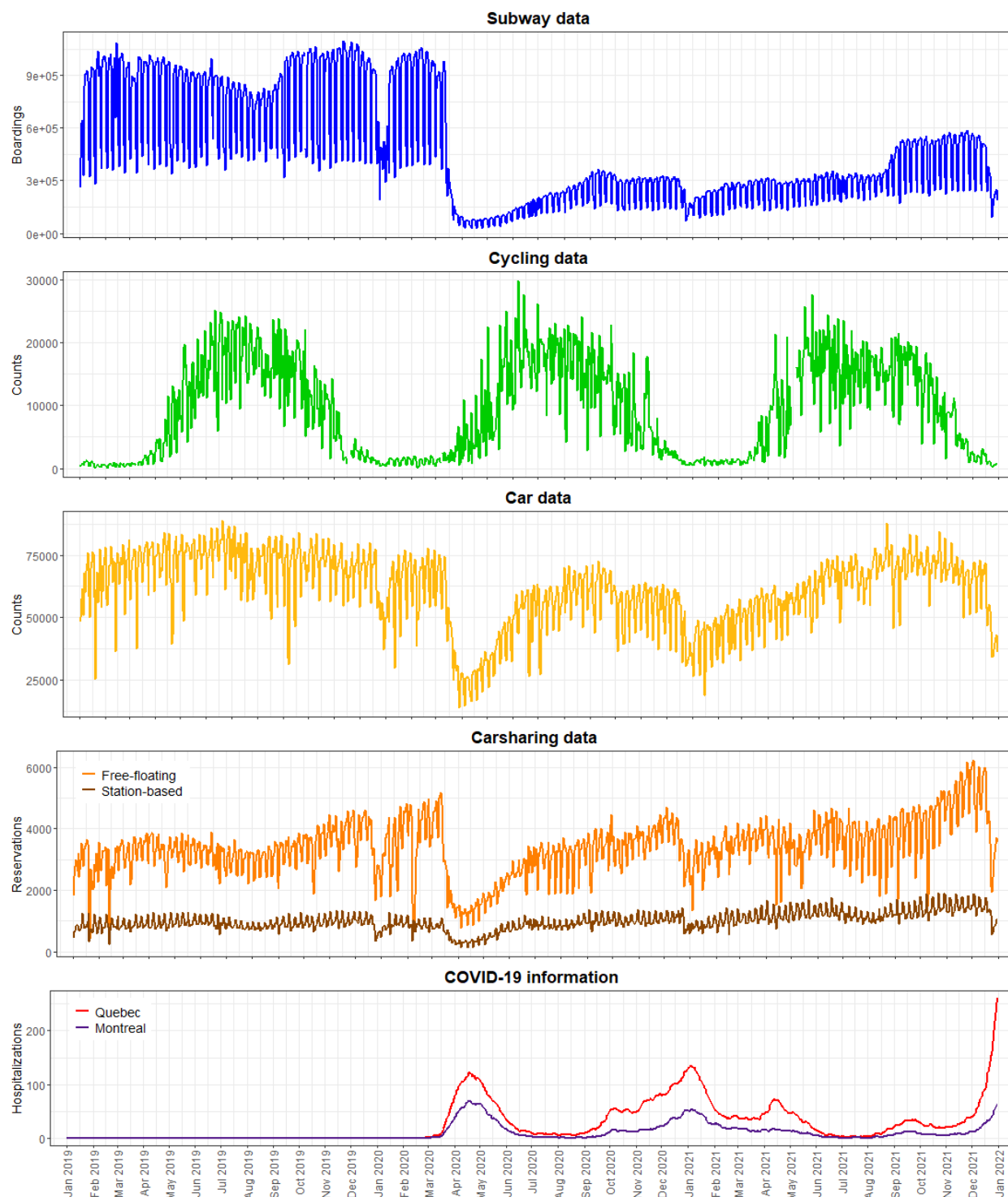


Figure 8.2 Daily time series for subway, bicycle, car, carsharing and COVID-19 hospitalizations

Table 8.2 Comparison of the average daily use of each mode in 2020 and 2021 vs. 2019

Comparison		2020 vs. 2019			2021 vs. 2019		
Mode	Indicator	Total	BD	NBD	Total	BD	NBD
Subway	# entries	-59.3%	-59.3%	-59.8%	-59.0%	-60.0%	-54.8%
Cycling	# counts	-0.4%	-10.9%	+27.9%	+5.0%	-2.7%	+25.5%
Car	# counts	-26.2%	-24.3%	-31.1%	-14.9%	-16.0%	-12.1%
FFCS	# reservations	-5.3%	-3.8%	-9.3%	+17.9%	+19.5%	+13.2%
	Hrs reservation	+37.7%	+44.1%	+28.9%	+75.4%	+75.6%	+75.7%
SBCS	# reservations	-6.9%	-3.5%	-13.6%	+34.3%	+37.6%	+28.0%
	Hrs reservation	+5.2%	+13.4%	-5.7%	+45.8%	+52.4%	+37.4%

Notes: BD = business days (working weekdays), NBD = non-business days (weekends + public holidays), FFCS = free-floating carsharing, SBCS = station-based carsharing, Hrs = hours

8.3.1.2 Count data for cycling

The city of Montreal collects cycling data with several counters from the Eco-counter system. These data include personally owned bicycles but also station-based bikesharing, operated by BIXI in Montreal. In this paper, 7 counters are selected based on the continuity of the information they gathered (the counters with the fewest missing observations are chosen). They are located at different places on the Montreal Island (see Figure 8.1) that were not affected by the “public health corridors” implemented during the pandemic. All counts (both directions of all counters) being summed together per day, almost no change is measured in the average daily use of bicycles between 2019 and 2020 (-0.4%), while an increase of 5% is recorded between 2019 and 2021. However, this increase is only observed on weekends (increase of +25-30% compared to 2019), whereas a decline is revealed in both 2020 and 2021 on weekdays (but in a smaller proportion in 2021 compared to 2020). As discussed in the literature review, cycling may have increased for recreational or physical activities during the pandemic.

8.3.1.3 Count data for private cars

Car data come from traffic counters (loops and radars) on the Jacques-Cartier Bridge, spotted in Figure 8.1 and managed by the Corporation “Jacques Cartier and Champlain Bridges Incorporated”. These counters are positioned at the entrances and exits of the bridges, but also on the different lanes of the bridge. Therefore, several combinations of the counts (linked by simple

additive equations) can be used to calculate the total flow of vehicles on the bridge. These equations were compared, and most differences were lower than 5%, except when the road works or equipment malfunctions resulted in data loss. To compensate missing data present in all counters for different periods, the counts are merged (i.e., the equations which give the total flow of vehicles are interchanged according to the availability of the data) to have continuous information over the whole study period. In addition, as the counted vehicles were classified into different categories by the devices, only cars (without trucks, but they can include shared vehicles) are selected. Based on this information, the average daily number of cars on the bridge in 2020 and 2021 remained lower than in 2019, but this decline was less pronounced in 2021 than in 2020; this suggests that the flow is increasing, and Figure 8.2 shows that it almost reached its 2019 level at the end of the study period. The decrease in car counts was larger on non-business days than on business days in 2020, but again the opposite was observed in 2021.

8.3.1.4 Reservation data for carsharing

Carsharing use is derived from the transactional data of Communauto, the Montreal carsharing company. Data is available for both free-floating (FF) and station-based (SB) services. The former allows one-way trips (without reservations in advance), whereas the latter (with reservations in advance) requires a return to the starting station. The stations and free-floating zone are exhibited in Figure 8.1. A reduction of about 5% for FF carsharing (7% for SB carsharing) in the average daily number of reservations was recorded in 2020 compared to 2019. Therefore, at first glance, despite special hygiene measures, carsharing in Montreal also suffered from the COVID-19 pandemic, especially on weekends. However, this does not consider the reservation duration. In a second step, we calculated the total number of reservation hours per day from the timestamp of the start and end of each reservation. Based on this new indicator, carsharing (both services) usage increased on average, except on weekends for SB carsharing. This growth in total time of use with a decrease in the number of reservations indicates a longer average reservation duration in 2020 than in 2019. The higher increase for FF (compared to SB) carsharing may be explained by a lack of vehicles in stations: indeed, Communauto experienced capacity issues during the pandemic (the old vehicles could not be replaced because of delivery delays). As a result, carsharing members used FF vehicles instead of SB vehicles to make trips which are typically longer. In 2021, an

increase was reported for both services (SB and FF) and both indicators (reservation hours and numbers).

8.3.1.5 COVID-19 information

The daily numbers of hospitalizations due to the coronavirus (more reliable than the cases) are drawn from the INSPQ open data for the Montreal region and the whole province of Quebec. The 7-day moving averages are displayed in Figure 8.2. Five waves are apparent: the first began in March 2020 and ended in July 2020, while the second (longer and larger, with small ups and downs) lasted from September 2020 to March 2021. The third wave spanned from March 2021 to July 2021 in several Quebec regions but did not affect Montreal. The fourth wave (July 2021 – December 2021) was smaller but was followed by a very important fifth wave due to the Omicron variant. The effect of the lockdowns described above on the number of hospitalizations are clearly visible (except for the last one since it was implemented at the end of the study period). In this paper, the question is: did they also impact the subway ridership and the use of other modes, and to what magnitude?

8.3.2 Analysis strategy

Going beyond simple descriptive statistics, an analysis strategy in three steps is proposed:

1. Initially, only the subway data are processed. CPD methods are applied to detect breakpoints objectively and automatically in subway ridership and split the daily time series into several segments (or phases). The breakpoints are then linked to the COVID-19 timeline of Montreal; this allows determining which measures had more impacts on the subway and quantifying the corresponding systemic effects (direction, magnitude, and delay of the impacts). Such an approach is particularly relevant for Montreal since many changes in the rules happened during the pandemic, and not all of them necessarily had similar impacts on the subway ridership. In addition, CPD allows lag effects and interactive effects of a combination of policies (Bian et al., 2021).
2. The variations of the detected impacts are then examined (for the subway only again). Deeper analyses are performed (at a more disaggregate level of time) to check whether the temporal structure of the travel patterns changed between phases. To this end, the magnitude of the impacts is appraised by phase for several subgroups of trips (by day type

and period of the day). Moreover, time-frequency analyses are carried out using the wavelet transform (applied to the hourly time series of boardings) to monitor the usual daily and weekly rhythms over the study period. The average strength of each frequency is calculated by phase and the resulting distributions are compared to the 2019 baseline. Therefore, this second step allows assessing if all times and frequencies were affected with the same magnitude in each phase.

3. Finally, interactions between subway ridership and the use of bicycles, private cars and carsharing services are estimated using ITS models. The goal is to evaluate whether the measures which influenced the subway ridership also had impacts on the use of other modes (cross-analysis of the impacts). Furthermore, correlation between the subway boardings and the cycling counts, car counts or carsharing reservations is measured by phase to figure out whether the measures also affected the relationships between these modes.

Methodological details are provided below for each of these steps.

8.3.2.1 Change point detection (CPD)

CPD aims to identify abrupt changes in longitudinal data based on changes in the time series characteristics. Several types of methods exist (Aminikhanghahi & Cook, 2017). In this paper, structural CPD is used to detect (multiple) pure structural changes in offline data based on a regression structure model (note: other methods based on the likelihood ratio were also tested and gave similar results). This design has been applied in various domains but also in several transportation-related studies, for instance to evaluate the effects of bus policies (Yeh & Lee, 2019) or time lag effects of COVID-19 measures (Bian et al., 2021) on ridership. For k breakpoints, and then $k + 1$ segments (or phases) in which the regression coefficients are constant, the equation of the associated segmented regression is:

$$y_t = \beta_j x_t^\top + \varepsilon_t \quad t = T_{j-1} + 1, \dots, T_j \quad j = 1, \dots, k + 1 \quad (1)$$

where y_t is the outcome (dependent variable), $x_t^\top$ a vector of regressors (independent variables), β_j the corresponding vector of coefficients for segment j and ε_t the error term at time t . $T_1, \dots, T_k$ are the time indexes of the breakpoints, while $T_0 = 0$ and $T_{k+1} = n$ (number of observations). See (Bai & Perron, 2003) for more theoretical details and (Zeileis et al., 2003) for practical applications

using the R statistical software package *strucchange*, that is used in this work. The algorithm, based on the dynamic programming method, computes the dates of the breakpoints (and the corresponding regression coefficients) given a number of breaks. Moreover, a confidence interval is calculated for each breakpoint. The optimal number of points can be determined based on the Bayesian Information Criteria (BIC) criterion. In addition, supF-test allows testing the hypothesis of m against $m + 1$ breakpoints. In our model, the outcome is the subway ridership: we take the logarithm of the daily number of boardings to approximate a normal distribution. There are no regressors (x_t is equal to unity in equation 1), so changes in means of the time series segments are modeled (note: other more complex equations were also tried and gave similar results). In the algorithm, the minimum duration of a phase is set at 14 days. To validate our results, Bayesian CPD is also investigated using the R package *bcp* (Erdman & Emerson, 2007). The method provides posterior probabilities of the predictions which indicate how likely a detected change point is a “true” change point.

8.3.2.2 Wavelet analysis

Wavelet analysis, based on the wavelet transform, is a processing technique widely used for signal decomposition to identify subsignals with certain frequencies in data and monitor their variations over time. Unlike the Fourier transform, which can only detect global frequencies that persist over an entire signal (in other words, the time series structure is assumed stationary), the wavelet transform can capture local spectral and temporal information simultaneously. It is also a better alternative approach to windowed Fourier transform (or short-time Fourier transform), which consists in performing a Fourier transform on different segments of the time series, thus on (too) small subgroups of data. The wavelet transform of a discrete time series signal x_i with equal time spacing δ_t and $i = 0, \dots, N - 1$ is defined as the convolution of x_i with a scaled and translated copy of a mother wavelet function ψ :

$$W(s, t) = \sum_{i=0}^{N-1} x_i \psi * \left[\frac{(i - t)\delta_t}{s} \right] \quad (2)$$

where $(*)$ is the complex conjugate. A wavelet is a wavelike oscillation characterized by its scale s (or dilatation) and location in time t . Basically, the wavelet transform computes how much any wavelet of a particular scale is in a signal by sliding it across the entire signal, i.e., by varying its

time index. The wavelet power spectrum $|W(s, t)|^2$, i.e., the squared magnitude of the wavelet transform, is then estimated to provide the strength of a particular frequency (intrinsically linked to the scale) and show how this strength varies with time. The resulting plot is called a time-frequency plot or a spectrogram. It looks like a heat map where frequency (y-axis) is represented against time (x-axis) and the power of each frequency at each time position is given by a color. Black contour lines can also be drawn to indicate significance. In this paper, we apply the Morlet wavelet (as mother wavelet function) for continuous wavelet transform (CWT), i.e., frequency is continuous, of the subway time series. Validations are disaggregated per day-hour to extract periods (inverse of the frequencies) in hours, for instance the 24-hour period of daily patterns, and their evolution throughout the pandemic is explored. To this end, the R package *dplR* (Bunn et al., 2022), based on the methodological guide of Torrence et Compo (1998), is used.

8.3.2.3 Interrupted time series (ITS) regression with autoregressive errors

ITS regression is useful to evaluate the impact of a population-level intervention (so without any control group) that have been implemented at a defined point in time. When longitudinal data were available before and after the intervention date, many public health interventions were analyzed based on this approach (Bernal et al., 2017). It was also applied in a transportation context to assess the impacts of transit strikes on bikesharing (Fuller et al., 2019) or the effects of specific policies on traffic safety (Ignacio Nazif-Muñoz et al., 2018; Song, Yu & Noyce, 2019). The method is based on the following segmented regression model:

$$y_t = \beta_0 + \beta_1 I_t + \beta_2 x_t + \beta_3 I_t \times x_t + \varepsilon_t \quad (3)$$

where y_t is the outcome variable, I_t a dummy variable for the pre-intervention period (coded 0) or the post-intervention period (coded 1), x_t a variable for which an interaction with the intervention is modeled (x_t is often the time elapsed since the start of the study period, to measure changes in the trend before/after the intervention), β_0 is the baseline level of the outcome, β_1 the change in outcome level after the intervention, β_2 the correlation with x_t before the intervention, and β_3 the change in this correlation after the intervention. More than one intervention implemented at different times can be specified in the model.

In our study, the outcome is the logarithm of the daily mode ridership (cycling counts or car counts or carsharing reservations) from January 1, 2020 to December 31, 2021 (2019 is kept to control for

seasonality; see below). The dates of the interventions correspond to the breakpoints (1/0 = before/after the breakpoint) and x_t is the logarithm of the subway ridership. In this way, changes in levels (not in the trend since the study period is too short to estimate a trend) and changes in the interaction with the subway in each phase are assessed. These changes are linked to previously detected change points instead of a priori COVID-19 measures. This means we first checked whether and when changes have occurred using CPD instead of introducing all events of Table 8.1 in the model. The associated equation is:

$$\begin{aligned} \log(M_t) = & \beta_0 + \beta_1 BP_1 + \dots + \beta_k BP_k + \beta_{k+1} \log(S_t) \\ & + \beta_{k+2} BP_1 \times \log(S_t) + \dots + \beta_{2k+1} BP_k \times \log(S_t) \\ & + \sum C_t + \varepsilon_t \end{aligned} \quad (4)$$

where β_j ($j = 1, \dots, k$) is the change in level following the j^{th} breakpoint BP_j , β_{k+1} estimates the correlation between the mode use (M_t) and the subway ridership (S_t) before the COVID-19 pandemic and β_{k+j+1} the change in this correlation following the j^{th} breakpoint BP_j (or during the phase $j + 1$). Additional independent variables (C_t) are introduced in the equation to control for other important effects: day type (1 for weekdays, 0 for weekends and holidays), weather conditions (average daily temperature and daily total rainfalls), and annual seasonality. The latter is considered by calculating the logarithm of the 7-day moving average of the 2019 daily mode ridership for the same day of the year (i.e., January 1, 2020 and January 1, 2021 are regressed against January 1, 2019).

The full linear regression is first trained (with all breakpoints and all control variables), then stepwise selection based on the BIC criterion is carried out to keep only the most important variables in the equation. This results in a more parsimonious model. Furthermore, as the error term ε_t may be correlated to error terms at p preceding time points (autocorrelation of order p), the coefficients of the proposed ITS model are corrected by estimating the final regression equation with autoregressive errors:

$$\varepsilon_t = \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + \dots + \varphi_p \varepsilon_{t-p} + \varepsilon_t \text{ with } \varepsilon_t \sim N(0, \sigma^2) \quad (5)$$

where φ_i counts for the correlation between the error term at time t and its i^{th} previous error term. p is determined by autocorrelation function (ACF) and partial autocorrelation function (PACF)

plots of the error terms in the regression model for each mode. The final ITS regression model with autoregressive errors is computed using the *gls* function of the R package *nlme* (Fox & Weisberg, 2019).

8.4 Results

Results are decomposed in three parts. First, the breakpoints detected in the subway ridership are qualified and the impacts of the COVID-19-related measures are quantified. Then, the temporal variability of these impacts is analyzed by phase. Finally, the changes in interactions between subway and other transportation modes are modeled.

8.4.1 Breakpoints in the subway ridership and magnitude of the impacts

Several segmentations of the regression structure model presented above were tested and compared based on the BIC criterion and supF tests. The results are documented in Table 8.3 with the corresponding breakpoint time indexes (time elapsed since January 1st, 2019). According to the minimum BIC value, the optimal number of breakpoints is 11. However, supF tests show that it is not worth adding a new breakpoint from 7 (the p-value then becomes greater than 0.05). In addition, the 10th and 11th breakpoints found by the method are the beginning and the end of the 2019 Christmas holidays period, whose effect is not of interest in this paper, while the 8th and 9th may indeed be linked to COVID-19-related events. Therefore, 9 breakpoints (and therefore 10 phases) are chosen in the following analyses. They are dated in Figure 8.3 (top plot). It is worth noting that no breakpoints are identified in 2019, hence the method succeeds in measuring changes due to the COVID-19 crisis. The corresponding 95% confidence intervals (in red), as well as the 10 phases and the associated average daily subway ridership (in blue), are also exhibited in Figure 8.3. As the intervals are narrow, the predictions are accurate, except maybe for the start date of phase 6 (breakpoint #5). Furthermore, the bottom of Figure 8.3 provides the results of Bayesian CPD. The posterior probability of a change point occurring (calculated for every observation point in time) increases for the 9 detected breakpoints, hence this confirms the quality of our predictions.

Table 8.3 Test results for different number of breakpoints k

k	Breakpoints time indexes (days since January 1 st , 2019)	BIC	supF tests	
			Test	p-value
1	437	1651.3	0 versus 1	0.000
2	442,544	1159.4	1 versus 2	0.000
3	442,530,950	993.6	2 versus 3	0.000
4	443,509,558,964	954.8	3 versus 4	0.000
5	437,451,509,558,964	920.8	4 versus 5	0.000
6	437,451,509,558,964,1081	897.4	5 versus 6	0.011
7	437,451,509,551,874,979,1081	884.6	6 versus 7	0.017
8	437,451,509,558,722,768,964,1081	862.5	7 versus 8	0.115
9	437,451,509,544,579,722,768,964,1081	859.8	8 versus 9	0.149
10	353,369,437,451,509,558,722,768,964,1081	845.1	9 versus 10	0.205
11	353,369,437,451,509,544,579,722,768,964,1081	841.7	10 versus 11	0.258
12	353,369,437,451,509,544,579,722,761,880,979,1081	842.6	11 versus 12	0.324
13	245,353,369,437,451,509,544,579,722,761,880,979,1081	845.5	12 versus 13	0.402
14	245,353,369,437,451,509,544,579,722,740,768,880,979,1081	850.1	13 versus 14	0.484
15	14,251,353,369,437,451,509,544,579,722,740,768,880,979,1081	857.4	14 versus 15	0.574

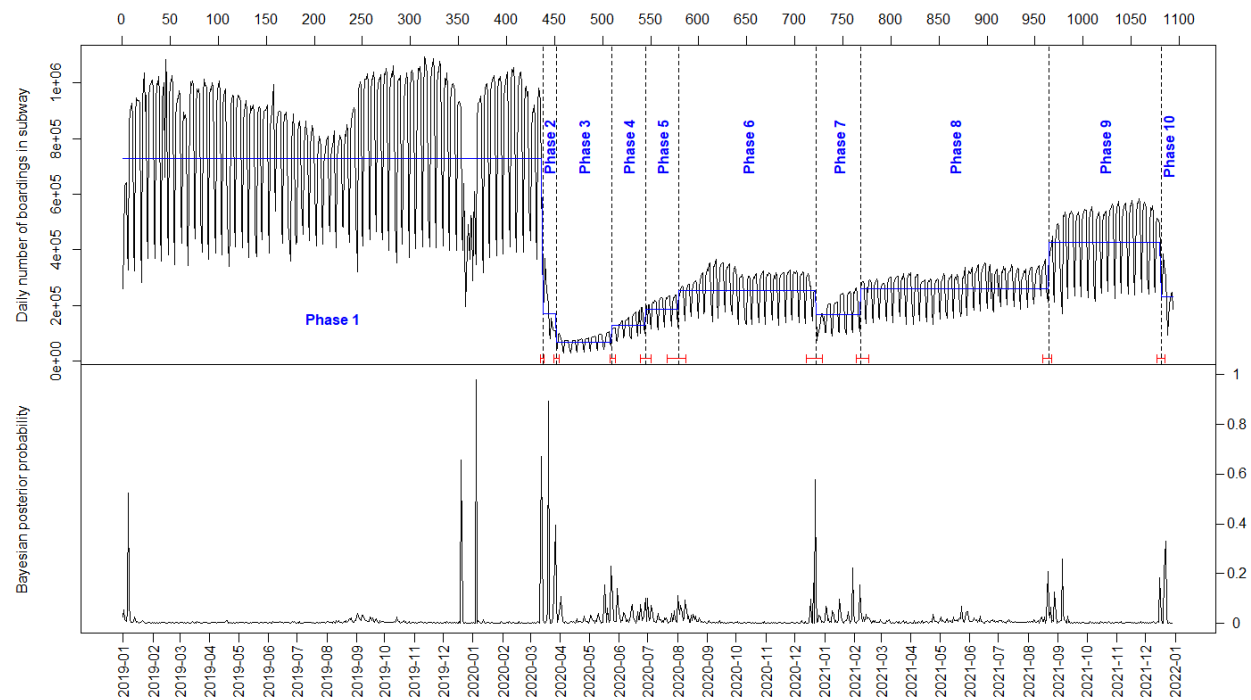


Figure 8.3 Structural change point detection results and associated Bayesian posterior probabilities

Table 8.4 presents the dates of the detected breakpoints, the COVID-19 measures to which they are potentially linked, and the magnitude of their impacts. The magnitude of impact is calculated as the percentage change in average daily ridership from the pre-COVID phase 1, but also between the two sequential phases separated by the breakpoint. The first two breakpoints correspond to the declaration dates of health emergency in the Quebec Province (start of the first lockdown) and in

the Montreal City. Both triggered important reductions in average subway use, but most of the decline in ridership occurred in the short second phase before stabilizing in phase 3. In the end, the average decrease between the pre-COVID phase 1 and the beginning of the first wave of the pandemic (phase 3, breakpoint #2) reached -91%. The second lockdown, introduced on October 1st, 2020, does not appear to have resulted in a significant effect. A small decline in ridership is visible in Figure 8.3, but is compensated (in the mean) by the increasing trend at the beginning of phase 6. In contrast, the third and fifth lockdowns (the fourth lockdown was not applied in Montreal), announced for the 2020 and 2021 Christmas holidays respectively, caused a significant decrease in ridership (see breakpoints #6 and #9), but to a lesser extent than for the first one: the average number of boardings in phases 7 and 10 was indeed not as low as during phase 3 (see the means in Figure 8.3). Ridership dropped by 77% and 68% compared to pre-COVID phase 1, thus the negative effect of the lockdowns is decreasing. This could be because people were more “used” to the risk and then did not react as drastically as for the first lockdown, or because the rules were less strict. One could remark that no time lag is observed between the lockdown implementations and their impact on subway ridership, indicating immediate effect: this was also observed by Shakibaei et al. (2021) for transit in New York City.

Table 8.4 Magnitude of the impacts of COVID-19-related events based on estimated breakpoint dates

#	Breakpoint date	Linked event(s)	Impact	
			phase <i>i</i> vs. phase 1	phase <i>i</i> vs. phase <i>i-1</i>
1	2020-03-13 (Friday)	Health emergency declared in Quebec. First lockdown.	- 75.6%	- 75.6%
2	2020-03-27 (Friday)	Health emergency declared in Montreal.	- 91.0%	- 63.1%
3	2020-05-24 (Sunday)	Distribution of free reusable face covers begins on 2020-05-25. Progressive reopening of several types of businesses. Authorization for outdoor gatherings (with conditions).	- 83.1%	+ 87.9%
4	2020-06-28 (Sunday)	Return of fare validation on buses on 2020-06-26. Authorization for on-site restaurant services and indoor gatherings.	- 75.4%	+ 45.5%
5	2020-08-02 (Sunday)	Summer holidays. Relaxed conditions for the events. Gradual return to the office (maximum 25% of workplace capacity).	- 66.0%	+ 38.1%
6	2020-12-23 (Wednesday)	Third lockdown. Closure of all nonessential businesses during Christmas holidays.	- 77.0%	- 32.5%
7	2021-02-07 (Sunday)	Reopening of nonessential businesses.	- 65.3%	+ 51.0%
8	2021-08-22 (Sunday)	Back to school in person.	- 42.4%	+ 66.0%
9	2021-12-17 (Friday)	Fifth lockdown. Tightening of the health measures during Christmas holidays.	- 68.4%	- 45.2%

However, this is not the case for the release measures at the origin of the recovery phases after the first lockdown, meaning their influence was probably more gradual. Breakpoints #3, #4 and #5 may be linked to several events, or a combination of them. Nevertheless, two of them (#3 and #4) may be, at least partially, due to specific measures implemented for transit (in blue in Table 8.4). Although the increase in daily number of boardings began at the end of phase 3, as there were several relaxations in May 2021, the recovery is only significant from phase 4 onwards. The start date of this phase (breakpoint #3) coincides with the start of the distribution of free reusable face covers, which could have helped people regain trust in the subway. Such a result proves the efficacy of the measure and could be of particular interest for practitioners. Moreover, breakpoint #4 is just after the return of fare validation on buses (interrupted during the first lockdown to avoid human contacts). This suggests that subway users (where validation was still required) may have shifted to buses to avoid a controlled validation, and then gone back to subway when fare supports had to be validated again on buses. Breakpoint #5 cannot be linked to any specific event, which could explain why its date is associated with the widest confidence interval. However, phase 6 corresponds to the summer and fall of 2020, when many measures were relaxed and gradual return to the office was permitted (up to 25% of the workplace capacity). Unlike the precedent three breakpoints (3, 4 and 5), breakpoints #8 and #9 are clearly identifiable: they correspond to the date when nonessential businesses reopened after the third lockdown, and to the date when all students (including graduate students) went back to school in person in September 2021, respectively. Both had a quick positive effect on subway use. An interesting thing to notice is that all these five breakpoints (with a positive impact, i.e., increase in subway ridership) fall on Sundays; thus, the effects were effective on Mondays. Furthermore, despite the various moments of crisis that interrupted the recovery, ridership has continued to increase and to approach its pre-COVID level (the difference in each recovery phase compared to phase 1 is lower and lower), and it peaked in the fall of 2021, i.e., in phase 9 when the school year started.

8.4.2 Variations of the impacts by phase

Based on the previous breakpoints, deeper analyses are carried out by phase to differentiate the impacts over time. The magnitude of the impacts (in average daily ridership) is assessed by day type and for different periods of the day in Table 8.5. Business and non-business days (i.e., weekdays versus weekends and public holidays) are separated, as well as five time periods

(morning, morning peak, day, afternoon peak, evening, and night) defined in Table 8.5. According to the results, business days were more impacted than non-business days at the first breakpoint (phase 2), whereas the reverse is true at the second breakpoint (phase 3). This indicates that non-discretionary activities declined before recreational activities. In these first two phases, subway boardings strongly decreased on all day periods, especially (morning and afternoon) peak hours and evening. The former can be explained by the lower number of commuting trips, while the latter is due to the closure of restaurants and bars. Rather, early morning and night were the least affected periods. This is probably related to the schedules of essential workers (for instance, in the food supply or health care sectors) who remained active during the lockdown. Non-business days, as well as day and evening periods (off-peak periods), suggesting leisure travel, were the greatest winners of the following two recovery phases (phases 4 and 5, or breakpoints #3 and #4). However, subway ridership increased on business days, especially in the morning peak period, after the fifth breakpoint (in phase 6). This may be due to the gradual return (up to 25% of workplace capacity) in government buildings and private companies during summer 2020. A decline in travel was observed in the evening after the sixth breakpoint (in phase 7) because of the implementation of a curfew. In the fourth recovery phase (phase 8), non-business days, as well as day and evening periods (off-peak periods), had again the highest increases in ridership. However, subway use increased on business days and peak periods in phase 9 with the start of the school year, before being again the most affected times by the fifth lockdown in phase 10. Phase 9 can be considered as the period closest to a certain “normality”. Nevertheless, the balance (phase 9 compared to pre-COVID phase 1) shows that the pandemic ultimately had more impacts on business days than non-business days, with an average reduction of about 43% against 41% of the daily boardings. Regarding the periods of the day, the highest decreases were observed on the peak hours and evening, whereas ridership during the off-peak hours in the early morning and during the day decreased less. Therefore, the times when transit was initially most frequent were the most affected. This can have impacts on the planning of future transit services.

Table 8.5 Magnitude of the impacts (differences in average daily ridership between pairs of phases) by type of day and period of the day

Phase i/j*	All days	TYPE OF DAY		PERIOD OF THE DAY (business days only)					
		Business days	Non- business days	Morning 4:00 → 6:00	Morning peak 6:00 → 9:00	Day 9:00 → 15:00	Afternoon peak 15:00 → 18:00	Evening 18:00 → 23:00	Night 23:00 → 4:00
2/1	-75.6%	-78.6%	-61.8%	-48.4%	-78.7%	-77.5%	-79.5%	-81.6%	-68.0%
3/2	-63.1%	-57.5%	-73.4%	-41.8%	-58.1%	-57.8%	-59.3%	-58.5%	-43.5%
4/3	87.9%	77.9%	104.0%	44.6%	64.1%	86.9%	82.4%	91.1%	44.8%
5/4	45.5%	45.1%	47.0%	11.9%	32.4%	48.5%	45.1%	61.3%	44.6%
6/5	38.1%	40.5%	22.7%	20.8%	62.3%	35.9%	51.5%	19.5%	3.0%
7/6	-32.5%	-29.0%	-35.4%	-12.3%	-25.0%	-26.5%	-27.4%	-44.4%	-31.8%
8/7	51.0%	41.5%	75.4%	14.0%	30.7%	40.9%	37.2%	79.3%	43.4%
9/8	66.0%	70.7%	40.1%	8.2%	71.3%	69.6%	68.2%	86.8%	57.3%
9/1	-42.4%	-43.4%	-40.6%	-36.4%	-47.1%	-37.2%	-44.0%	-47.8%	-39.9%
10/9	-45.2%	-49.2%	-26.1%	-37.4%	-64.9%	-37.4%	-49.7%	-50.7%	-45.0%

*Phase i is compared to j

The spectrogram resulting from the application of the wavelet transform on the hourly time series of the subway tap-in validations is provided in Figure 8.4. The power spectrum heat map is plotted below the raw data (number of boardings per day-hour). A color is assigned according to the power of each period, i.e., how much the period (inverse of the frequency, converted in days) is observed in the data signal at a particular time. The color scale is cut by the 25th, 50th, 75th, and 95th percentile values of power. Purple is the color of the weakest periods while red is the color of the strongest periods. Moreover, the significant powers are surrounded by black lines. The analysis highlights the usual seasonal patterns in the pre-COVID phase, namely the 8-hour (1/3-day), 12-hour (1/2-day), 24-hour (1-day) and 168-hour (7-day) periods, which reproduce the 8-hour and 12-hour work shifts, as well as the daily and weekly rhythms. A 3.5-day period is also apparent along the year (except in summer); this may be because the hourly distribution of trips on Wednesdays is closer to that of weekend days. The power of all these periods varied a lot across the COVID-19 pandemic. No significant periodicity was observed in the data during all phases of declining ridership. The daily and weekly rhythms resumed in the summer of 2020 (phase 6, when gradual return to the office was allowed) and in the fall of 2021 (phase 9, after the start of the school year), but the 3.5-day and intraday periods never became significant again. In addition, the spectral density functions per phase are illustrated in Figure 8.5 to evaluate how much each period was

observed during the different phases. These distributions were produced by averaging the power of each period across all days of each phase. Each phase is associated with a colored line, to be compared with the pre-COVID phase 1 (dotted dark line). The outcome demonstrates that all periods are much less important in all phases compared to pre-COVID phase 1, especially in phases 3 and 4 during the first wave of the pandemic. It is in phase 9 that the strongest usual rhythms are measured. However, the average power of the daily periodicity is higher than the average power of the weekly periodicity, while the opposite was observed in the pre-COVID phase 1. This means that travel is more similar from one day to another than from one week to another. In other words, the structure of the hourly distribution of the boardings has remained more analogous across the day than across the week. This may be due to the impact of teleworking some days per week (the days selected to go to work changing from one week to another) on the weekly seasonality. Therefore, this confirms that the pandemic had strong impacts on the subway user behavior, at least at the systemic level.

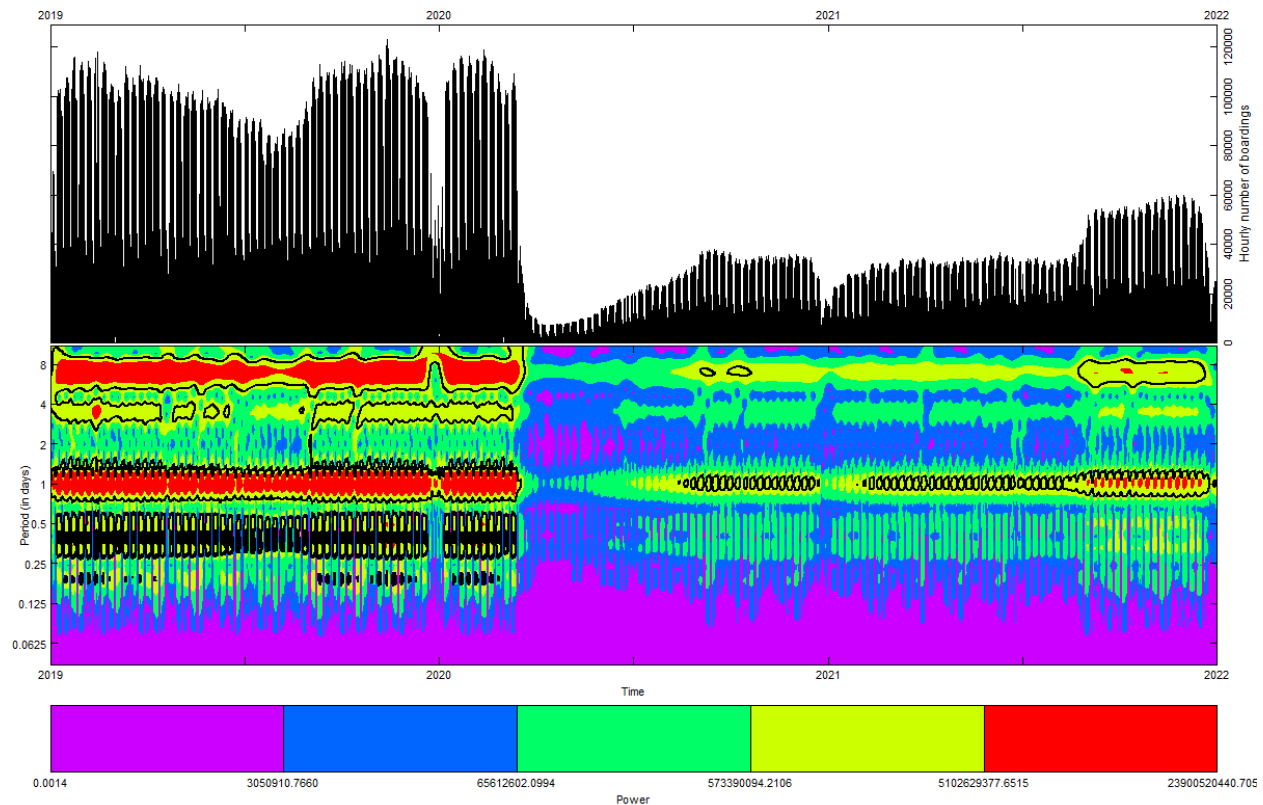


Figure 8.4 Wavelet power spectrum of the hourly subway time series

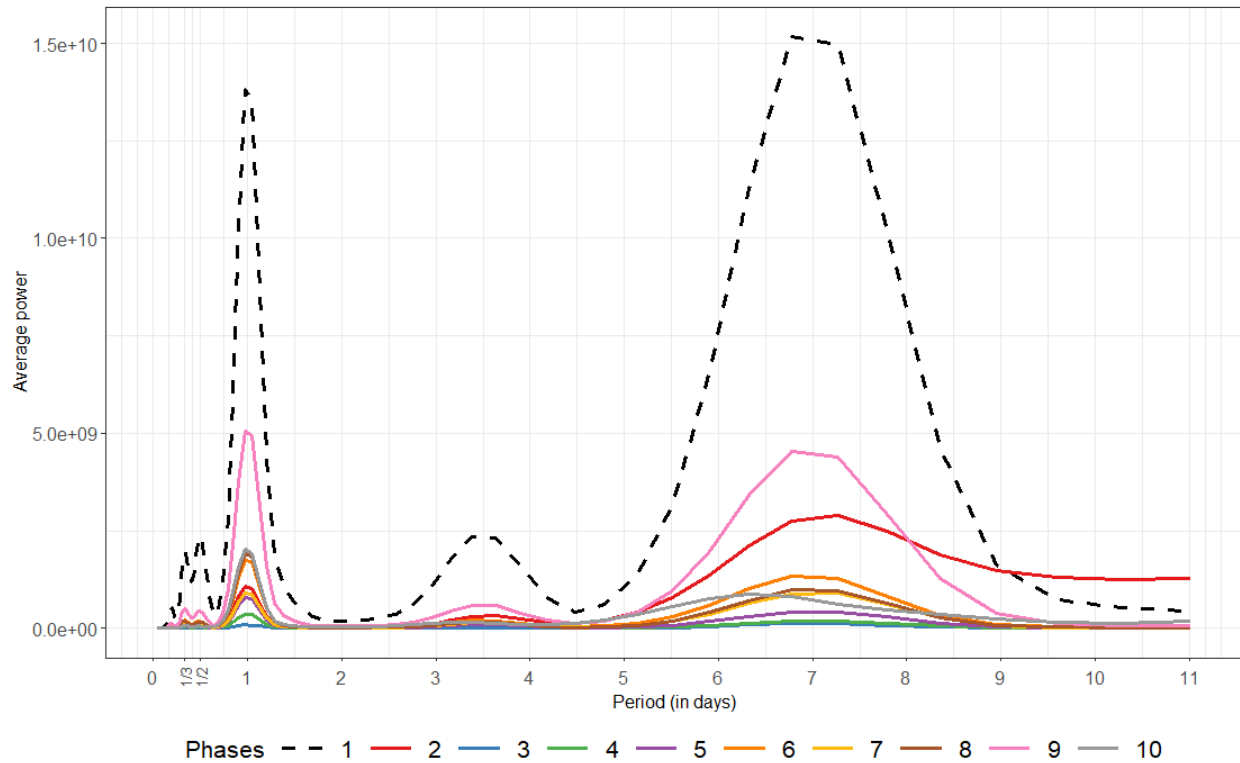


Figure 8.5 Spectral density function per phase

8.4.3 Interactions with other transportation modes

Finally, the ITS models with autoregressive errors trained by mode are provided in Table 8.6. The model for the subway is also reported, but without interaction terms. Only the variables chosen by the stepwise selection method (based on the BIC criterion) applied to the full ITS regression equations were included in the final models. The variables that were not selected are marked with a dash. An autocorrelation of order 7 was found in the errors of all regression models, which can be explained by weekly seasonality. We observed that adding these autocorrelated error terms in the final models mainly changed the significance level of the effects (i.e., the coefficients of certain variables were no more significant after correcting for autocorrelation), but not the direction (i.e., the coefficients remained positive or negative). Therefore, autocorrelation should be considered when applying the ITS method to better appreciate the importance of the impacts. All models have high pseudo R^2 , indicating good fit, but the carsharing models are less satisfying. Nevertheless, the indicator of use in time (hours of reservation) gave better results than the number of reservations for both services (FF and SB).

Table 8.6 Interrupted time series models by transportation mode

MODE	SUBWAY		CYCLING		CAR		FF CARSHARING		SB CARSHARING	
Dependent variable (indicator of use)	Log(# boardings)		Log(# counts)		Log(# counts)		Log(# hours)		Log(# hours)	
Independent variables	Coeff	P-value	Coeff	P-value	Coeff	P-value	Coeff	P-value	Coeff	P-value
Intercept	6.885	0.000 ***	-14.977	0.000 ***	4.747	0.000 ***	3.059	0.001 **	0.125	0.881
Log(7-MA of daily mode use in 2019)	0.452	0.000 ***	0.492	0.000 ***	-	-	0.373	0.000 ***	1.064	0.000 ***
Day type	0.588	0.000 ***	-0.369	0.000 ***	-0.111	0.000 ***	-0.471	0.000 ***	-0.390	0.000 ***
Average temperature (°C)	0.001	0.623	0.038	0.000 ***	0.005	0.000 ***	0.007	0.000 ***	-	-
Total rainfalls (mm)	-0.003	0.000 ***	-0.046	0.000 ***	-0.004	0.000 ***	-	-	-	-
Phase 2	-1.335	0.000 ***	17.915	0.000 ***	0.167	0.010 **	-	-	-6.270	0.000 ***
Phase 3	-0.774	0.000 ***	-	-	-2.394	0.000 ***	-1.689	0.019 *	0.240	0.008 **
Phase 4	0.431	0.000 ***	-	-	1.156	0.016 *	-	-	0.216	0.002 **
Phase 5	0.400	0.000 ***	-6.028	0.000 ***	-0.151	0.000 ***	0.014	0.888	-0.004	0.954
Phase 6	0.242	0.000 ***	-	-	-0.074	0.032 *	-0.323	0.000 ***	4.442	0.000 ***
Phase 7	-0.230	0.000 ***	-	-	-	-	0.123	0.026 *	0.215	0.000 ***
Phase 8	0.336	0.000 ***	-	-	-	-	-	-	-	-
Phase 9	0.362	0.000 ***	-5.017	0.001 ***	4.001	0.000 ***	4.232	0.000 ***	-	-
Phase 10	-0.127	0.162	-	-	-0.193	0.000 ***	0.100	0.445	-	-
Log(subway ridership)	Not applicable		1.428	0.000 ***	0.475	0.000 ***	0.205	0.000 ***	-0.028	0.591
Phase 2 x log(subway ridership)			-1.301	0.000 ***	-	-	-	-	0.482	0.000 ***
Phase 3 x log(subway ridership)			-	-	0.225	0.000 ***	0.172	0.003 **	-	-
Phase 4 x log(subway ridership)			-	-	-0.096	0.026 *	-	-	-	-
Phase 5 x log(subway ridership)			0.482	0.000 ***	-	-	-	-	-	-
Phase 6 x log(subway ridership)			-	-	-	-	-	-	-0.371	0.000 ***
Phase 7 x log(subway ridership)			-	-	-	-	-	-	-	-
Phase 8 x log(subway ridership)			-	-	-	-	-	-	-	-
Phase 9 x log(subway ridership)			0.367	0.002 **	-0.318	0.000 ***	-0.320	0.000 ***	-	-
Phase 10 x log(subway ridership)			-	-	-	-	-	-	-	-
Autoregressive errors										
ε_{t-1}	0.43		0.36		0.57		0.88		0.96	
ε_{t-2}	0.27		0.12		0.18		0.46		0.69	
ε_{t-3}	0.27		0.12		0.12		0.22		0.39	
ε_{t-4}	0.23		0.12		0.08		0.08		0.11	
ε_{t-5}	0.16		0.08		0.06		0.08		-0.12	
ε_{t-6}	0.21		0.14		0.07		0.15		-0.15	
ε_{t-7}	0.41		0.15		0.13		0.20		0.12	
Goodness-of-fit										
Pseudo R ² (Cox and Snell)	0.80		0.71		0.69		0.40		0.64	

Notes: *** 0.1% significance level ($p < 0.001$), ** 0.1% significance level ($p < 0.01$), * 0.5% significance level ($p < 0.05$). Otherwise, the variables are not significant. “-” indicates variables that were not selected in the stepwise model selection.

We will focus on the impacts of the COVID-19 pandemic on each mode and their interaction (by phase), but before doing so, we can quickly comment on the coefficients of the control variables. The 7-day moving average of the 2019 daily mode use was kept in the stepwise variable selection for all modes except car. This means that there is strong annual seasonality in the use of these modes, and thus it is important to control for it. Day type is also present in all models, with a positive impact on subway boardings but a negative impact for all other modes. Therefore, subway daily use tends to be higher on weekdays than on weekends and holidays, while the opposite is true for the other modes. The weather impacts (positive impact of temperature and negative impact of rainfalls) are not significant for all modes, and when they are, the corresponding coefficients are very small. This may be because they are partly captured in the annual seasonality.

The direct impacts of the breakpoints (and associated measures) that were previously identified on the use of each mode are evaluated in the models thanks to binary variables that indicate the different phases (coded 0 before, and 1 after the breakpoint). In addition, their impacts on the systemic interactions between the subway and the other modes are appraised by correlation terms added in the models. A positive correlation states that when subway use increases over time, use of the other mode tends to increase as well. This is a kind of global complementarity, but further analysis would be required to specify the relationships between the two modes. For instance, they may be used in the same trip to solve the first-mile/last-mile problem, i.e., one mode is used to reach another (intermodality), or they may be combined by individuals over time (multimodality), for different activities (Deschaintres et al., 2021).

Obviously, only the direct impacts are considered in the subway model (we cannot measure interaction between the subway and itself). The results were expected: as before, four declining phases (with negative coefficients) and five recovery phases (with positive coefficients) are obtained. However, the coefficient of the 10th phase (fifth lockdown announced for the 2021 Christmas holidays) is not significant. This may be due to the low number of observations in the last phase (because of data unavailability), or the decrease in subway ridership during the Christmas holidays may be captured in the annual seasonality and the negative effect of the lockdown may actually be negligible. Information over the first days of 2022 is needed to better assess the real impact of this breakpoint.

Only three breakpoints had significant effects on cycling: breakpoints #1 (first lockdown), #4 (summer 2020, return of fare validation on buses) and #8 (fall 2021, back to school in person), or phases 2, 5 and 9. The first had a positive impact on bicycle counts, whereas the other two had a negative impact. Therefore, all these impacts have the opposite direction on subway use and on cycling: when subway ridership decreased, the use of bicycles increased and vice versa. This may suggest potential modal shifts from the subway to cycling. This is confirmed by the evolution of the correlation term between the two modes. The effects of each phase on cycling (direct impact) and on the interaction with the subway also have opposite signs. For instance, the first lockdown (phase 2) had a positive impact on cycling, but a negative impact on the interaction with the subway. The positive correlation with the subway in the pre-COVID phase is almost canceled when we add the negative coefficient of phase 2. This correlation increased again afterwards (in phases 5 and 9) but did not reach its pre-pandemic value (the sum of all coefficients of the interaction terms is still negative). Therefore, subway and bicycle daily usage were less positively correlated in 2020 and 2021 than in 2019.

Three phases (2, 5 and 6) in 2020 also had opposite impacts on subway and car use. These three phases correspond to periods when subway boardings were primarily affected on weekdays. Therefore, it is when the use of the subway decreased (or increased) the most on business days that car counts significantly increased (or decreased). This could indicate a substitution for commuting trips, but the correlation with the subway did not change during these phases. The effects in phases 3 and 4 are in the same direction for the subway and the car, as well as in phases 9 and 10. The two later impacts reveal a strong increase in car use when the school year started, but a decrease after the 2021 Christmas lockdown. Again, they are linked to events that mainly affected commuting. Although the impacts on the two modes had the same direction, a decline is observed in the correlation between subway boardings and car counts during these phases. This means that the increase (or decrease) in use of the two modes is less dependent or less similar (for instance, the daily variations or the magnitude of the increase/decrease may be different).

A negative impact on the use (in time) of the two carsharing services was first reported at the beginning of the pandemic, in phase 3 for FF and 2 for SB. However, the hours of reservation increased in all following significant phases (except phase 6 for FF). This delayed growth in carsharing use may be due to the registration period for new enrollment and/or to the inter-regional travel restrictions in place at the beginning of the pandemic (carsharing, especially SB, is used a

lot to get out of Montreal). The highest increase was observed in phase 6 (summer 2020) for SB versus in phase 9 (fall 2021, back to school in person) for FF. This refers to a different usage of the two services, the former being more used for recreative activities and the later for commuting. Regarding the correlation with the subway, it first increased (in phase 3 for FF and phase 2 for SB). Therefore, the decline in carsharing use was strongly correlated with the decline in subway use in the first phases of the pandemic. However, this correlation was reduced later, in phase 9 for FF and phase 6 for SB, i.e., the phases when the use of the two services increased the most. Subway ridership also increased during these phases, but in a less complementary way with carsharing.

8.5 Discussion

Based on several passive data streams available over three complete years (from 2019 to 2021) and processed into daily time series, this paper evaluated the impacts of COVID-19-related measures on the subway and its correlations in use with cycling, car and carsharing services. Change points were first detected in subway tap-in validations using structural and Bayesian CPD methods. Four of them occurred at the same time as the health emergency announcements and/or at the beginning of the three main lockdowns in Montreal. Such measures had a negative and immediate (without lag) effect on subway use, but the magnitude of the impact decreased over time, indicating a better management of the risk perception among transit passengers. Consequently, the average daily subway ridership steadily increased in the intermediate recovery phases. The impacts of several release measures were first combined, leading to gradual increase in subway boardings (in three phases). Two quick recovery phases were also due to important governmental relaxations, including the return to school in person. In addition, two transit-specific measures may have facilitated the increase of subway use: the free distribution of masks, which then had the expected effect (i.e., to improve trust in transit hygiene), and the return of fare validation in buses (suspended to avoid contacts during the first lockdown). The latter impact was not necessarily desired but is similar to the impact of a free transit policy. Such a measure could have temporally attracted new passengers in buses, from subway but maybe also from other modes. This could be further examined to investigate the effects of free transit. However, the main limitation of these first analyses (breakpoint detection) is that it is impossible to isolate the effect of each measure when several were implemented at approximately the same time. The share of decrease or increase attributable to one particular measure among others cannot be identified.

The impacts by phase (between each pair of breakpoints) were then examined by day type (business versus non-business days), time period (peak versus off-peak periods), and hour to explore variations over time. To this end, descriptive statistics by trip subsets were calculated, and a spectrogram in the time and frequency dimensions was plotted after applying the wavelet transform. Overall, the highest decreases in subway boardings were observed on business days and peak hours, as well as in the evening. This is probably due to the adoption of teleworking and fewer activities after work, but the results should be combined with travel purposes and job types to validate this justification. The traditional work shifts, daily and weekly patterns, characterized by 8-hour, 12-hour, 24-hour and 7-day periods, were also strongly affected. The daily rhythms in subway boardings (defined at the hourly level) reappeared the most in late 2021 (before the fifth lockdown). Indeed, although subway ridership decreased in absolute during these periods, the morning and afternoon peak hours have been a little flattened but still stand out on weekdays (based on the hourly distribution of the subway boardings, available upon request). On the contrary, the weekly regularity is much weaker than before the pandemic, probably because days were chosen randomly to go to work in person. In summary, the time and frequency scales most affected are those that were most responsive to 5-day-a-week commuting needs. This outcome can question the role of transit in a post-pandemic world. Indeed, transit planning and service operation have long focused on commuting during the peak hours of an average weekday, but with teleworking likely to continue after the pandemic (at least partially), these typical average patterns are no more a reference. Other needs and a broader range of activities should be targeted to increase transit ridership (and reduce car dependency). Furthermore, the transit fare policy should be rethought to meet these new needs, as weekly and monthly passes have become much less attractive. However, analyses at a more disaggregate level are required to better understand intra-personal and inter-personal variations in subway use (Deschaintres et al., 2019) in the post-COVID new normal. Another lack in our paper is that we did not consider the spatial dimension. Variability of the impacts in space would also be interesting to explore. Some descriptive analyses by subway station were performed by the authors (available upon request), and the interaction with land use was also investigated. The results were as expected, therefore we decided to not present them in this paper for length issues. They revealed that the impacts of the detected breakpoints varied in space, depending on the station location and the activities allowed during each phase (all stations were not affected with the same magnitude in each phase). Overall, the presence of hospitals, open

businesses and schools near the stations was associated to lower declines in station use. However, these results are only based on 68 subway stations. Bus stops should be included to get a better picture of the spatial variability. The influence of the station/stop neighborhood on the response to COVID-19 could then be modeled, for instance using latent class modeling.

Regarding the other modes, bicycle, private car and carsharing (both FF and SB) were all more resilient to the crisis than the subway in Montreal. Nevertheless, it is important to keep in mind the important differences in scale: cycling and carsharing still support a much smaller number of trips than subway and car. As in other cases reported in the literature, modal shifts from the subway to these individual modes may have occurred. To validate this assumption, ITS models with autoregressive errors were trained to cross-analyze the impacts of the breakpoints detected in subway boardings on cycling counts, car counts and carsharing reservation hours (controlled for by annual seasonality, day type and weather conditions). In addition, changes in correlation between mode use and subway ridership by phase were modeled by introducing interaction terms. All direct impacts had an opposite direction on subway use versus bicycle use, and the positive correlation between the two modes has decreased from the first lockdown. This suggests a potential substitution, and this may also be related to a shift in travel purpose, activity location and mobility utility: recreative and sportive trips, made by cycling, may have replaced commuting trips made by subway because of the adoption of teleworking. Car use was influenced by measures that mainly affected commuting, but the correlation between these impacts and the subway boardings was not established. The impact of the COVID-19 pandemic on carsharing use (in time) was first negative but the hours of reservations started to recover later and continued to increase. This growth may have been delayed by the registration period of new enrollment and/or the inter-regional travel restrictions. When carsharing use increased the most, the correlation with the subway declined. Therefore, this growth was done independently of the increase in subway ridership. Overall, summing all coefficients of the interaction terms, the correlation between each mode use and subway ridership tended to decrease (except for SB carsharing); this means that the evolution of their use is less complementary. Some subway trips could have been substituted by other modes, and these modal shifts could explain a part of the decline in ridership. However, there is room for precisising these interactions. They are measured as global correlations at the system level in this paper, while specific types of interactions, intermodality for instance, should be studied. Specific times, peak hours for instance, should also be singled out to better understand when and why the

interactions between modes changed. In addition, they are drawbacks in the data and methodology used. The representativeness of the count data for cycling and private cars can be questioned since only some sites were selected, and changes in carsharing vehicle fleet should be considered. Although resupply was more difficult during the pandemic, the number of carsharing vehicles probably ended up increasing and this could also explain some of the delayed increase in total reservation hours. This additional independent variable could then improve the models obtained for this mode. Changes in transit services and in space sharing for bicycles could also be considered to assess their effects. Finally, the impact of each breakpoint was measured in correlation only, not in causality because the COVID-19 pandemic and the associated population-level events fall within non-experimental research.

8.6 Conclusion

Therefore, with a longitudinal and multimodal perspective, this paper provides a methodological framework to identify when changes have occurred because of the COVID-19 pandemic, and then allows quantifying and modeling these changes and their interaction with several modes. The results are specific to Montreal, but they confirm findings reported in the literature and the methodology can easily be replicated in other contexts. The outcome of such analyses could have practical implications for policymakers, operators, and planning partners. Indeed, the results yield a better understanding of the impacts of COVID-19-related events on mobility behavior and transportation systems. This could assist practitioners in adjusting planning and operational responses in case of new wave, based on a better knowledge of the magnitude of the effects of different types of measures. Such waves are still likely because of the emergence of virus variants, depending on the effectiveness of vaccines against the new forms of virus. The results could also help practitioners be better prepared to handle future pandemics or other similar crises and improve transit resilience. Moreover, as a slow recovery is being observed for transit, this paper could support practitioners in forecasting future ridership and proposing mobility projection scenarios. To that end, short-term effects (due to the temporary university and business closure, or the greater number of visits in the hospitals) and possible long-term effects should be distinguished. For instance, teleworking (if partially maintained) and home relocations are more likely to have a long-term impact. In addition, modal shifts generated by the pandemic could last in time if they have allowed people to discover a mode that is more convenient for them. Transit ridership is still far

from the pre-pandemic levels in many cities, but high congestion levels on the roads could lead to a turnaround in transit use. Furthermore, we believe that this crisis should be considered as an opportunity to redefine the role of transit. All conclusions show that a holistic approach that addresses all needs, not just commuting, and includes all modes of transportation should be adopted. In this line, the coverage of new demand corridors, based on the aggregation of desire lines for different travel purposes (Bahbouch et al., 2017), as well as the complementarity between modes, could be improved to help transit recovery.

Future work should investigate all these issues and possible outlooks. More recent longitudinal data should be acquired to update the results, including 2022, and confirm structural changes in subway use and interactions with other modes. This would allow studying the impact of the reopening after the fifth lockdown, which led to a period even closer to “normal”, as well as the impact of the new schedules in the companies (including teleworking or hybrid mode), which were mainly implemented at the beginning of the year 2022 in Montreal. With additional passive datasets, other modes, taxi and bus for instance, could be added into the analyses. The interruption of fare perception on board buses during the first lockdown could be an issue for this mode, but CPD methods still works with missing data: the corresponding days must be removed from the time series, but the methods will then identify the same breakpoints in the available information and the correct dates can be associated. Moreover, analyses at the individual level (per card) or at least by fare type (student pass, for instance) should be carried out to assess more disaggregate impacts (instead of systemic impacts). Indeed, the different phases of the pandemic may have had different effects on different people regarding the choice of using transit (Chen et al., 2022). Such analyses could be crossed with survey results to appreciate the changes by socio-demographic attributes and delve into passengers’ perceptions. Finally, forecasting models should be developed and applied to different scenarios to better grasp the most likely paradigm shifts in activity systems and travel behaviour.

Acknowledgments

The authors wish to thank all the providers of the passive datasets used in this paper: *STM – Société de Transport de Montréal* (Montreal transit authority) for AFC data, *Ville de Montréal* (municipal organization of the city of Montreal) for cycling count data, *PJCCI - Ponts Jacques Cartier et Champlain Incorporée* (private company responsible for several bridges) for traffic count data, and

Communauto (carsharing operator) for reservation data. They also acknowledge the financial support of the Canada Research Chairs (CRC), Quebec Research Funds (FRQNT) and Institute for Data Valorization (IVADO) programs.

CHAPITRE 9 ARTICLE 6 : ASSESSING COMPLEMENTARY AND COMPETING INTERACTIONS BETWEEN TRANSIT AND SHARED TRANSPORTATION MODES

Elodie Deschaintres, Catherine Morency, Martin Trépanier

Accepté pour présentation et recommandé pour publication dans *Transportation Research
Procedia : World Conference on Transport Research 2023*

Date de soumission initiale : 1 décembre 2022

Abstract

Shared transportation modes have been introduced in many cities in recent years. However, their interactions with the existing transit system, i.e., whether they complement or compete with it, are still unclear. This paper addresses this issue by assessing relationships between two shared modes (bikesharing and free-floating carsharing) and transit (subway and bus) using several passive data streams from Montreal. An original bio-inspired terminology is first proposed to outline different types of interactions based on their impacts on the ridership of each mode in the short or long term. In line with this conceptual framework, a rule-based algorithm is developed to classify bikesharing and carsharing disaggregate trips into different groups of potential complementary or competing relationships with transit. This classification depends on the feasibility of serving the trip by transit, but also the quality of the transit alternative calculated by a transit routing engine. The spatial and temporal distributions of trips in different categories are then analyzed to highlight travel demand corridors where transit services are potentially supplemented or substituted by the shared modes. Finally, the causal impacts of the competing bikesharing trips on daily route-level transit ridership are evaluated using a fixed effects difference-in-differences model. The results reveal that daily ridership on transit lines that would have been taken in the absence of bikesharing may have been reduced by 1.3% for every hundred bikesharing trips. However, the shared modes can also complement transit in time, space, or when the service is unsatisfying or saturated. Therefore, this paper provides tools to better inform operators and urban planners to develop more integrated transportation systems.

Keywords: transit; bikesharing; free-floating carsharing; complementarity; competition; difference-in-differences

9.1 Introduction

Over the past few years, the rapid development of information and communication technology led to a diversification of urban transportation services. For instance, shared modes such as bikesharing and carsharing, micromobility (e.g., electric scooters) and ride sourcing programs (e.g., Uber or Lyft) enhanced classic transportation (e.g., transit) supply. This increasing multimodality is desirable within the context of sustainable mobility since alternatives to the exclusive use of a private car may be provided (Spickermann et al., 2014). However, little research has been conducted to better understand how these various modes actually interact (i.e., whether they are complementing or competing). Considering that transit ridership is stagnating or declining in some North American cities (Erhardt et al., 2022; Graehler Jr et al., 2019), questions about the impacts of emerging services on traditional modes are still raised.

Against this background, this paper aims to give a deep dive into interactions between urban transportation modes at the trip level. More specifically, different types of relations between transit (subway and bus) and two shared modes (bikesharing and free-floating carsharing) are investigated for Montreal (Quebec, Canada). To this end, historical bikesharing and carsharing trips are classified into different interaction categories based on their potential complementary or competing relationship with transit. In addition, these relationships are analyzed over time and space, and their impacts on daily transit ridership per line is evaluated. Therefore, this paper provides tools to help clarify the role of several transportation modes in urban mobility and better integrate them in policy and planning strategies.

This paper is organized as follows. First, some background regarding the impacts of new modes on transit is presented and the corresponding methods are reviewed. The case study and data are then described, a conceptual framework for defining different types of interactions is proposed, and the methodological elements are detailed. The results of the trip classification, spatial-temporal analysis and impact modeling are reported. Finally, the paper concludes with a discussion of policy implications, limitations of the work and future investigations.

9.2 Background

Mobility as a Service (MaaS) is a concept that has become increasingly popular since Hietanen (2014) introduced it. This notion implies the integration of all modes of transport into a single

mobility service based on accessibility (rather than ownership) and tailored to the traveler's needs. MaaS is envisioned as an opportunity to promote more sustainable mobility by substituting car trips (Jittrapirom et al., 2017). This role is even more important in the post-COVID mobility to help transit recovery and discourage a further shift to private cars (Basu & Ferreira, 2021). However, several works in the literature revealed that, instead of improving modal connectivity with transit, shared modes can also compete with it.

In particular, a large proportion (40-60%) of survey respondents declared that their trips would have been made by transit if bikesharing did not exist, suggesting a modal shift between the two modes (Fishman et al., 2014; Fuller et al., 2013). Indeed, bikesharing can be preferred to transit when it offers a more direct (without transfers), faster (without waiting time) and more flexible alternative than transit for short distances. However, many bikesharing users also reported combining the two modes to connect their trip origin and/or destination to a subway station or bus stop (Fuller et al., 2013; Shaheen et al., 2013). According to a recent study in four U.S. cities, modal integration between transit and bikesharing is more likely among bikesharing subscribers, especially for commuting on weekdays (Kong et al., 2020). Likewise, Ma, Xiaolei et al. (2019) showed that the use of free-floating shared bikes decreased transit ridership on weekends whereas a positive impact was observed during the peak hours on weekdays. The relationship between bikesharing and transit may depend on the transit mode (Graehler Jr et al., 2019): bikesharing complements subway by playing a role in the first or last mile (Ma, T. & Knaap, 2019; Ma, Xinwei et al., 2018), or even both (Tarpin-Pitre & Morency, 2020), while it tends to compete with buses on trips that are shorter on average (Campbell & Brakewood, 2017). Furthermore, the effect of bikesharing may vary between major, dense cities and smaller, lower-density cities (Shaheen et al., 2013), with higher substitution rates in the former than in the latter. Variations were also measured within the same city, where transit use is reduced by bikesharing in the city center but increased in the urban periphery (Ma, T. & Knaap, 2019; Martin & Shaheen, 2014; Tang et al., 2021). The most common explanation is that bikesharing provides access to transit in underserved areas and complements it outside of service time and network area, while the shared mode may be more competitive and quicker than transit for short distances in dense areas, especially when the network is saturated.

The positive impacts of carsharing to reduce car ownership and vehicles-miles traveled have been largely demonstrated in the literature, but the interactions with transit have been much less

analyzed. Similarly to bikesharing, carsharing was found to both complete (Ceccato & Diana, 2021) and compete with transit (Papu Carrone et al., 2020). Even if, overall, higher modal shares of transit are observed among carsharing members (Clewlow, 2016), this may be affected by self-selection and simultaneity biases (Mishra et al., 2019). In other words, the detected impacts may be due to (observed or unobserved) differences between members and non-members of the service (e.g., choice of residence), not to the subscription itself, or they may be biased because there is a reverse causality between an explanatory variable and the dependent variable. For example, if people with higher transit use are more likely to start using carsharing, then the outcome variable (transit use) may also influence the causal variable (carsharing membership). On the one hand, carsharing can extend the existing transit network but, on the other hand, it can compete with transit by providing the comfort and flexibility that the rider could have with a private car (Silvestri et al., 2021). The impact may depend on car ownership: members who reduced their car ownership by joining a carsharing service tend to use transit more, while previously non-motorized members who gained access to a car through carsharing decreased their use of transit (Lane, 2005; Martin & Shaheen, 2011). Furthermore, free-floating carsharing is more likely to substitute transit than station-based carsharing since it allows for daily one-way trips (Becker et al., 2017; Le Vine et al., 2014). However, there is still room to better understand the variations of these interactions in time and space.

Most of the existing works to assess interactions between transit and shared transportation modes are based on cross-sectional surveys. Beyond response and representativeness biases to which they are subject, these data generally allow only qualitative analyses. Survey respondents are asked to indicate which mode is replaced by another most often or for a specific trip (e.g., the last one), allowing only one type of interaction (substitution or complementarity) per respondent (while the same respondent may both combine and substitute the modes under different conditions), and no quantitative information on the frequency of the relationship is collected. At best, only retrospective stated frequencies are compiled. Therefore, the general trend estimated at the respondent level may not reflect the real impact on total ridership. In addition, although they provide useful contextual insights, surveys lack spatial and temporal details to investigate the variability of the relationships between modes in time and space. Rather, passive operational data streams contain high-resolution spatial and temporal information by collecting the timestamp and location of the start and end of each mode usage. Such data were leveraged to measure systemic

correlations between mode ridership (Morency et al., 2020; Morency et al., 2018; Tang et al., 2021) or between the bikesharing trip origins/destinations and the number of transit boardings/alightings (Ma, Xiaolei et al., 2019). Longitudinal data also allow for the use of causal inference techniques, such as the difference-in-differences method, which reduces the confounding selection biases between a treatment and a control group (Campbell & Brakewood, 2017; Ma, Xiaolei et al., 2019; Ziedan et al., 2021).

However, these analyses at the system level (generally with trips aggregated by zone) evaluate only the global impact on transit ridership, which can be the sum of different individual effects (competing and complementary relationships may coexist in the same place), and they do not account for interactions which may not influence transit ridership in the short term (e.g., if the other modes complete transit outside service time or service area). More importantly, they do not consider if the trip substitution/complement is really feasible. Only the spatial relationship between transit and the shared mode (based on the proximity between the transit system and mode infrastructure) is examined, overlooking transit planned schedules and the quality of the transit alternative. Therefore, very recent but few studies investigated interactions at the trip level (Kong et al., 2020; Luo, H. et al., 2021; Song, Ying & Huang, 2020) for bikesharing and e-scooters (but to the best of our knowledge, no such analyses exist for carsharing, probably due to the difficulty of accessing disaggregate data). Nevertheless, the feasibility of the relationship was poorly considered using similar start and end times of the trips or simple transit route search algorithms (allowing one transfer maximum), and/or the quality of the transit alternative was ignored. Our work will use a more developed routing tool and will demonstrate the value of considering the transit route that would have been taken if the shared mode was not present to assess the impacts of this mode on transit ridership.

In this way, this paper will contribute to filling gaps in understanding the relationships between modes, which have so far been based mainly on cross-sectional surveys and systemic analyses (i.e., without explicit study of interactions). Furthermore, an original bio-inspired terminology will be proposed to establish different types of interactions according to their impact on transit systems (distinguishing between short-term and long-term effects). There is no such standardization in the transportation field, yet it is of great importance to better define the role of the different modes in a multimodal context.

9.3 Data and methodology

In this paper, interactions between transit and shared transportation modes are explored at the trip level to distinguish all possible relationships, which may have impacts or not on transit ridership, and to consider the feasibility of the complementary or competing interaction. To this end, several passive datasets containing disaggregate information are harnessed. A conceptual framework is first defined to clarify different types of relationships between modes. Based on this framework, a trip classification algorithm is then proposed to classify the trips into different interaction categories, and the temporal and spatial distributions of the trips in each category are analyzed. Finally, the impacts on transit ridership are evaluated using a fixed effects difference-in-differences (FE-DID) regression model.

9.3.1 Case study and datasets

Montreal (Quebec, Canada) is the case study of this work. It is a large-sized city with a well-developed transit network, composed of four subway lines (named by color) and about two hundred bus lines. This network, represented in Figure 9.1, is operated by the STM (*Société de Transport de Montréal*) transit authority. The Montreal bikesharing system, called BIXI, has been exploited since 2009. It is a station-based bikesharing system, consisting of about 700 stations (located in Figure 9.1) and 9000 bikes (including e-bikes) in 2021. Every year, the system operates from April 15 to November 15 and is closed during the winter season. The carsharing system was implemented in Montreal in 1995 by the Communauto company. Both station-based and free-floating carsharing services are available, but only free-floating (called FLEX) is considered in this paper as only this service allows for direct one-way trips. In 2021, the free-floating zone (in which the vehicles must be picked up and dropped off) covered about 115 km² of the Montreal Island (see Figure 9.1). In addition, there are FLEX drop-off points, which are on-street or off-street parking spaces located in the city center (spotted in Figure 9.1), and therefore outside the service area, where it is also possible to start and end a trip with a FLEX vehicle.

Trip-level data (disaggregated trips or trip legs) are mined for these three modes. Transit transactional data were provided by STM. They were collected by an automatic fare collection system (OPUS system) which gathers all tap-in validations (at boarding only) of smart cards and tickets. Bus data were compiled at the route level because bus validations are not geolocated at the

stop level. Subway data were available at the station level, but we also aggregated them by route. For stations at the intersection of several lines, we applied average daily percentages (taken from the 2018 Montreal regional household survey) to distribute the validations on the different lines. The data sample used in this paper covers several weeks in the fall of 2021; this was the first “normal” (or close to normal) period after the COVID-19 pandemic because most activities were reopened. More specifically, the study period includes the five weeks (34 days) of September 2021, spanning from Monday, August 30 to Sunday, October 3, without Monday, September 6 that we removed because it was a Quebec public holiday. Moreover, a control period (with BIXI closed) is selected to be used in the model which will be specified later. This control period covers the last half of November 2021 and the first half of December 2021, from Thursday, November 16 (one day after the closure of BIXI) to Sunday, December 19. This is also 34 days with the same number of each day type as in the five weeks of September 2021 (4 Mondays and 5 days for the other weekdays and weekend days). Therefore, the aggregation (sum of all validations) of the entire transit data by day and route (without distinction of the direction) led to 216 lines (common lines to both the study and control periods) x 68 days = 14,688 observations.

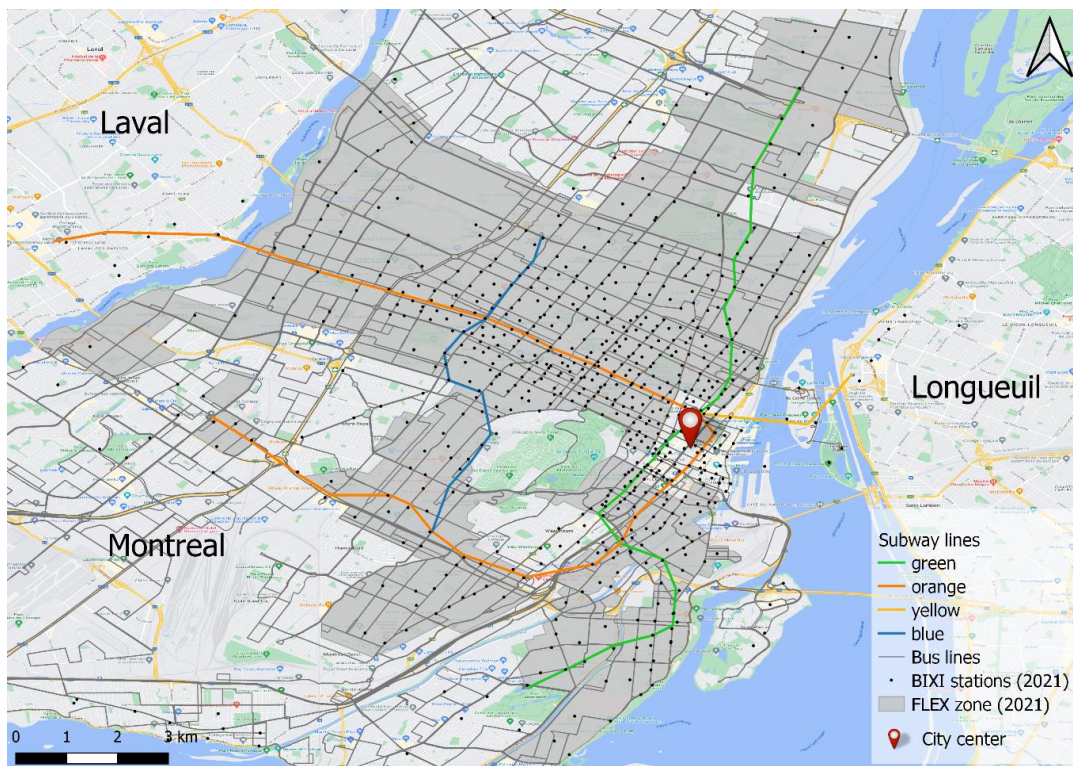


Figure 9.1 Case study city (Montreal) and the transit, bikesharing and free-floating carsharing systems

For the two shared transportation modes, open historical bikesharing data were downloaded from (BIXI, 2021) and carsharing data were obtained from Communauto. They are transactional data streams containing the timestamp and coordinates of the origin and destination of each transaction and trip duration. Trip distance is available in the carsharing data and was calculated based on GPS trajectory by the company, but for bikesharing we used the Transition platform (presented in section 9.3.3.1) to estimate the distance of the shortest path from origin to destination on the OpenStreetMap bicycle network. The data selected for both modes cover the five weeks of September 2021 and include only direct one-way trips (so that the alternative transit path can be calculated). Therefore, only the trips with different origin and destination stations were selected for bikesharing (this also cleaned up the immediate returns after a pickup due to a bike malfunction). The resulting data sample consists of over 1,053,500 trip records, or 96% of the original records. For carsharing, we applied several criteria from a previous thesis work that classified Montreal carsharing trips and cross-validated the results with several data sources (Wielinski, 2018). These criteria, whose goal is to remove round trips and one-way trips with an activity, are as follows:

- The number of trips made during the reservation must be equal to 1. This variable was calculated by Communauto based on the number of engine stops, but they warned us against the unreliability of the value as several stops may have been missed. Therefore, other validations must be performed.
- Total distance and total duration of the trip must be inferior to 100 kilometers and 3 hours respectively. They are the maximal distance and time thresholds needed to cross the entire Montreal Island.
- The Euclidean distance between the origin and destination must be superior to 760 meters (threshold based on the cited work).
- The ratio between the Euclidean distance and the real distance must be superior to 0.121 (threshold based on the cited work).

To validate the results, the average trip speed (distance divided by duration) was also computed, and we only kept the trips between the 5% and 95% percentiles (the lowest speeds may have indicated that an activity was performed during the reservation whereas the highest speeds were outliers). Over 85,600 trips (about 64% of the original records) were selected based on this method and this result is similar to that of Wielinski (2018).

Finally, GTFS (General Transit Feed Specification) data are handled to describe the transit network and schedule information. We gathered STM data, but also STL (*Société de Transport de Laval*) and RTL (*Réseau de Transport de Longueuil*) data for Laval and Longueuil because there are BIXI stations in these two cities, in the North and South of Montreal respectively. The scheduled vehicle-kilometers traveled per line for each day of the study period were also calculated from this data to be used as control variable in the future model. Other contextual information includes route-level attributes, aggregated in a 400 m buffer around each transit line: population and employment density (estimated from the 2018 Montreal regional household survey), median household income (from the 2016 Canadian census), as well as weather information at the day level (average temperature and total rainfalls) from governmental open data (Environment Canada). Descriptive statistics are reported in Table 9.1.

Table 9.1 Descriptive statistics

Variable	Mean	SD	Min	Max
Dependent variable				
Daily transit ridership per line	4,332.63	20,247.29	1	274,996
Treatment variable/intensity				
Number of bikesharing stations (treatment #1)	27.48	32.03	0	161
Bikesharing use (treatment #1 and 2)	701.30	1,603.85	0	9,706
Bikesharing use (treatment #3)	197.97	462.52	0	3,152
Bikesharing use (treatment #4)	28.15	136.81	0	2,638
Control variables				
Scheduled vehicle-kilometers	1,028.01	1,239.47	13.14	13,473.34
Population density	5,577.49	2,947.26	92.48	12,525.33
Employment density	5,636.56	8,656.79	521.51	54,150.75
Median household income	31,964.68	5,261.64	2,967.74	44,797.97
Average daily temperature (°C)	8.35	10.09	-9.30	24.80
Total daily rainfalls (mm)	2.76	5.84	0.00	28.20

9.3.2 Conceptual framework

Different types of interactions between two modes can be defined based on their impacts on each mode, which can be positive (+), negative (-) or neutral (0). In this paper, we propose an original conceptual framework with a terminology based on analogy with the six types of biological interactions between species (Odum & Barrett, 2005) described in Table 9.2. This typology of interactions has already been transposed to the interactions between different technologies (Sandén & Hillman, 2011). Although the interactions between two modes can be expressed in terms of

resources (competition in space, infrastructure, energy, etc.), here they will be defined based on the impacts on market share and ridership of each mode. As with interactions between technologies, dynamic (long-term) and quasi-static (short-term) interactions between modes must be distinguished.

Table 9.2 Biological interactions applied to two modes A and B based on the impacts on their ridership: positive (+), negative (-) or neutral (0)

INTERACTION	Impact on (old) MODE A	Impact on (new) MODE B
Competition	-	-
Mutualism	+	+
Neutralism	0	0
Parasitism	-	+
Commensalism	0	+
Amensalism	0	-

A dynamic interaction refers to the arrival of a new mode B (for instance, a shared mode) when an old mode A (for instance, transit) is already well established, or to long-term changes in the market shares of each mode. Mode B can attract the market share of mode A, leading to a modal shift between the two modes, also called *parasitism* in biology, or cannibalization in economy (Moorthy & Png, 1992). However, mode B can benefit from a market niche brought out by mode A without having impacts on the ridership of the old mode. For instance, on-demand collective taxis (mode B) can be used for paratransit, a special type of transit (mode A). In this case, we have a situation of *commensalism*. *Amensalism* occurs if the market is locked out by mode A or if mode B suffers from the bad reputation of mode A. In this way, mode B is inhibited, while mode A is not affected. *Mutualism* exists when the two modes (for instance, transit and bikesharing) complement each other (mode B fills the mobility gaps of mode A) and create a new market by causing a modal shift from another mode (for instance, private car); this is the most desirable situation for a transition towards vehicle-free lifestyles. Mode B may also have a neutral relation (*neutralism*) with mode A if it develops an independent market (competing with another mode or creating a new demand) without interacting with mode A. Finally, two emerging technologies may also compete for a common market (*competition*).

Quasi-static interactions can take the form of one of the three symmetric interactions (competition, mutualism or neutralism) and correspond to a short-term quasi-stable situation (the new mode B has been established for some time, and the market shared by the two modes is assumed to be

almost fixed). These are the interactions that will be assessed in this paper (between a shared mode as mode B and transit as mode A) because the modes were implemented several years ago, the study period is short (one month), and the interactions are analyzed at the daily level (daily variations). The two modes can evolve independently without influencing each other (no overlap in demand); this is the *neutralism* case. The two modes can also compete for the same customers (*competition*). In this case, when people travel, they make a choice between mode A and mode B, based on the availability of the service (for a specific time and location), the quality of the trip and their preferences. On the contrary, the two modes can be used in complementarity, until perhaps they become mutual dependent (*mutualism*). The two modes can be integrated within the same trip, with one main mode and another mode used to make the first/last mile connection; this is also called modal integration, or intermodality, and has a positive impact on each mode ridership. Nevertheless, other types of complementarities may exist without affecting the demand of mode A, when mode B supplements mode A in different (independent) trips: this amounts to *commensalism*. For instance, if the trip is not feasible with mode A because both origin and destination are outside the service area (complementarity in space) or trip start is outside the service time (complementarity in time), or the trip is not feasible within reasonable trip duration, trip distance, waiting time and transfer time (complementarity in quality).

In this paper, all trips of mode B (bikesharing or carsharing) will be classified according to if they have no impact on, compete with or complement mode A (transit). For more clarity, the neutralism relation (which means no interaction) will correspond to the “no impact” category. The “competition” category is already explicit, and “complementarity” will be decomposed in three types: complementary in space will group together first/last mile connections and trips made outside the transit network area, complementarity in time will gather the trips made outside the transit service time (at night for instance), and complementarity in quality will correspond to the trips for which the transit alternative is not competitive.

9.3.3 Methodological elements

9.3.3.1 Trip classification algorithm

The rule-based algorithm used to classify the bikesharing and carsharing trips into different categories of potential interactions (complementarity, competition or no impact) with transit is

presented in Figure 9.2. It is important to note that only potential relationships between bikesharing (or carsharing) and transit are measured based on several chosen rules and parameters ($\theta, r, \alpha, \beta, \gamma, \delta, \epsilon$) which are required to create links between the independent datasets of the modes. In addition, the shared modes could substitute modes other than transit. Therefore, the detected relationships can be viewed as the upper bound of the actual complementary and competing interactions. The classification depends on the spatial-temporal feasibility of the trip with transit and on the quality of the alternative in transit. This is evaluated using, on the one hand, the trip start/end time and location and, on the other hand, the transit infrastructure and schedule. Therefore, the input data are the historical bikesharing and carsharing trip records and the GTFS data introduced before.

The algorithm is inspired by the one of (Luo, H. et al., 2021), used to assess trip-level interactions between e-scooters and bus, but we include more categories, consider walking network distances instead of Euclidean distances and, rather than a simple route search algorithm (allowing for one transfer only) which may be insufficient for other modes with longer trip distances, we benefit from a more complex routing tool to calculate the transit alternative, namely trRouting (Bourbonnais, 2019). This transit routing engine, based on the Connection Scan Algorithm (Dibbelt et al., 2013), is open source and was integrated in the Transition transit network simulation platform (Bourbonnais et al., 2019). This platform also makes use of the Open Source Routing Machine (Luxen & Vetter, 2011) to compute shortest paths in the bicycle and pedestrian networks of the OpenStreetMap data (with the default walking and bike speeds of 5 and 15 km/h respectively). For each given Origin-Destination (OD) pair and departure time of a bikesharing (or carsharing) trip, Transition allowed us to calculate the transit alternative with the earliest arrival time based on the imported GTFS data. The returned data include the times and distances for each component of the trip (access and egress, waiting, in-vehicle and transfers), as well as the sequence of transit lines used.

The trip classification algorithm works as follows. For each bikesharing (or carsharing) trip i , we first look at the total walking travel time t_{tot}^{walk} (as if the whole trip was made on foot) to remove walkable trips. If this time is inferior to the threshold $\theta = 10$ minutes, the trip is classified into the “no impact” category because we assume that the mode is more likely to replace walking than transit in this case. Indeed, the results of a two-wave longitudinal survey of dock-less e-bike share

users showed that the dominant mode substitution for trips of less than 1 mile (equivalent to about a 10-min walk) for most trip purposes was walking (Fukushige et al., 2021).

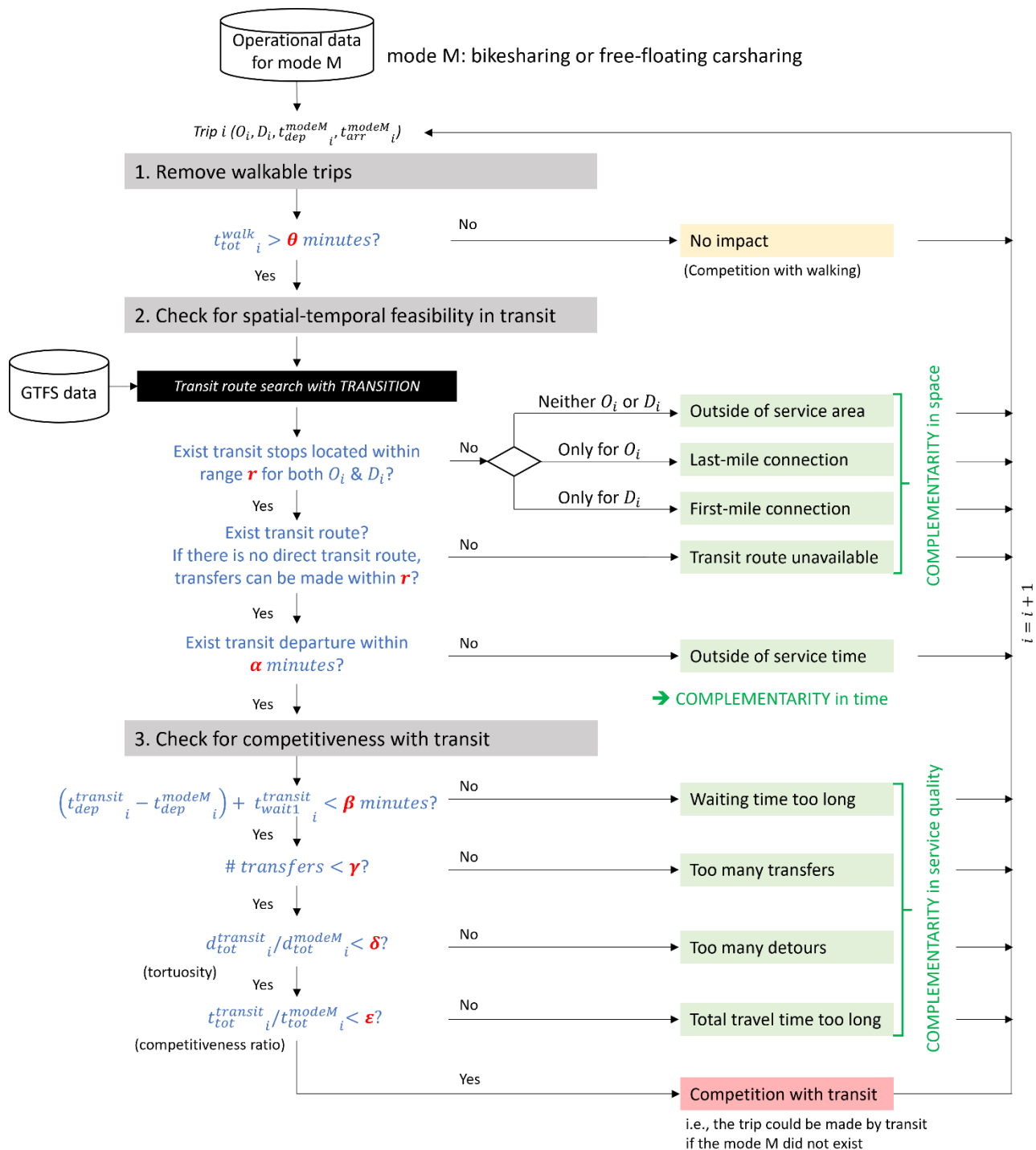


Figure 9.2 Rule-based classification algorithm to assess potential complementary and competing interactions with transit

Then, the spatial-temporal feasibility of the trip in transit is checked. If there are no transit stops located within a range r (walking time on the pedestrian network) of the trip origin (O) and destination (D), we consider the shared mode was used to complement transit “outside of service area”. If stops exist within the surrounding of either the origin or destination but not both, the trip is identified as a “first-mile connection” or “last-mile connection”. For a trip with stops near both O and D, the Transition platform evaluates if a transit path exists (direct route or with transfers). If not, the trip falls within the “transit route unavailable” category. The maximum access, egress and transfer walking times are all set to $r = 5$ minutes. This corresponds to the commonly used threshold of 400 meters (or 0.25 miles) to measure the catchment area of transit (Campbell & Brakewood, 2017; Kong et al., 2020; Tarpin-Pitre & Morency, 2020) with a constant walking speed of 5 km/h. Regarding time parameters, the minimum waiting time before boarding is set to 3 minutes to account for random service delays. This means the user must arrive at least 3 minutes early at every stop. The maximal waiting time is 35 minutes, which corresponds to the maximum headway for most of the STM transit lines. Beyond this time, the trip is categorized as “outside of service time”. All the previous rules of spatial-temporal feasibility were coded in Transition (with the possibility to change the threshold values) to be inspected directly in the platform. For these five interaction categories, the trip path is not calculated, and error messages are printed to indicate why (no access at origin and/or destination, no routing found or no service from origin). Nevertheless, for the trips in the “outside of service time” category, we changed the departure time to validate that a path exists. Otherwise, they were moved to the “transit route unavailable” category.

For the following steps, intended to check competitiveness with transit, we use the information returned by Transition to compare the trip with the corresponding transit alternative according to several criteria which are typical trip attributes included in the utility function for mode choice modeling. First, waiting time is examined. The first waiting time (at the first stop) provided by Transition ($t_{wait1}^{transit}$) is always optimized (minimized) by the algorithm, adjusting the departure time of the user. However, the initial trip departure time is not the true departure time since the user walked to the bikesharing station (or free-floating carsharing vehicle). Therefore, we measure waiting time as the sum of the initial lost time at the departure (difference between the departure times by transit and with the shared mode) and the optimized waiting time returned by Transition. If the total is superior to $\beta = 15$ minutes, then waiting time is considered as unacceptable due to

the limited transit frequency and the trip is classified into the “waiting time too long” category. If the transit path includes “too many transfers”, transit may have been considered inconvenient and the shared mode may also have been preferred. The maximum acceptable number of transfers is set at $\gamma = 2$ for bikesharing as in other studies (Kong et al., 2020; Luo, H. et al., 2021), but at $\gamma = 3$ for carsharing to account for longer trips. The maximal tortuosity (travel distance ratio) and the maximal competitiveness (travel time ratio) between the bikesharing (or carsharing) trip and the transit alternative are set at $\delta = 2$ and $\varepsilon = 2$ respectively, which was the threshold also applied by Luo, H. et al. (2021). If these values are exceeded, we assume the shared mode was used to avoid “too many detours” or a “total travel time too long” in transit due to fixed routes and planned schedule (in other words, the shared mode proposes a significantly more direct or faster option). Finally, all remaining trips fall within the “competing with transit” category. Therefore, to be a potential competing trip (and thus potentially substitute transit), the transit alternative must be feasible (in time and space), but also competitive (of good quality), i.e., the trip could be served by transit with reasonable duration, distance, waiting time and transfer time if the shared mode did not exist.

Several thresholds ($\theta, r, \alpha, \beta, \gamma, \delta, \varepsilon$) are used in this algorithm and the selected values are the following: $\theta = 10$ minutes, $r = 5$ minutes, $\alpha = 35$ minutes, $\beta = 15$ minutes, $\gamma = 2$ for bikesharing and 3 for carsharing, $\delta = 2$, $\varepsilon = 2$. Sensitivity analyses were performed by Luo, H. et al. (2021) and showed that, due to the reduced candidate transit stops pool around the trip OD, lower values of r decreased the share of competing trips and increased the share of trips complementary in space and in quality (especially with waiting time too long). The values of the parameters to assess the quality of the transit alternative may also vary according to individual preferences. Overall, less restrictive (higher) values lead to more competing trips since they allow more trips to be served by bus. Despite this high dependence on the chosen parameters, the competing relationship will be validated using a causal DID inference method.

9.3.3.2 Temporal and spatial analysis

The results of the trip classification are then analyzed and their variations in time and space are examined. Temporal distributions of the trips in each class by day type (from Monday to Sunday) and hour of the day are plotted. For the spatial dimension, we do not map the origins (or destinations) of the trips as in all other papers in the literature, but rather OD pairs (or travel desire

lines) to better identify the gaps in transit services. These OD straight lines are intersected with a hexagonal grid (hexagons of 100-m diameter), and densities of desire lines of each interaction category are then calculated per hexagon. In addition, for the competing relationship, the transit paths (sequence of transit lines) that would have been taken if the shared mode was not present is exploited to estimate a number of competing trips per line. This allows the detection of the most potentially affected transit lines.

9.3.3.3 Fixed effects difference-in-differences (FE-DID) regression model

Finally, a FE-DID regression model is built to validate the classification results and quantify the impacts of the interactions on transit ridership. DID model is part of statistical techniques specifically designed to make causal inferences in a non-randomized experimental setting. The method is used to measure the impact of a treatment, policy or intervention on outcomes by comparing the changes over time between a treatment and a control group.

In this paper, the effect of interest is the short-term impact of the competing trips on daily route-level transit ridership. All types of interactions were analyzed (in time and space) but only competing relationships were considered in the model because, as explained in the conceptual framework, complementarity does not necessarily have an impact on transit ridership in the short term. Only first-mile and last-mile connections constitute a mutualistic interaction (win-win situation for each mode ridership), but they were hard to detect with the proposed method in the Montreal context (see the results for more details). This limited assessment of the positive impact of complementarity between modes could be improved with long-term analyses; this point will be further discussed in the conclusion. In other transportation-related works (Campbell & Brakewood, 2017; Luo, H. et al., 2021; Ma, Xiaolei et al., 2019), DID models were applied to analyze the launch of a new service (for instance, a new bikesharing system) based on data before and after the implementation. Therefore, they measured dynamic interactions between modes. However, in Montreal as in other many developed cities, shared modes were introduced several years ago (in 1995 for carsharing and in 2009 for bikesharing) and we do not have data to cover these changes. Nevertheless, the Montreal context offers a situation in which we can assess quasi-static (short-term) interactions. We will use September 2021 as the period with treatment (BIXI open) and November-December 2021 as the period without treatment (BIXI closed). There is no such control period for carsharing, that is why only bikesharing interactions with transit will be modeled.

However, as the same trip-classification algorithm was applied to the two modes, we can assume that the classification will be valid for both. In fact, for carsharing we could have used the evolution of the free-floating carsharing zone and selected a place where the zone was extended (before/after the extension), but the other problem is that free-floating carsharing reservations were not geolocated by Communauto before 2021.

The model is estimated using the following regression equation:

$$\begin{aligned} Y_{it} &= \text{LogTransitRider}_{it} \\ &= \alpha + \beta_1 \text{Period}_t + \beta_2 \text{Group}_i + \beta_3 \text{Period}_t * \text{Group}_i * \text{BikeUse}_{it} \\ &\quad + \gamma \text{Controls}_{it} + \delta \text{DateFE}_t + \lambda \text{RouteFE}_i + \varepsilon_{it} \end{aligned}$$

where $\text{LogTransitRider}_{it}$ is the logarithm of daily transit ridership for the route i and date t . β_3 is the impact of interest, known as the average treatment effect on the treated (ATT). It is the interaction between Period_t (dummy variable for the period: $\text{Period}_t = 1$ in September 2021 when BIXI is in operation, $\text{Period}_t = 0$ in November/December 2021 when BIXI is closed) and Group_i (dummy variable for the group membership: $\text{Group}_i = 1$ if the transit route i is in the treatment group and 0 if it is in the control group), weighted by the bikesharing use intensity within the 400 m buffer of the route i on date t . β_3 is also the double difference (hence the name of the DID method) between the two periods and the two groups:

$$\begin{aligned} \beta_3 &= (E[Y_{it} | \text{Group}_i = 1, \text{Period}_t = 1] - E[Y_{it} | \text{Group}_i = 1, \text{Period}_t = 0]) \\ &\quad - (E[Y_{it} | \text{Group}_i = 0, \text{Period}_t = 1] - E[Y_{it} | \text{Group}_i = 0, \text{Period}_t = 0]) \end{aligned}$$

This formulation shows that the method removes selection biases resulting from permanent differences between the treatment and control groups, as well as biases caused by temporal trends or seasonal effects having the same impact on both groups.

The treatment group gathers the transit lines which potentially face more the impact of bikesharing. Four definitions of group membership and treatment intensity are compared. In the literature, infrastructure availability (for instance, in number of stations or docks or bikes) is often chosen as measurement to define the treatment group. However, service use (with trip-level data) helps to account for daily variations and better quantify the impacts. The two cases are examined in treatment #1 and treatment #2. Furthermore, Luo, H. et al. (2021) demonstrated the importance of accounting for the competing relationship (rather than assuming all trips to have the same effects)

in evaluating the impact of shared e-scooters on bus ridership. This will be validated for bikesharing in this paper (treatment #3). In addition, we also investigate the value of considering the transit path that would have been taken in the absence of bikesharing using the sequence of lines of the transit alternative returned by Transition (treatment #4). Therefore, the four treatment measurements differ in the grouping variable ($Group_i$) and in the treatment intensity ($BikeUse_{it}$) as follows:

- Treatment #1: $BikeUse_{it}$ is the number of bikesharing trips originating near the line i on date t and $Group_i$ is based on the number of bikesharing stations nearby the line;
- Treatment #2: $BikeUse_{it}$ is the number of bikesharing trips originating near the line i on date t and $Group_i$ is based on the daily average of the number of bikesharing trips nearby the line;
- Treatment #3: $BikeUse_{it}$ is the number of bikesharing potential competing trips originating near the line i on date t and $Group_i$ is based on the daily average of the number of bikesharing potential competing trips nearby the line;
- Treatment #4: $BikeUse_{it}$ is the number of (potential missing) boardings on the line i and date t based on the transit paths of the bikesharing potential competing trips and $Group_i$ is based on the daily average of the number of boardings on the line.

Descriptive statistics on these different variables are provided in Table 9.1. In the four cases, the transit lines are classified into the treatment or control group based on the median value of the grouping variable (for instance, for treatment #1, the lines with a number of bikesharing stations superior to the median value of the number of stations per line are classified into the treatment group, while the remaining lines are in the control group). In this way, there is the same number of lines in the treatment and in the control group (108 and 108). Please note that other thresholds (other quantiles) were also tested and gave similar conclusions.

Control variables ($Controls_{it}$) include route-level variables (population density, job density, median household income and mode: subway versus bus), day-level variables (average temperature and total rainfalls), as well as the logarithm of the scheduled vehicle-kilometers traveled per day and per route (both area- and time-varying variable), which allows accounting for the differences in service frequency and route length. In addition, the clustered structure of the data (observations

nested into routes) is taken into account by including fixed effects in the model. Indeed, hierarchical data are generally modeled using random-effects or fixed-effects multilevel models, but fixed effects models can better account for omitted variables (Allison, 2009), and therefore they are typically used with DID models to reduce confounding and generate unbiased estimates of the treatment effect. Here, as in other studies on interactions between modes (Campbell & Brakewood, 2017; Wang, Mingshu & Zhou, 2017; Ziedan et al., 2021), we introduce date ($DateFE_t$) and route ($RouteFE_i$) fixed effects in the model. They take the form of dummy variables, one for every route (except one) and one for every date (except one), to control for unobserved route-level time-invariant characteristics and daily systemic variations respectively.

The main assumption of the DID approach is the parallel trend hypothesis: both the treatment group and the control group must have similar trends before treatment. This was validated visually, for instance with Figure 9.3 for treatment #4, and with an ANOVA analysis: the interaction term between the group variable and the time variable (for linear trend) was not significant. Figure 9.3 also shows that treatment routes have higher average ridership than control routes and increase in average ridership after the closure of BIXI is slightly higher in the treatment group than in the control group. Normality of the independent variable was achieved thanks to the logarithm, and correlations between the independent variables were also checked.

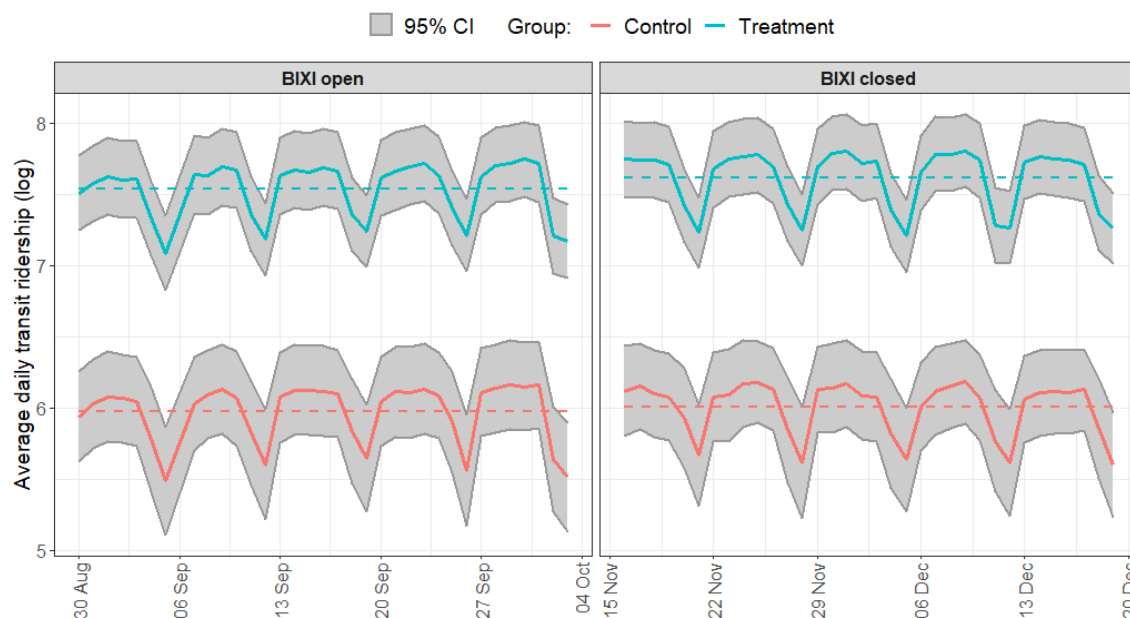


Figure 9.3 Average daily transit ridership for the lines in control and treatment groups (based on treatment #4), with BIXI open and closed

9.4 Results

The general results of the classification of the bikesharing and carsharing trips in different interaction categories based on their potential complementary or competing relationship with transit are first described. Then, these results are analyzed in time and space, and they are validated with the FE-DID regression model by evaluating if the competing trips have actually led to a decrease in transit ridership.

9.4.1 General results of the trip classification

The bikesharing and free-floating carsharing trips are distributed per interaction class in Table 9.3. According to these results, about 12% of the total bikesharing trips were feasible in less than 10 minutes by walking and are therefore assumed to have no impact on transit. This proportion is almost null for carsharing since this mode is used for longer trips and is therefore unlikely to replace walking. There is little spatial complementarity between the two shared modes and transit ($< 2\%$ of the total trips of each mode). This may be explained by the large overlap between the bikesharing (or carsharing) and transit service areas, as shown in Figure 9.1. However, based on the previous survey-based literature, more trips serving as first-mile or last-mile connection to the transit system were expected, especially for bikesharing. This difference may be attributed to the Montreal context, because the shared modes are concentrated in the urban core and are not very present in the urban periphery (where, according to previous studies, the potential for intermodality is highest), but also to the trip-level method. It is indeed difficult to identify the legs of a same trip using independent trip datasets and the proposed thresholds may be too permissive. Shorter access and egress walking times (i.e., inferior to 5 minutes) could be applied because users have already walked from their actual trip origin to the BIXI station or Communauto vehicle and may have parked closer to connect to the transit stop. Similarly, few trips began outside of the transit service time. However, most of the complementary trips played a role in addressing deficiencies in transit service quality, especially waiting time and travel time too long. Based on the probability densities in Figure 9.4, these trips tend to have a shorter total trip duration and therefore may be more direct and faster than with transit. The condition on distance based on tortuosity did not hold many trips, but this may be because they have already been categorized in the “too many transfers” class which comes before in the algorithm. In the end, about 25% of the bikesharing trips and 40% of the carsharing trips could potentially compete with transit. We do not have comparative works for

carsharing, but the proportion obtained for bikesharing is lower than in another trip-level study on four US cities (Kong et al., 2020). However, the quality of the transit alternative was not considered. These competing trips are longer trips (see Figure 9.4) for which using transit is a good alternative in trip duration, trip distance, waiting time and transfer time. Taking into account trip duration, they represent about 32% of total bikesharing travel hours and 47% of total carsharing travel hours.

Table 9.3 Distribution of the bikesharing and free-floating carsharing trips by interaction class

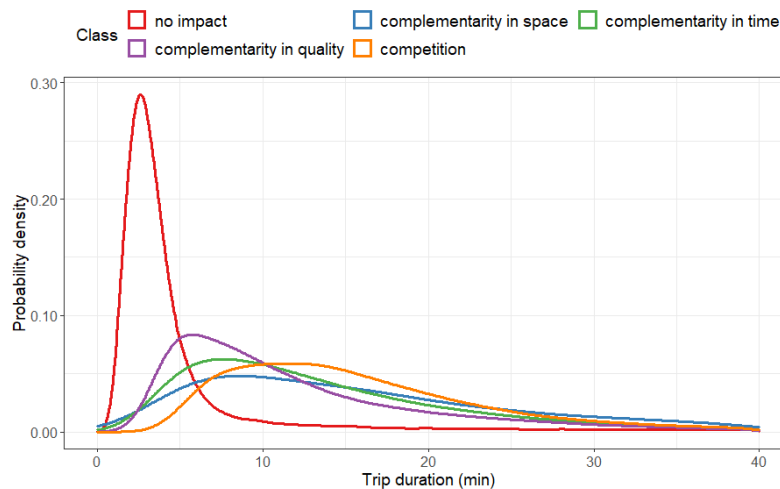
Class	Bikesharing	Carsharing
No impact	11.62%	0.01%
Complementarity in space	1.31%	1.99%
Outside service area	0.002%	0.004%
Last-mile connection	0.36%	0.53%
First-mile connection	0.35%	0.53%
No transit route	0.60%	0.93%
Complementarity in time	5.52%	4.77%
Outside service time	5.52%	4.77%
Complementarity in quality	56.77%	53.03%
Waiting time too long	16.65%	21.15%
Too many transfers	15.81%	11.43%
Too many detours	2.24%	0.94%
Total travel time too long	22.07%	19.51%
Competition	24.78%	40.20%

9.4.2 Temporal analysis

The bikesharing and carsharing trips in each interaction category are also distributed by day (all trips) and hour (week trips only) in Figure 9.5, in parallel with the average trip duration by mode. The relative shares of the interaction classes are quite stable from Monday to Friday, but more differences are observed between weekdays and weekends. The two modes serve more travel needs outside the transit service time on weekends, hence when transit frequency is lower, but also on Fridays. The latter is due to more night trips on this day. Trips for which the transit alternative is not convenient because of waiting time too long are also more likely on weekends. This results in a reduction in the “too many transfers” and “total travel time too long” categories. There are more potential competing trips on weekdays than on weekends for carsharing, but this difference is not clear for bikesharing, although the average trip duration increases on weekends. As expected, most of the trips classified as “outside service time” on weekdays occurred at night, when transit services

are closed or very limited. Complementarity in quality decreases while competition increases during the morning and afternoon peak hours, when transit frequency is at its highest, because more trips can be served by transit within acceptable conditions. The average trip duration also tends to increase for both modes during these hours, making transit more attractive. We can notice that the afternoon peak hour is later for bikesharing than for carsharing, and the midday peak, which is more visible for bikesharing than for carsharing, is also accompanied by an increase in the number of competing trips. Although it is not very apparent on the graph, most of the detected first-mile and last-mile connection using bikesharing also took place during the peak hours. Trips for which Transition did not find an equivalent transit route are mostly observed in early morning and late evening; this may be due to trips made outside the urban core or some confusion with the “outside service time” class.

a)



b)

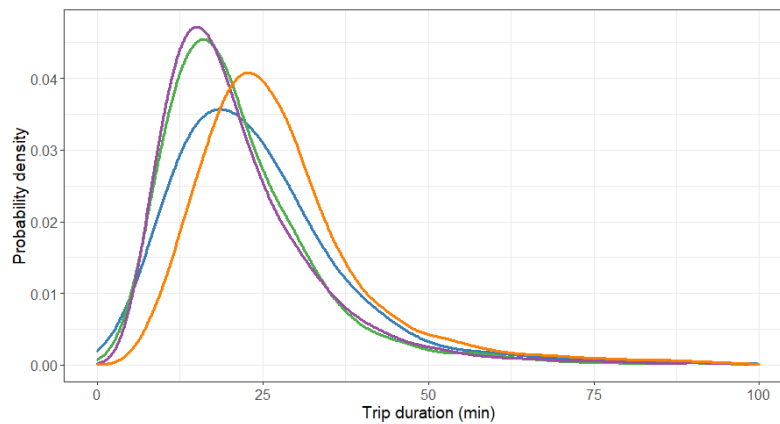


Figure 9.4 Trip duration distribution by interaction category of the a) bikesharing and b) carsharing trips

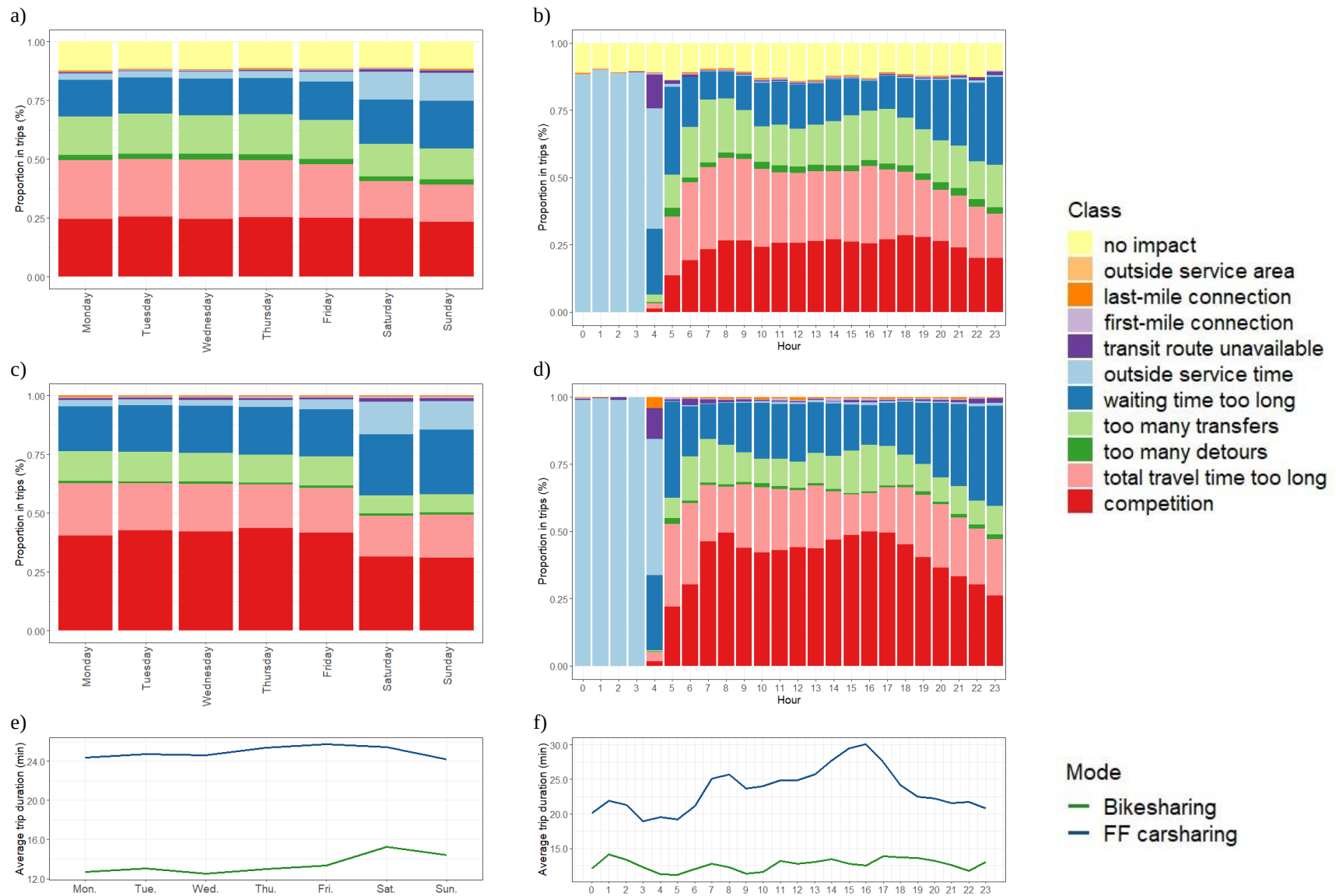
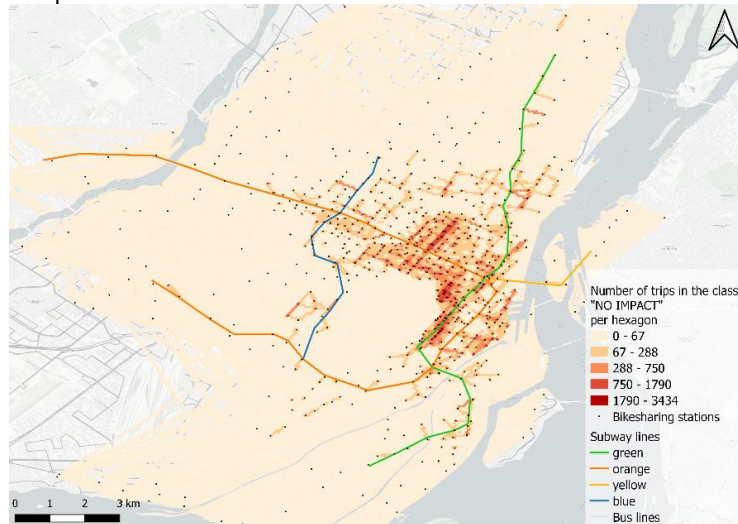


Figure 9.5 Distribution of all trips by day and distribution of all week trips by hour for bikesharing (a and b) and free-floating carsharing (c and d), with average trip duration by day (e) and hour (f)

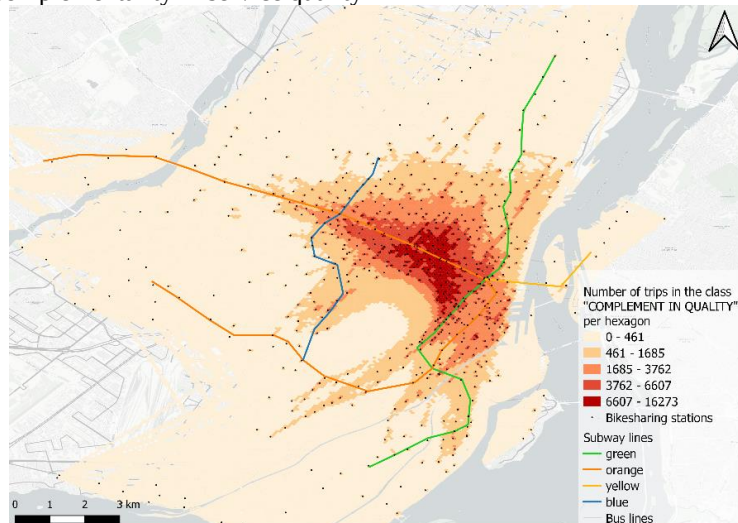
9.4.3 Spatial analysis

Maps of the potential interactions of different categories are displayed in Figure 9.6 for bikesharing and in Figure 9.7 for carsharing. The bikesharing trips in the “no impact” class (potentially replacing walking) are short trips which are mostly located in the city center and in dense, bike-friendly areas, such as the borough called Le Plateau-Mont-Royal. Some of these trips could also have been made for the first/last-mile, especially to connect to/from the subway lines. Figure 9.6 b (complementarity in space) highlights bikesharing stations located more than a 5 min-walk from the transit system: in the East of Montreal, known to be less served by transit, on small islands in the South (Île Sainte-Hélène, Île Notre Dame and Île des Sœurs), but also on a university campus, positioned on top of a hill with quite a long walk to access the subway blue line, and on a point near the green and orange subway lines where the transit network is less dense. The latter point is also detected in the carsharing map (Figure 9.7 a) and thus signals travel needs from/to this area which are not met by the transit system. Other carsharing desire lines come from or go to schools and commercial malls, suggesting trip purposes (dropping off/picking up children or shopping) that are typically more convenient with a car, especially where the transit network is unsatisfactory. Many OD lines also pass through the East. For bikesharing, the highest densities of quality complementary trips (Figure 9.6 c) form a diagonal between the city center and Le Plateau-Mont-Royal. Here, bikesharing may be more convenient and faster than the transit alternative because it avoids a transfer between the subway orange and green lines. It is precisely in this diagonal that a new subway line project is studied by the municipal organization of the city of Montreal. Competition (Figure 9.6 d) is most located in the downtown area (served by a dense transit network and therefore trips could be satisfied by the transit system if the shared modes did not exist), especially along the green subway line. Therefore, this line is the most affected in Figure 9.8. For carsharing, complementarity in quality and competition (Figure 9.7 b and c) are almost located at the same place, with the latter being slightly further south, closer to the city center. For both modes, the transit lines which face more competing trips (based on the transit alternatives to the bikesharing and carsharing trips) are the green and orange subway lines, as well as the parallel bus lines. They are the most saturated lines in Montreal, so this may encourage people to switch to other modes. However, in this case, the competition could also be seen as a kind of complementarity to help unload the transit network.

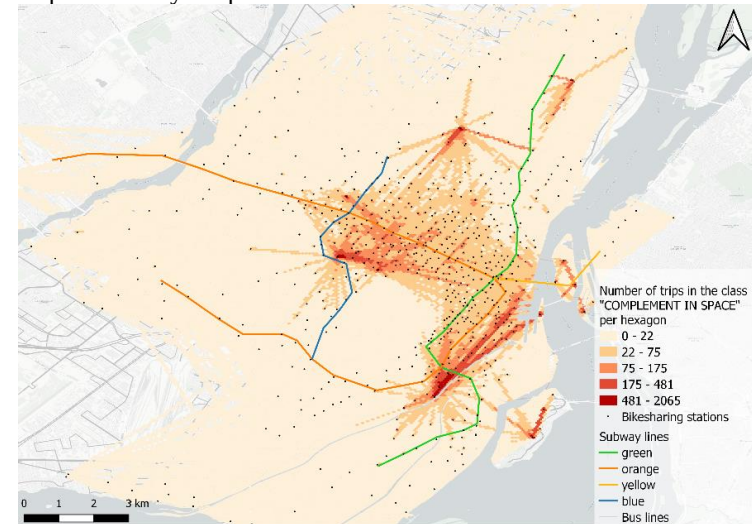
a) No impact



c) Complementarity in service quality



b) Complementarity in space



d) Competition

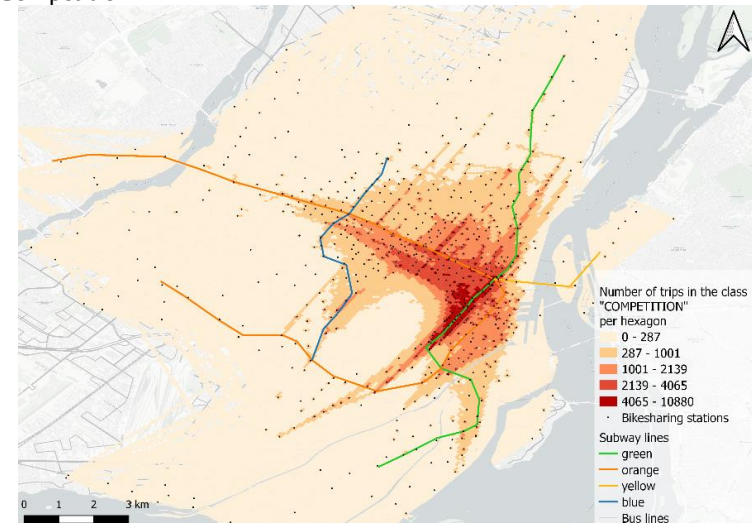
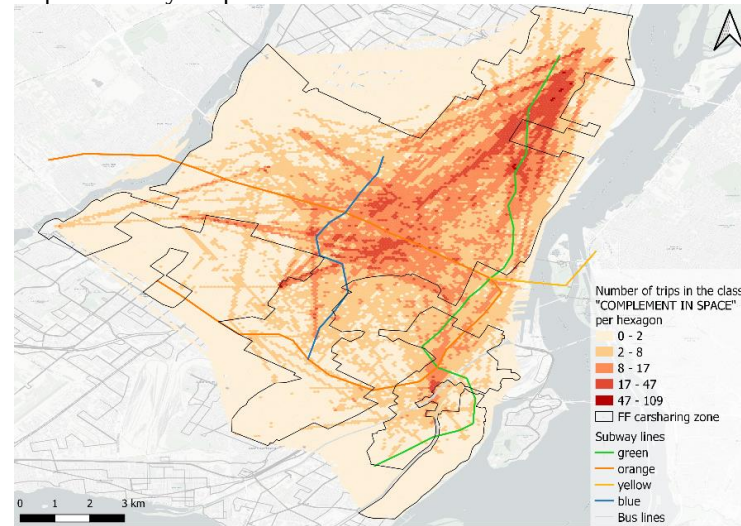
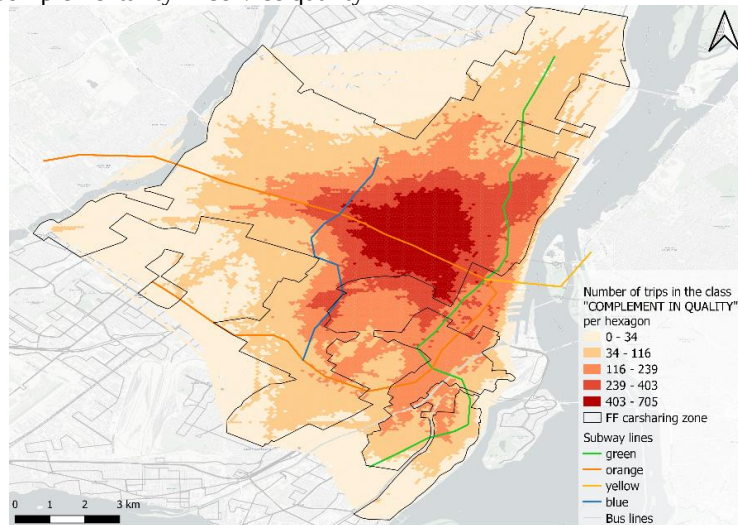


Figure 9.6 Spatial representation in a hexagonal grid (100-m diameter hexagons) of the density of bikesharing Origin-Destination desire lines belonging to different interaction classes: a) no impact, b) complementarity in space, c) complementarity in service quality, d) competition

a) Complementarity in space



b) Complementarity in service quality



c) Competition

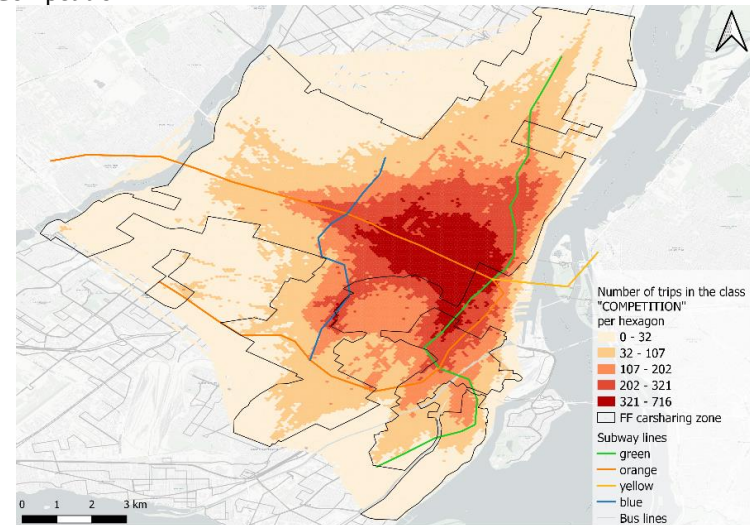


Figure 9.7 Spatial representation in a hexagonal grid (100-m diameter hexagons) of the density of free-floating carsharing Origin-Destination desire lines belonging to different interaction classes: a) complementarity in space, b) complementarity in service quality, c) competition

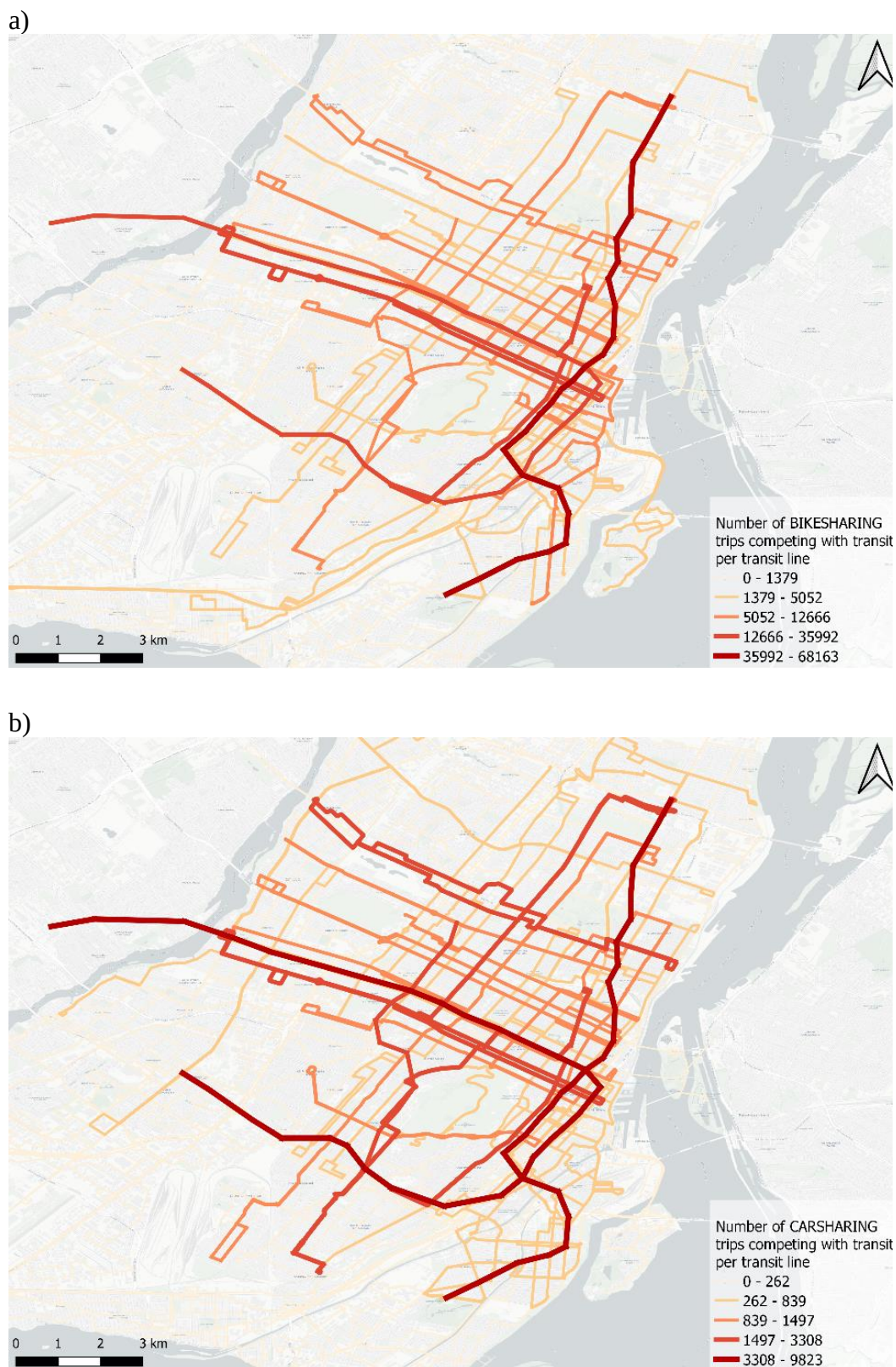


Figure 9.8 Number of a) bikesharing and b) free-floating carsharing trips competing with transit per line based on the transit alternatives calculated by Transition

9.4.4 Modeling results

Finally, the results of the FE-DID model for four treatment definitions are reported in Table 9.4. All models have the same high coefficient of determination ($R^2 = 97\%$), indicating a good fit to the data. However, these models were trained for interpretative purposes only and should not be used for future ridership forecasts as we did not control for overfitting. The control variables have the expected effects, but we are going to focus on the impact of bikesharing, represented by the interaction term between period (September 2021, BIXI open versus November/December 2021, BIXI closed), group (treatment versus control) and bikesharing use. All four treatment measurements led to a significant negative coefficient. This means that bikesharing usage has a negative impact on transit ridership whether the competing relationship is considered or not. On the contrary, the change was found insignificant by Luo, H. et al. (2021) when the competing relationship was not considered (treatment #1 and #2). This may be because the impact was measured at a finer scale (at the stop level), or because the other interactions (with positive and no impacts) do not compensate for the reduction in transit ridership in the Montreal context. The coefficients of the treatment group variable ($Group_i$) are identical in the first three models. This is because the transit lines which belong to the treatment group are very similar. However, they are more different in the fourth treatment definition (considering the sequence of lines used in the transit alternative), and the transit lines in the treatment group are more affected (lower coefficient). In addition, the impact of the treatment intensity ($Period_t * Group_i * BikeUse_{it}$) differs. The coefficient is lower and lower from treatment #1 to treatment #4, indicating an increasing impact on transit ridership. The difference is higher between treatment #3 and #4 (considering the competing relationship) and treatment #1 and #2 (not considering the competing relationship), because the trips with no impact or with a positive impact which could offset the negative impact of the competing trips are removed from the intensity, but the coefficient is also multiplied by 1.2 from treatment #3 to treatment #4. Therefore, this validates the trip classification results (the competing trips did have competing effects) as well as the fact that considering the transit path adds value in evaluating the impact of bikesharing on transit ridership.

Applying the function $x \mapsto 100(e^x - 1)$ to the bikesharing impact coefficient [$100(e^{-1.31E-04} - 1) = 0.013\%$] yields the percentage change in the initial dependent variable (daily transit ridership per line) for one-unit difference in the independent variable. In other words, compared to the

control group, every hundred competing bikesharing trips that would have been made using the transit line caused an additional reduction of 1.3% in daily ridership for the lines in the treatment group. Based on the maximal number of competing trips provided in Table 9.1 (corresponding to the boardings on the green subway line on a given day), this can explain up to 35% of the decrease in daily transit ridership of the impacted line.

Table 9.4 Results of the FE-DID model for four treatment definitions

Variable	Treatment #1	Treatment #2	Treatment #3	Treatment #4
Intercept	-8.74E+00 .	-8.72E+00 .	-7.54E+00	-9.89E+00 .
Treatment variables				
Period	9.52E-03	9.77E-03	6.40E-04	-1.94E-02
Group	-4.05E+00 ***	-4.05E+00 ***	-4.05E+00 ***	-4.12E+00 ***
Period x Group x Bikesharing use	-3.29E-05 ***	-3.31E-05 ***	-1.09E-04 ***	-1.31E-04 ***
Control variables				
Vehicle-kilometers (log)	1.81E+00 ***	1.81E+00 ***	1.81E+00 ***	1.81E+00 ***
Mode	6.74E-01 ***	6.74E-01 ***	6.84E-01 ***	7.58E-01 ***
Population density	5.01E-04 ***	5.01E-04 ***	5.03E-04 ***	5.02E-04 ***
Employment density	6.48E-05 ***	6.48E-05 ***	6.50E-05 ***	6.54E-05 ***
Median household income	-1.02E-04 ***	-1.02E-04 ***	-1.02E-04 ***	-1.02E-04 ***
Average daily temperature	1.14E-04	1.17E-04	1.78E-04	-8.38E-05
Total daily rainfalls	-3.64E-03 ***	-3.64E-03 ***	-3.66E-03 ***	-3.47E-03 ***
Route fixed effects	Yes	Yes	Yes	Yes
Date fixed effects	Yes	Yes	Yes	Yes
Goodness-of-fit				
R ²	0.971	0.971	0.971	0.971

Note: *** 0.01% significance level ($p < 0.001$), ** 1% significance level ($p < 0.01$), * 5% significance level ($p < 0.05$), . 10% significance level ($p < 0.10$).

9.5 Conclusion

The literature review identified several possible interactions between shared modes and transit, but their impacts are confused and were poorly measured, mainly using survey data without spatial-temporal and quantitative details, or at the aggregate (systemic) level with passive data. As a first attempt to clarify the matter, this work proposed a conceptualization by transposing the terminology of biological interactions to define dynamic (long-term) and quasi-static (short-term) interactions between modes based on the neutral, positive or negative impact on ridership. In this paper, for the Montreal case study, the analysis of long-term interactions was not possible because the two modes studied (bikesharing and free-floating carsharing) were introduced several years ago, and we do not have data before and just after their implementation. However, short-term trip-

level relationships were assessed based on one month of disaggregate data. The bikesharing and carsharing trips were first classified into different interaction categories, including the neutral relationship (no interaction), different types of potential complementarities (in time, space or quality) and potential competition. To this end, a rule-based algorithm was applied. The spatial-temporal distributions of the trips of different interaction categories were then analyzed to better understand when and where the shared modes may supplement or substitute transit. Finally, the causal impacts of the bikesharing competing trips on daily route-level transit ridership were quantified with a FE-DID model (using winter as a control period during which the bikesharing service is closed), which is robust to several biases. This also validated the trip classification results and the evaluation was improved by accounting for the transit alternative path calculated by a routing engine.

The findings of the paper can have several policy implications. Time periods, travel demand corridors and types of trips for which travel needs were poorly served by transit, and therefore for which the shared modes were a better option, were highlighted. These results could assist transit operators in improving the existing services and in better encouraging complementarity with the other modes which can fill these gaps. The latter may be a win-win situation (mutualism interaction, i.e., increasing the use of each mode) in the short term with intermodal trips, but also in the long term if it generates a modal shift from private cars. However, the outcome mentioned above can warn the Montreal transit authority against the current poor integration between transit and bikesharing: a negative impact of bikesharing use on transit ridership was measured even when the competing relationship was not considered, suggesting that the other interactions (with no impact or positive impact) do not compensate for the reduction. In particular, the large overlap between the shared modes and transit services in Montreal resulted in low spatial complementarity. The bikesharing station and carsharing vehicle position strategy could be revised to place more bikesharing stations and extend the free-floating zone in less dense areas, where shared modes can provide complementary spatial coverage, while competition is more likely in the city center, especially during the peak hours. Furthermore, payment integration, fare incentives and integrated trip planning platforms could help remove several barriers to the combined use of different modes. Nevertheless, the impacts of the interactions between transit and shared modes were assessed in the short term in this paper, while the positive impacts of complementarity between modes may be more visible in the long term. Complementarity in space (when both origin and destination are

outside the service area), time and quality leads to commensalism (no impact on transit ridership) in the short term but may turn into mutualism (positive impact for the use of both modes) in the long term. Indeed, by filling the spatial-temporal gaps of transit and by offering a better-quality alternative in certain situations (e.g., to avoid transfers), shared modes could be used in synergy with transit. Furthermore, competition could also have a positive impact in the long term if it helps to unload the transit system and reduce transit vulnerability (to service disruption, for instance). This could make transit more reliable and resilient, and thus potentially attract new users. In addition, beyond the concern about loss of transit patronage, from the traveler's perspective, shared modes can improve overall journey quality, and thus user satisfaction.

However, several analytical and methodological limitations of this work should be addressed in future investigations. First, only potential interactions were identified (based on preset rules) since there are no direct links between the independent datasets. Integrated payment methods and data sharing between the operators would be needed to capture actual relationships. In addition, transit and walking were considered as the only alternative modes to bikesharing and carsharing while the two shared modes may also replace other mobility options. Further sensitivity analyses should be conducted to better appreciate the impacts of the many thresholds used in this paper. In particular, the criteria of the trip classification algorithm, representing the users' tolerance in considering the transit alternative as feasible and good enough for their travel need, depend on individual preferences. Mode choice and modal utilities analyses could help better understand them. Moreover, in the historical bikesharing and carsharing data, we do not have the actual trip origin/destination but the trip start/end location, and the distance traveled on foot between them was neglected, leading to potential biases in the results. Other types of complementary situations could have been examined, in travel purpose (e.g., carsharing may be preferred to transit for shopping if the rider has to carry a load), or in relieving the saturated transit services. In the proposed DID model, the treatment measure based on bikesharing use accounted for daily variations and allowed better quantifying the impacts, but this may have caused reverse causality bias: transit ridership can also influence bikesharing use, especially when the network is overcrowded, hence again the importance of considering the capacity of the services. A longer period of data may also be needed to validate the trend assumption. Last, modeling was done at the route level due to available data, but the stop level could provide more detailed results. Bus and subway could also be distinguished.

More work needs to be done to have a broad overview of all the interactions between modes. The negative impact of the competing trips (e.g., in this paper) but also the positive impact (i.e., increase in transit ridership) of the complementary trips should be examined. This study evaluated the impacts of competing relationships on transit demand, but the determinants of the competing relationships could also be explored. Short-term interactions were investigated, but medium-term analyses should also be conducted to capture seasonal interactions (e.g., the change between summer and winter in the relationships between transit and biking), and long-term analyses (several years) should be carried out for the dynamic interactions defined above. Although it was applied to a specific case study, the modeling framework proposed in this paper could be replicated in all these different contexts. Finally, this research could contribute to improving urban mobility and transportation sustainability by leveraging all modes.

Acknowledgments

The authors wish to thank *STM - Société de Transport de Montréal* (Montreal transit authority) and *Communauto* (carsharing operator) for providing us with transit and carsharing data respectively, as well as the *ARTM - Autorité Régionale de Transport Métropolitain* (Metropolitan regional transportation authority) for giving access to the household survey. They also want to thank Geneviève Bastien for her help in handling the Transition platform and coding new features. Moreover, the authors acknowledge the financial support of the Canada Research Chairs (CRC), Quebec Research Funds (FRQNT), Institute for Data Valorization (IVADO) and WCTRS Y-II programs.

CHAPITRE 10 DISCUSSION GÉNÉRALE

Après un bref résumé des principaux résultats obtenus dans chacun des six articles présentés dans les chapitres précédents, les différentes contributions sont mises en commun et discutées en regard des objectifs fixés au début de cette thèse. Ce chapitre montre ainsi en quoi la recherche proposée permet d'augmenter les connaissances sur les interactions entre plusieurs modes de transport, et les aspects méthodologiques associés à leur modélisation à partir de différents ensembles de données sont synthétisés.

10.1 Synthèse des résultats

10.1.1 Article 1 : Comparaison entre enquête OD et données passives

Le premier article a permis d'analyser des variations journalières systémiques dans les comportements de mobilité à partir de deux types de données, à savoir l'enquête Origine-Destination de Montréal (repondérée) et des données passives (nettoyées et sélectionnées sur une période commune), et non l'un ou l'autre comme c'est généralement le cas dans la littérature. Malgré les différences intrinsèques entre ces deux types de données, les indicateurs proposés, basés sur la structure sous-jacente des séries temporelles calculées plutôt que sur les valeurs elles-mêmes, ont permis de mettre en évidence des variations longitudinales similaires dans l'utilisation des modes de transport étudiés. Les mêmes fluctuations (par type de jour, en particulier le jeudi) et les mêmes rythmes périodiques ont été observés, notamment pour l'automobile privée et le métro. Cette analogie est moins probante pour le vélo, ce qui peut être dû à un manque de représentativité des données et/ou à la plus grande irrégularité d'usage de ce mode. Néanmoins, une tendance décroissante semblable a été mesurée au cours de l'automne dans les deux types de données.

Ces résultats permettent, d'une part, de valider la méthode d'expansion développée pour étendre la représentativité de l'enquête Origine-Destination à toute la période de collecte. D'autre part, la variabilité des comportements de déplacement est vérifiée dans les deux sources de données, démontrant ainsi que la notion de journée moyenne n'est pas représentative de la réalité. Enfin, les similarités observées confirment le potentiel de fusion des deux types de données. Ce premier travail légitime l'article suivant (article 2) et justifie l'application des fluctuations extraites des données passives au jour moyen de semaine de l'enquête Origine-Destination.

10.1.2 Article 2 : Fusion entre enquête OD et données passives

Le deuxième article fournit une procédure automatisée de fusion de données entre des enquêtes transversales, basées sur des carnets de déplacements quotidiens, et des flux passifs. Cette méthode, robuste aux valeurs aberrantes et aux données manquantes, repose sur des informations sémantiques agrégées par jour : des comportements moyens d'un côté, et des variations longitudinales de l'autre.

La tendance détectée dans les données passives a d'abord permis de projeter l'utilisation de trois modes (et les parts modales correspondantes) entre deux enquêtes (interpolation entre deux points). Cette projection, produite à partir de l'enquête OD 2013, a été validée avec l'enquête OD 2018. L'erreur entre le nombre de déplacements moyen estimé pour l'automne 2018 et la valeur observée dans l'enquête est très faible, inférieure à celle obtenue avec la méthode de projection du MTQ. De plus, les saisonnalités (hebdomadaire et annuelle) identifiées dans les données passives ont permis d'annualiser le nombre moyen de déplacements par mode pour 2018 avec un profil qui varie en fonction du jour de l'année et du jour de la semaine. Alors que les variations estimées à long terme (sur cinq ans) sont très légères, de plus grandes fluctuations ont été mises en exergue au cours de l'année, et même au cours de l'automne.

Finalement, la méthode proposée permet d'obtenir des informations multimodales (pour plusieurs modes) et longitudinales (en continue sur une ou plusieurs années) grâce à la combinaison des parts modales moyennes de l'enquête OD et des fluctuations observées dans les données passives.

10.1.3 Article 3 : Multimodalité des comportements de mobilité

Le troisième article a permis d'examiner la multimodalité des comportements de mobilité individuels. Une segmentation en 311 combinaisons (ou types) des personnes enquêtées dans l'enquête Origine-Destination a d'abord été créée, puis le niveau de multimodalité de chaque type de personnes a été estimé en 2013 et en 2018 à partir d'une variante de l'indicateur de Dalton. Cet indicateur a été personnalisé de sorte à obtenir une distribution homogène et à mesurer une plus grande variété de niveaux de multimodalité : les intensités d'utilisation des modes (agrégées par type de personnes) ont finalement été évaluées en temps de déplacement, les différents modes de transport en commun ont été considérés séparément, et l'indicateur a été paramétré pour mieux tenir compte des modes les moins utilisés.

Les effets de différents facteurs sur cet indicateur ont ensuite été modélisés, d'abord via un arbre de régression puis avec un modèle DID. Les résultats ont révélé que l'âge et la composition du ménage avaient un impact important sur les niveaux de multimodalité estimés, mais le facteur le plus important est l'accessibilité aux services de transport : une meilleure accessibilité est liée à une plus grande diversité d'utilisation de modes. Par ailleurs, le niveau de multimodalité moyen a augmenté entre 2013 et 2018, et cette évolution peut en partie être expliquée par la bonification de l'offre de transport des modes partagés (transport en commun, vélopartage et autopartage).

10.1.4 Article 4 : Interactions systémiques intermodales

Le quatrième article a modélisé les interactions spatio-temporelles entre le métro de Montréal et son environnement urbain au niveau des stations. Un modèle multiniveau a été construit pas à pas en suivant les étapes prescrites pour ce genre de modèle. En outre, la prise en compte de l'autocorrélation et de l'hétéroscédasticité des résidus, permise par la flexibilité du modèle mixte (avec effets aléatoires), a fourni un meilleur ajustement aux données (comparé aux autres modèles de régression testés).

Le modèle final montre que l'utilisation quotidienne du métro diffère d'une station à l'autre en fonction de plusieurs paramètres liés aux caractéristiques du voisinage des stations (utilisation du sol, composition socio-démographique, localisation de la station dans le réseau). Aussi, le nombre de validations dans chaque station varie dans le temps en fonction du type de jour, certaines stations présentant une plus grande différence entre la semaine et la fin de semaine que d'autres. Des dynamiques pendulaires ont également été mises en évidence en segmentant les déplacements par période de la journée et par groupe de stations. De plus, des facteurs externes comme la tenue d'évènements socio-culturels à proximité des stations et les conditions météo ont des effets significatifs sur la variabilité intra-station.

En contrôlant pour toutes ces variables, des interactions avec d'autres modes ont aussi été détectées dans le modèle. Des corrélations positives entre les entrées dans le métro et les destinations d'autres modes ont été mesurées, signe de l'existence potentielle de relations intermodales aux abords des stations de métro. Cependant, cette complémentarité est surtout visible aux extrémités des lignes de métro, en périphérie de Montréal, alors qu'une certaine forme de compétition est plus probable pour les stations situées au centre-ville.

10.1.5 Article 5 : Impacts de la pandémie de COVID-19

Le cinquième article a proposé une analyse longitudinale et multimodale des impacts de la pandémie de COVID-19. Les résultats de la détection de points de rupture attestent que l'achalandage du métro a été fortement impacté à plusieurs moments. En effet, la diminution du nombre de validations a été abrupte et immédiate suite à la mise en place de trois confinements par le gouvernement québécois. Toutefois, cet effet a été décroissant dans le temps, l'achalandage moyen ne retombant jamais aussi bas que lors du premier confinement, et les périodes de crise ont été entrecoupées par des phases de reprise graduelle. Deux mesures spécifiques au transport en commun ont d'ailleurs probablement eu un impact positif sur le retour des usagers du métro. Finalement, l'augmentation de l'achalandage est lente mais bien visible.

Globalement, les échelles temporelles (type de jour, période de la journée, rythmes périodiques) et les stations qui ont été les plus affectées sont celles qui répondaient le plus à des besoins de déplacement domicile-travail. En effet, les déplacements durant la semaine, en particulier en périodes de pointe du matin et de l'après-midi, sont ceux qui ont le plus diminué. En outre, la crise sanitaire a eu un impact important sur la régularité hebdomadaire des comportements (la période typique de 7 jours est beaucoup moins forte qu'avant la pandémie).

En revanche, d'après les analyses croisées réalisées, les autres modes étudiés (automobile, vélo et autopartage) ont été plus résilients que le métro. En outre, il y a potentiellement eu un transfert modal du métro vers ces modes de transport individuels, notamment le vélo. Dans le modèle de séries temporelles interrompues obtenu, la corrélation systémique entre tous les modes et le métro a diminué au cours des différentes phases de la pandémie, ce qui indique que l'utilisation de ces modes varie de manière moins complémentaire.

10.1.6 Article 6 : Complémentarité et compétition au niveau du déplacement

Le sixième article a précisé les interactions systémiques entre plusieurs modes de transport à une échelle plus désagrégée. D'après la classification proposée, les déplacements en vélopartage et en autopartage (véhicules en libre-service) sont faits en compétition avec le transport en commun (bus et métro) dans environ 25% et 40% des cas. Cette compétition survient surtout en semaine, pendant les heures de pointe, pour des déplacements suffisamment longs qui se concentrent dans le centre-ville. Les lignes les plus affectées sont ainsi les plus saturées. Les modes analysés sont peu

complémentaires dans l'espace du fait du chevauchement des différentes zones de services. Cette disposition a rendu difficile la distinction des déplacements multimodaux (lorsque les modes partagés sont utilisés pour servir le premier ou le dernier kilomètre d'un trajet principalement réalisé en métro) par la méthode proposée. En revanche, des déplacements complémentaires en qualité, pour lesquels les modes partagés sont plus compétitifs que le transport en commun (en termes de temps de déplacement notamment), ont été mis en évidence dans plusieurs corridors de demande. Ainsi, des lacunes dans le service de transport en commun et des opportunités de complémentarité avec les autres modes pour combler ces lacunes ont été identifiées.

L'impact de l'usage du vélopartage sur le nombre de validations par jour et par ligne a également été modélisé. Cet impact s'est révélé négatif dans tous les scénarios de traitement analysés, ce qui peut signifier que les modes ne sont pas suffisamment combinés entre eux pour compenser les déplacements en transport en commun remplacés par le vélopartage, mais il est encore plus important lorsqu'on considère seulement les déplacements en compétition. Cela valide donc les résultats de l'algorithme de classification. De plus, la prise en compte du trajet qui aurait été emprunté en l'absence du service du vélopartage améliore la mesure de l'impact de ces déplacements. Chaque centaine de déplacements en vélopartage pourrait entraîner une diminution moyenne de 1,3% du nombre de montées sur les lignes les plus sujettes à la compétition (par rapport aux lignes du groupe contrôle).

10.2 Discussion générale des résultats

Somme toute, les résultats résumés précédemment permettent de répondre aux objectifs formulés dans la section 1.3 et de combler ainsi plusieurs lacunes dans la littérature. Les sections suivantes rappellent en quoi les six articles proposés permettent d'intégrer des sources diversifiées de données, de qualifier et de quantifier différents types d'interactions entre les modes, ainsi que de modéliser ces interactions.

10.2.1 Intégration de sources diversifiées de données

Premièrement, cette recherche est basée sur la valorisation et la combinaison de différents ensembles de données provenant à la fois d'enquêtes et de flux passifs. Cette condition était nécessaire afin de pouvoir analyser plusieurs modes tant au niveau des comportements de mobilité individuels que des systèmes de transport complets, avec des informations socio-démographiques

pour contextualiser les interactions mais aussi des détails spatio-temporels pour explorer leur variabilité. Plusieurs enquêtes Origine-Destination de Montréal ont été mises en correspondance via des variables socio-démographiques communes ayant permis la création de types de personnes qui se répètent d'une enquête à l'autre (méthode d'appariement de l'article 3). De même, différents flux de données passives ont été combinés grâce à des unités d'agrégation spatio-temporelle communes, par exemple par jour et par station (article 4) ou par jour et par ligne (article 6).

Enfin, les deux types de données (enquêtes et flux passifs) ont été fusionnés dans un objectif de suivi longitudinal de l'utilisation de plusieurs modes de transport et des parts modales correspondantes (article 2). Ce type de fusion entre les deux grands principaux types de données du domaine des transports est rare dans la littérature, et encore plus dans la direction où ce sont les données d'enquêtes qui sont enrichies par les données passives (et non l'inverse). Ici, les informations transmises d'un type de données à l'autre (des données passives aux enquêtes) sont des informations temporelles liées à la variabilité (long terme et saisonnière) des comportements de déplacement, largement reconnue par la communauté scientifique mais non prise en compte dans le jour moyen obtenu grâce aux enquêtes. 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.

Toutefois, ces intégrations de différentes sources d'informations ont généralement conduit à agréger les données et à considérer des variations journalières systémiques seulement, alors que la variabilité des comportements de déplacement est surtout visible au niveau individuel (variabilité inter et intra-individuelle). En outre, la méthode de fusion de données proposée peut être remise en question avec la pandémie de COVID-19 : des cassures importantes dans la tendance sont à prévoir, et il ne sera probablement même plus possible d'identifier une tendance long terme du fait des changements rapides survenus pendant la crise sanitaire. C'est pourquoi ces impacts ont été examinés dans l'article 5. Pour continuer à être applicable, la méthode de fusion proposée devrait donc intégrer les effets de rupture liés à la pandémie dans une composante supplémentaire.

10.2.2 Qualification et quantification de différents types d'interactions entre les modes

Les différents types d'interactions étudiés dans cette thèse sont résumés dans le Tableau 10.1 ci-dessous (par article), en fonction du niveau d'analyse (individu, système, déplacement), de la période durant laquelle les interactions ont été examinées (court terme versus moyen/long terme) et en fonction des modes considérés. Dans les analyses effectuées à partir des enquêtes Origine-Destination de Montréal (articles 1, 2 et 3), seuls des modes classiques ont été étudiés car les nouveaux services de mobilité ne sont pas distingués ou sont sous-représentés dans ces données. Dans les analyses des interactions au niveau systémique à partir de données passives (articles 4, 5 et 6), l'emphasis a été mise sur les relations entre le transport en commun (surtout le métro, en tant que mode structurant par excellence dans les villes) et des modes partagés (vélopartage et autopartage). Néanmoins, les relations intermodales entre le métro et le taxi ont aussi été traitées dans l'article 4. De plus, les impacts de la pandémie de COVID-19 sur la corrélation entre le métro et des modes classiques individuels (automobile privée et vélo) ont été modélisés dans l'article 5.

Tableau 10.1 Synthèse des interactions analysées dans chaque article

(N° article) Thème	Niveau			Période		Modes de transport							
	Individu	Système complet	Déplacement	Court terme	Moyen/long terme	Métro	Bus	Automobile privée	Taxi	Vélo	Vélopartage	Autopartage VLS	Autopartage stations
(1) Comparaison EOD/données passives		X		X		X		X		X			
(2) Fusion EOD/données passives		X			X	X		X		X			
(3) Multimodalité des comportements	X				X	X	X	X	X	X			
(4) Interactions systémiques intermodales		X		X		X			X		X	X	
(5) Impacts de la pandémie de COVID-19		X			X	X		X		X		X	X
(6) Interactions au niveau du déplacement			X	X		X	X				X	X	

La combinaison de plusieurs ensembles de données a généralement mené à agréger les données et donc à des résultats au niveau systémique. Toutefois, des analyses plus désagrégées par type de personnes (article 3) et par déplacement (article 6) ont aussi été proposées. De même, les interactions modales ont souvent été étudiées à court terme à cause d'un manque de données avant et après l'arrivée des nouveaux modes. En effet, les premiers services de vélopartage et d'autopartage de Montréal ont été mis en place il y a plusieurs années (en 2009 pour BIXI et en 1995 pour Communauto) et les systèmes de collecte de données passives n'étaient alors pas aussi développés qu'actuellement. Par conséquent, des interactions quasi-statiques et leurs variations journalières ont été évaluées sur quelques semaines (articles 4 et 6), et la période de fermeture de BIXI pendant l'hiver a servi de période contrôle pour mesurer un effet de causalité (article 6). Toutefois, l'évolution des parts modales (article 2) et de la multimodalité des comportements de mobilité (article 3) a été examinée à plus long terme (sur cinq ans), de même que les impacts de la pandémie (article 5) ont été analysés longitudinalement entre début 2019 et fin 2021 (sur trois années complètes).

Ces différents types d'interactions entre les modes ont également été quantifiés. Au niveau individuel, un niveau de multimodalité a été estimé (article 3), tandis que des proportions de déplacements dans différentes catégories de complémentarité ou de compétition entre les modes ont été calculées au niveau du déplacement (article 6). Au niveau du système, des corrélations ont été mesurées et analysées dans le temps et dans l'espace. Ces corrélations ont été évaluées en fonction de l'intensité d'utilisation des modes, et non seulement avec une variable binaire indiquant la présence d'un nouveau service de mobilité comme c'est souvent le cas dans la littérature. Des travaux précédents avaient déjà été réalisés pour étudier des corrélations d'usage entre plusieurs modes au cours d'une année complète et par MIAZ (zone d'analyse d'interaction de la mobilité) dans le contexte montréalais (Morency et al., 2020; Morency et al., 2018). Cependant, cette thèse propose des analyses complémentaires : des parts modales ont été mises en relation sur cinq ans (article 2), des changements dans les corrélations entre les volumes de plusieurs modes ont été examinés au cours du temps pendant la pandémie de COVID-19 (article 5), et la corrélation entre les destinations d'un mode et les origines d'un autre mode a été utilisée comme proxy pour estimer des relations intermodales (article 4). En outre, les origines des déplacements de modes partagés ont aussi été croisées avec les entrées dans le transport en commun pour mesurer l'impact de la compétition entre les modes (article 6).

En résumé, les résultats concernant ces différents types d'interactions permettent de conclure que les niveaux de multimodalité ont augmenté avec la bonification des services de transport dans la région de Montréal au cours des dernières années, ce qui insinue que les comportements de mobilité sont plus diversifiés (utilisation d'une plus grande variété de modes) et moins auto-dépendants (article 3). L'autopartage est le mode qui a eu l'impact le plus important sur cette évolution et est donc peut-être un vecteur du changement de comportement. Un fort potentiel d'intermodalité a également été montré entre le métro et d'autres modes en périphérie du réseau (article 4), et une forme de complémentarité saisonnière entre le métro et le vélo peut être soupçonnée dans les parts modales des deux modes : l'utilisation du vélo augmente et celle du transport en commun diminue en été, tandis que l'inverse se produit en hiver, suggérant une combinaison des deux modes par certains individus au cours de l'année (article 2). De plus, des situations de complémentarité (dans le temps, l'espace ou la qualité) entre le transport en commun et les modes partagés ont été mises en évidence dans les comportements de déplacement (article 6). L'impact positif de cette complémentarité sur l'achalandage du transport en commun est peu visible à court terme, mais des analyses à plus long terme devraient permettre de déceler un bénéfice pour l'utilisation de tous les modes (mutualisme). En effet, en comblant les lacunes spatio-temporelles du transport en commun et en offrant une alternative de meilleure qualité dans certaines situations (par exemple, pour éviter des correspondances), les modes partagés sont susceptibles d'être utilisés en synergie avec le transport en commun et d'entraîner un transfert modal depuis l'automobile. Il s'agirait alors d'un transfert modal de l'automobile vers un cocktail modal (plutôt que de l'automobile vers le transport en commun uniquement), engendrant une augmentation du niveau de multimodalité telle qu'observée dans la région de Montréal au cours des dernières années (article 2). Cependant, la complémentarité spatiale entre les modes peut encore être améliorée car les zones de service actuelles se chevauchent (article 6). En outre, des situations de compétition ont aussi été identifiées entre les modes partagés et le transport en commun (article 6). Toutefois, cette compétition à court-moyen terme pourrait également être bénéfique à plus long terme si elle permet de décharger les systèmes de transport en commun lorsqu'ils sont saturés, et de réduire leur vulnérabilité (aux interruptions de service par exemple) en fournissant des alternatives. En rendant le service plus fiable et plus résilient, cette compétition pourrait finalement attirer de nouveaux utilisateurs vers le transport en commun. Néanmoins, toutes ces conclusions devront être revalidées car les

interactions entre les modes pourraient avoir été fortement impactées par la pandémie de COVID-19 (article 5).

10.2.3 Modélisation des interactions entre les modes

Au niveau méthodologique, deux grandes familles de modèles ont été utilisées pour expliquer, contextualiser les interactions entre les modes et évaluer leurs impacts sur la demande de transport. Tout d'abord, des méthodes de modélisation longitudinale, basées sur des séries temporelles, ont été appliquées, notamment l'algorithme de décomposition STL (article 2) et des modèles multiniveaux à effets fixes et/ou aléatoires. Les effets fixes ont servi à réduire les biais d'omission de variables non observées dans l'estimation d'un impact causal (article 6), mais l'ajout d'effets aléatoires a fourni une certaine flexibilité dans la modélisation des résidus qui a permis d'améliorer l'ajustement aux données (article 4). De plus, différents indicateurs ont été employés pour extraire et comparer des phénomènes structurels (tendance, variations saisonnières, périodicité) dans des données longitudinales (article 1).

Par ailleurs, des méthodes d'inférence causale ont également été exploitées dans des contextes non expérimentaux et non aléatoires. Des effets systémiques à l'échelle d'une population, par exemple l'impact de la pandémie de COVID-19, ont été évalués à l'aide de méthodes de détection de points de rupture et d'un modèle de séries temporelles interrompues (article 5). En outre, des impacts plus locaux, basés sur des différences entre un groupe traité et un groupe contrôle, ont été mesurés via la méthode des doubles différences. Cette dernière a été appliquée dans sa forme la plus simple, entre deux groupes et deux périodes (article 3), mais aussi dans sa forme généralisée avec des données longitudinales (article 6). Dans les deux cas, l'effet estimé (interaction entre une variable de groupement et une variable de période) a été pondéré par un indicateur d'intensité du traitement.

Ces différentes méthodes ont permis d'analyser la variabilité des interactions modales dans le temps, entre différentes phases ou différents moments de la journée, mais aussi dans l'espace, en fonction de la localisation des stations par exemple. De surcroît, les impacts de différents facteurs sur ces interactions, notamment les attributs socio-démographiques et l'évolution de l'offre de transport (article 3), ou les mesures et événements liés à la crise sanitaire (article 5), ont pu être modélisés. Ces facteurs ont aussi parfois été introduits en tant que variables contrôles dans les modèles car ils peuvent également expliquer l'utilisation des modes. Par ailleurs, les impacts de ces interactions sur la demande de transport ont été évalués en termes de corrélation (article 4) ou

de vraie relation de causalité (article 6). Cet impact est positif lorsqu'il s'agit de relations intermodales (article 4), négatif lorsqu'il y a substitution d'un mode par une autre (article 6), ou nul à court terme mais positif à long terme lorsqu'un mode permet de combler les lacunes d'un autre mode, par exemple en termes de couverture spatio-temporelle.

CHAPITRE 11 CONCLUSION ET RECOMMANDATIONS

Pour conclure, les résultats discutés précédemment contribuent à mieux comprendre, à évaluer et à modéliser (en corrélation et en causalité) différents types d'interactions entre plusieurs modes de transport (classiques et émergents) à partir de deux principaux formats de données (des enquêtes ménages déplacements et des flux passifs) et de leur fusion. Néanmoins, les analyses proposées restent limitées par certaines barrières, et de nombreuses pistes d'amélioration et travaux futurs peuvent être suggérés. Dans ce dernier chapitre, les contributions, limites et perspectives associées à chaque article sont listées dans des tableaux de synthèse, puis discutées de manière plus globale dans chacune des sections suivantes.

11.1 Contributions

Les contributions aux niveaux analytique, méthodologique et tactique/stratégique de chaque article sont résumées dans le Tableau 11.1. Tout d'abord, cette thèse met à disposition un cadre conceptuel original pour définir différents types de relations entre les modes de transport. Ce squelette permet ainsi de regrouper toutes les interactions modales qui ont été examinées de manière éparsée dans la littérature. De plus, il permet d'expliquer les résultats contradictoires parfois observés d'une étude à l'autre par des différences dans les niveaux d'agrégation et les échelles spatio-temporelles. Les analyses produites contribuent finalement à augmenter les connaissances sur les interactions entre les modes, les comportements associés, ainsi que leur variabilité dans le temps et dans l'espace. De nombreux modes et services de mobilité ont pu être considérés grâce à la grande quantité de données rassemblées pour cette recherche.

Au niveau méthodologique, cette thèse fournit des outils pour évaluer le niveau de multimodalité, de corrélation ou de complémentarité/compétition entre les modes. Ces outils prennent la forme d'indicateurs, d'algorithmes de classification, mais aussi de modèles longitudinaux (basés sur des séries temporelles) et de méthodes d'évaluation des impacts (basées sur des corrélations ou des inférences causales). Par ailleurs, différents défis méthodologiques liés aux données et à leur fusion ont été relevés. Au-delà d'une simple comparaison descriptive, ces données ont été valorisées et enrichies dans l'objectif de répondre à un besoin d'informations longitudinales et multimodales. Ainsi, cette recherche a présenté un cas d'application de la complémentarité entre des données d'enquêtes et des flux passifs pour le suivi de différents indicateurs de mobilité multimodale.

Tableau 11.1 Synthèse des contributions de la thèse

(N° article) Thème	Contributions		
	Analytiques	Méthodologiques	Tactiques/stratégiques
(1) Comparaison EOD/données passives	Analyse croisée de la variabilité des comportements de mobilité à partir de deux types de données	- Développement d'une méthode d'expansion longitudinale de l'enquête OD - Proposition d'indicateurs de comparaison systématique entre deux types de données	- Suivi des indicateurs de mobilité au cours de la période d'automne - Validation des données de l'enquête OD pour la modélisation des comportements de déplacement
(2) Fusion EOD/données passives	Analyse longitudinale des parts modales de plusieurs modes de transport	Développement d'une méthode de fusion de données	Suivi des indicateurs de mobilité entre deux enquêtes ménages déplacements
(3) Multimodalité des comportements de mobilité	Augmentation des connaissances sur les comportements de multimodalité	- Élaboration d'une méthode d'appariement par type de personnes - Quantification du niveau de multimodalité (indicateur) - Modélisation de l'impact causal d'une amélioration des services de transport	Évaluation de l'impact d'une stratégie de bonification de l'offre de transport
(4) Interactions systémiques intermodales	Analyse spatio-temporelle des relations intermodales dans le voisinage des stations de métro	- Combinaison de plusieurs sources de données en deux dimensions (espace et temps) - Modélisation des effets de l'environnement urbain en tenant compte de la structure hiérarchique du réseau de métro - Modélisation des termes d'erreurs	Aide pour mieux intégrer les nouveaux services de mobilité avec le métro

Tableau 11.1 Synthèse des contributions de la thèse (suite et fin)

(N° article) Thème	Contributions		
	Analytiques	Méthodologiques	Tactiques/stratégiques
(5) Impacts de la pandémie de COVID-19	Analyse longitudinale et multimodale des impacts de la pandémie	Identification, quantification et modélisation des impacts	- Contribution à une meilleure compréhension des changements dus à la pandémie - Soutien pour améliorer la résilience du TC dans d'autres situations de crise et encourager la reprise de l'achalandage
(6) Complémentarité et compétition au niveau du déplacement	Analyse désagrégée (au niveau du déplacement) de différentes situations de complémentarité ou de compétition entre les modes	- Développement d'un algorithme de classification des interactions (avec utilisation d'un calculateur de chemin) - Modélisation de l'impact causal de la compétition entre le vélopartage et le transport en commun	- Informations sur les interactions entre les modes et leurs impacts sur la demande de transport - Identification d'opportunités pour mieux intégrer les modes partagés avec le transport en commun

Au niveau décisionnel, les travaux proposés permettent d'informer les opérateurs (niveau tactique) et les planificateurs (niveau stratégique) des systèmes de transport sur l'évolution des comportements de mobilité (augmentation du niveau de multimodalité, changements dus à la pandémie de COVID-19, etc.), mais aussi sur les opportunités de complémentarité entre les modes. Une meilleure intégration entre le transport en commun et les nouveaux services de mobilité est possible et souhaitable car elle permettrait entre autres de retirer des déplacements automobiles. Cette intégration peut se faire dans l'espace, par le biais d'un positionnement stratégique des infrastructures, dans les horaires des services, et/ou dans les cas où le transport en commun n'est pas compétitif. Ces différentes situations de complémentarité sont gagnantes pour les opérateurs puisqu'elles pourront se traduire en une augmentation d'utilisation des modes impliqués, à court terme grâce à des relations intermodales, et/ou à long terme en créant une synergie. Pour les faciliter, une plus grande collaboration entre les parties prenantes est néanmoins nécessaire. De plus, des décisions stratégiques doivent être prises en faveur de politiques et d'interventions visant

à augmenter l'offre de transport des modes alternatifs, qui permet une augmentation des niveaux de multimodalité, et à améliorer différentes formes d'intégration (opérationnelle, informationnelle, tarifaire, technologique, etc.). Finalement, ce projet participe à encourager les différents acteurs du transport à travailler collectivement dans le but de réduire la dépendance à l'automobile et d'assurer un développement durable de la mobilité.

11.2 Limites

Les limites de chacun des articles sont regroupées dans le Tableau 11.2. Elles sont séparées en deux parties, selon si elles se rapportent aux données ou à la méthodologie.

Tableau 11.2 Synthèse des limites de la thèse

(N° article) Thème	Limites	
	Données	Méthodologie
(1) Comparaison EOD/données passives	• Période d'enquête courte • Manque d'informations sur la fin de semaine • Représentativité des données du vélo	• Analyses des variations au niveau systémique/agrégé
(2) Fusion EOD/données passives	• Sous-estimation des nouveaux services de mobilité dans les enquêtes	• Formulation paramétrique • Hypothèse de linéarité locale • Méthode tendancielle moins pertinente dans un contexte de pandémie
(3) Multimodalité des comportements de mobilité	• Une seule journée enquêtée par individu • Manque d'informations sur la marche combinée à d'autres modes	• Mesure d'effets agrégés • Hypothèse de la méthode des doubles différences • Rayon de 500 mètres à valider par des analyses de sensibilité
(4) Interactions systémiques intermodales	• Indépendance des bases de données → relations intermodales analysées en termes de corrélations seulement	• Variance inter-station mal expliquée dans le modèle • Sensibilité de la définition du « voisinage » autour des stations à vérifier • Non prise en compte de l'autocorrélation spatiale • Hypothèse de linéarité des effets

Tableau 11.2 Synthèse des limites de la thèse (suite et fin)

(N° article) Thème	Limites	
	Données	Méthodologie
(5) Impacts de la pandémie de COVID-19	• Représentativité des données de comptage	• Impacts évalués en corrélation seulement (et non en causalité) • Impossibilité d'isoler l'impact de chaque mesure • Interactions mesurées comme des corrélations systémiques globales • Non prise en compte de l'évolution de l'offre de transport
(6) Complémentarité et compétition au niveau du déplacement	• Indépendance des bases de données → complémentarité ou compétition <u>potentielle</u> seulement • Pas de données avant/après l'implantation des modes partagés • Données du bus non géolocalisées → modélisation à l'échelle d'une ligne (versus arrêt ou station)	• Beaucoup de paramètres et de valeurs seuils • Amalgame entre origine/destination et début/fin du déplacement • Biais de causalité inverse possible • Hypothèse de la méthode des doubles différences

Au-delà des enjeux habituels de représentativité des informations recueillies, les données exploitées dans cette recherche ont des lacunes qui ont parfois limité les analyses possibles. Tout d'abord, l'agrégation nécessaire pour faire correspondre les différentes bases de données entre elles a contraint à réaliser la majorité des analyses au niveau systémique. De même, l'indépendance des ensembles de données utilisés a souvent conduit à mesurer des interactions potentielles seulement. En particulier, les relations intermodales (plusieurs modes dans un même déplacement) sont difficiles à détecter dans des données passives puisque chaque segment de déplacement fait avec un mode différent est collecté dans une base de données distincte. L'intermodalité au niveau individuel n'a pas non plus pu être étudiée car très peu de déplacements multimodaux sont disponibles dans les données des enquêtes Origine-Destination de Montréal (environ 2% des déplacements de l'enquête de 2018 combinent des modes de transport autres que des modes de transport en commun entre eux). De plus, la seule journée d'observation par individu et le manque d'informations sur les services de mobilité émergents et sur la marche lorsqu'elle est combinée avec d'autres modes ont restreint l'analyse de la multimodalité des comportements. Aussi, l'indisponibilité des données avant et après l'implantation des modes partagés à Montréal a empêché de mesurer les effets des interactions entre les modes à long terme. Seuls des impacts temporaires (à court terme) ont pu être évalués grâce à une période de fermeture des services. Ainsi,

l'impact positif d'une synergie entre les modes sur leur utilisation (complémentarité à long terme) n'a pas pu être estimé. Certains modes (Uber par exemple) n'ont également pas pu être inclus du tout dans cette thèse du fait de l'absence de données accessibles.

Au niveau méthodologique, les limites de cette recherche sont essentiellement dues aux hypothèses des méthodes, notamment à l'hypothèse de linéarité des modèles appliqués. En effet, ces derniers sont principalement des modèles de régression linéaire améliorés (ou généralisés) pour prendre en compte l'autocorrélation des données longitudinales et les effets confondants de variables non observées qui nuisent à la mesure de la causalité. Néanmoins, cette lacune pourrait être dépassée avec des transformations de variables ou des modèles non paramétriques (des algorithmes d'apprentissage machine par exemple). En outre, beaucoup de seuils préétablis ont été utilisés pour définir le voisinage des objets étudiés (autour du lieu de domicile des personnes, ou autour des stations/arrêts de transport en commun) ou pour dépeindre les préférences des usagers en matière de choix modal. De même, des seuils ont été nécessaires pour distinguer le groupe traité du groupe contrôle dans les applications de la méthode DID. Des analyses de sensibilité plus poussées devraient être effectuées pour évaluer les impacts de ces seuils sur les résultats obtenus, ou des méthodes spécifiques, par exemple un modèle d'arbres de décision entraînés par *gradient boosting* (Ding et al., 2019), pourraient être appliquées pour déterminer ces seuils de manière objective et automatique. Par ailleurs, les interactions entre les modes ont été modélisées dans cette thèse avec une approche « bottom-up », c'est-à-dire que l'utilisation de chaque mode a été examinée séparément, puis les résultats ont été sommés pour produire des parts modales. Une démarche alternative (« top-down ») consisterait à prévoir le nombre total de déplacements (une courbe enveloppe de la mobilité) puis à les distribuer par mode via la modélisation de proportions plutôt que de volumes. Cette approche intégrerait donc explicitement les effets de répartition de la demande entre les modes.

Pour finir, les analyses proposées plaçaient généralement le transport en commun comme le mode central des interactions alors qu'il peut aussi servir de mode d'accès. Par exemple, un véhicule d'autopartage peut être emprunté pour atteindre un arrêt de transport en commun, mais l'interaction contraire est aussi possible (utilisation du transport en commun pour aller chercher une voiture partagée). De même, le vélopartage peut influencer l'usage du transport en commun, et réciproquement le transport en commun peut inciter l'utilisation du vélopartage, par exemple lorsque le service est saturé. Par conséquent, bien que celle-ci ait permis de mieux évaluer

l'ampleur de l'interaction (en termes d'impact sur la demande), l'étude des relations entre les modes en termes d'utilisation (avec des volumes) est sujette à un possible biais de causalité inverse.

11.3 Perspectives

Outre le dépassement des limites soulignées précédemment, de nombreuses pistes d'exploration futures sont envisageables. Les perspectives analytiques et méthodologiques déjà énoncées dans chaque article sont rappelées dans le Tableau 11.3. De manière plus générale, la thématique abordée dans cette thèse est très large et cette première exploration fournit une preuve de concept qui ouvre la voie pour de nombreux autres travaux. En particulier, chacun des types d'interactions définis pourrait faire l'objet d'un sujet de recherche à part entière.

Une enquête spécifique (sur plusieurs jours, au moins une semaine) pourrait être menée afin de mieux comprendre les interactions modales au niveau individuel. Cette enquête pourrait collecter les déplacements de manière passive, par exemple via une application mobile, mais les modes devront alors être estimés à partir des traces GPS et des vitesses imputées (Yang, M. et al., 2022). Toutefois, une longue période de collecte (ou au moins des enquêtes répétées par panel) serait nécessaire pour examiner les interactions à long terme. Globalement, tout format de données permettant le suivi longitudinal des individus et/ou la collecte de déplacements complets (plutôt que de segments de déplacements) sur un réseau de transport multimodal serait pertinent pour pouvoir réaliser des analyses plus désagrégées et étudier de vraies interactions modales (et non seulement des interactions potentielles). Une autre alternative serait de faire communiquer les différents ensembles de données passives recueillis de manière indépendante par un lien direct, par exemple grâce un moyen de paiement unique (une carte de crédit ou toute autre carte à puce) permettant d'accéder à tous les modes (Ma, Xinwei et al., 2018), à condition que les différents opérateurs des services acceptent ensuite de se partager les données. Dans tous les cas, une meilleure collaboration entre les acteurs du transport est souhaitable pour pouvoir approfondir l'analyse des interactions entre les modes.

Des investigations supplémentaires doivent également être réalisées pour pouvoir modéliser les interactions de tous les modes en même temps. Dans cette thèse, les relations ont été examinées entre deux modes à la fois, sans considérer les autres. En outre, les analyses proposées se sont centrées sur l'utilisation des modes (demande), mais les interactions pourraient aussi être évaluées

en termes d'offre : la complémentarité/compétition entre les modes pourrait être estimée en fonction de l'espace alloué aux infrastructures ou en fonction de la synchronisation des services, et l'indicateur de diversité proposé précédemment pourrait être basé sur les niveaux de service plutôt que sur les intensités d'utilisation de chaque mode afin de quantifier un niveau d'offre multimodale. De tels indicateurs de performance permettraient de comparer plusieurs villes entre elles. Par ailleurs, cette recherche a permis de mesurer l'impact d'une bonification de l'offre au niveau systémique, mais des effets plus locaux pourraient être explorés. Pour ce faire, il est recommandé de cibler certains éléments du réseau seulement (par exemple via un projet pilote) avant d'étendre l'intervention à tout le réseau. En effet, la comparaison d'informations avant/après entre le groupe ciblé et un groupe contrôle permet une évaluation moins biaisée de l'impact (en causalité) grâce aux méthodes présentées dans ce document. Les effets d'une meilleure intégration entre les modes devraient également être mesurés. La complémentarité temporelle et spatiale (dans les heures de service et les zones de couverture) est importante, mais les impacts d'autres formes d'intégration, telles que l'intégration tarifaire (impacts de la création de partenariats entre les modes) ou opérationnelle (impacts de l'implantation d'infrastructures, des stationnements vélo par exemple, permettant de faciliter les correspondances), devraient également être étudiés.

Enfin, des recherches plus approfondies sont requises pour évaluer les changements dans les interactions modales dus à la pandémie de COVID-19. En particulier, même si certains modes semblent avoir été préférés à d'autres (notamment les modes individuels par rapport aux modes collectifs et partagés), il sera important de déterminer s'il s'agit d'une compétition temporaire et conjoncturelle dans les choix modaux individuels (court terme) ou de vrais transferts modaux (long terme). Des effets rebonds liés à l'augmentation de la congestion automobile pourraient également être visibles. Le questionnaire amorcé sur le rôle des différents modes dans la nouvelle normale est une opportunité pour améliorer leur intégration. En plus d'aider la reprise de l'achalandage du transport en commun, une plus grande complémentarité entre les modes serait bénéfique aux opérateurs des autres services de mobilité dans deux situations : en tant que moyens d'accès pour la desserte d'un mode de transport lourd (dont les volumes transportés restent plus élevés), et en tant que partenaires du cocktail modal pouvant entrer en compétition avec l'automobile privée.

Tableau 11.3 Synthèse des perspectives de la thèse

(N° article) Thème	Perspectives	
	Analytiques	Méthodologiques
(1) Comparaison EOD/données passives	Analyser la variabilité d'utilisation des différents modes au niveau individuel (inter et intrapersonnelle)	- Combiner les deux types de données en se basant sur des périodicités communes (plutôt que sur des variables communes) - Fusionner les deux types de données pour étendre l'enquête OD au-delà de l'automne
(2) Fusion EOD/données passives	- Appliquer la méthode avec d'autres modes de transport - Réaliser des analyses plus désagrégées : par zone ou station (dimension spatiale), par type de titre de transport et pour certaines heures de la journée	- Développer des facteurs de variabilité pour corriger l'enquête OD de Montréal - Modéliser les résidus ou ajouter une composante supplémentaire dans la décomposition - Développer une méthode de prévision de la série temporelle désaisonnalisée
(3) Multimodalité des comportements de mobilité	- Analyser la multimodalité individuelle (par personne plutôt que par type de personnes) avec d'autres données - Appliquer l'indicateur proposé pour mesurer la multimodalité de l'offre de transport (basé sur des niveaux de service plutôt que des niveaux d'utilisation par mode) - Construire un modèle d'inférence causale à un niveau plus agrégé (systémique), par zone spatiale	- Hiérarchiser les modes en fonction de leur niveau de durabilité dans l'indicateur de multimodalité - Développer des seuils de niveau de service minimal pour qu'il y ait un effet sur les comportements
(4) Interactions systémiques intermodales	- Analyser les relations intermodales du dernier kilomètre (plutôt que du premier kilomètre) en inférant les sorties du métro - Explorer d'autres types de complémentarité entre les modes à un niveau plus désagrégé	- Construire un modèle sur une plus longue période - Prendre en compte la capacité des systèmes (effet de la saturation du métro au centre-ville et de l'indisponibilité des services partagés sur les relations modales)

Tableau 11.3 Synthèse des perspectives de la thèse (suite et fin)

(N° article) Thème	Perspectives	
	Analytiques	Méthodologiques
(5) Impacts de la pandémie de COVID-19	• Mettre à jour les résultats pour suivre l'évolution des changements • Inclure d'autres modes de transport • Étudier les impacts par type de produit tarifaire ou de manière encore plus désagrégée • Croiser les résultats avec des données d'enquête pour mieux comprendre les changements dans les comportements individuels • Examiner les impacts sur certains types d'interactions spécifiques	• Développer des scénarios de projection des changements dans les systèmes d'activités et les comportements de déplacement pour prévoir la demande de transport future
(6) Complémentarité et compétition au niveau du déplacement	• Inclure d'autres modes de transport • Analyser d'autres types de complémentarité (par motif de déplacement par exemple) • Évaluer l'impact causal positif des déplacements complémentaires	• Mesurer les déterminants des relations de complémentarité ou de compétition entre les modes • Modéliser les interactions en distinguant les impacts à court, moyen et long terme

RÉFÉRENCES

- Abduljabbar, R. L., Liyanage, S., & Dia, H. (2021). The role of micro-mobility in shaping sustainable cities: A systematic literature review. *Transportation Research Part D: Transport and Environment*, 92, 102734. doi:10.1016/j.trd.2021.102734
- Agard, B., Morency, C., & Trépanier, M. (2006). Mining public transport user behaviour from smart card data. *IFAC Proceedings Volumes*, 39(3), 399 - 404. doi:10.3182/20060517-3-FR-2903.00211
- Alam, F., Mehmood, R., Katib, I., Albogami, N. N., & Albeshri, A. (2017). Data Fusion and IoT for Smart Ubiquitous Environments: A Survey. *IEEE Access*, 5, 9533-9554. doi:10.1109/ACCESS.2017.2697839
- Alexander, L., Murga, M., González, M. C., & Jiang, S. (2015). Origin-destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C: Emerging Technologies*, 58, 240-250. doi:10.1016/j.trc.2015.02.018
- Allison, P. D. (2009). *Fixed effects regression models*: SAGE publications.
- Alsger, A., Tavassoli, A., Mesbah, M., Ferreira, L., & Hickman, M. (2018). Public transport trip purpose inference using smart card fare data. *Transportation Research Part C*, 87, 123-137. doi:10.1016/j.trc.2017.12.016
- Aminikhanghahi, S., & Cook, D. J. (2017). A Survey of Methods for Time Series Change Point Detection. *Knowledge and information systems*, 51(2), 339-367. doi:10.1007/s10115-016-0987-z
- 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.
- An, D., Tong, X., Liu, K., & Chan, E. H. W. (2019). Understanding the impact of built environment on metro ridership using open source in Shanghai. *Cities*, 93, 177-187. doi:10.1016/j.cities.2019.05.013
- Anderson, M. L. (2014). Subways, Strikes, and Slowdowns: The Impacts of Public Transit on Traffic Congestion. *The American Economic Review*, 104(9), 2763-2796. doi:10.1257/aer.104.9.2763
- ARTM. (2020). COVID-19 et transport collectif. Tiré de <https://www.artm.quebec/covid-19-et-transport-collectif/>
- ARTM. (2022). Vers de nouvelles mobilités. Tiré de <https://www.artm.quebec/grands-projets/nouvelles-mobilites/>
- Asparouhov, T., Hamaker, E. L., & Muthen, B. (2018). Dynamic Structural Equation Models. *Structural Equation Modeling-A Multidisciplinary Journal*, 25(3), 359-388. doi:10.1080/10705511.2017.1406803
- Astroza, S., Bhat, P. C., Bhat, C. R., Pendyala, R. M., & Garikapati, V. M. (2018). Understanding activity engagement across weekdays and weekend days: A multivariate multiple discrete-

- continuous modeling approach. *Journal of Choice Modelling*, 28, 56-70. doi:10.1016/j.jocm.2018.05.004
- Atmani, D. (2015). *Affectation dynamique dans les systèmes de transport multimodaux*. (Université Paris-Est). Tiré de <https://theses.hal.science/tel-01294620/document>
- Austin, P. C., & Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine*, 34(28), 3661-3679. doi:10.1002/sim.6607
- Babar, Y., & Burtch, G. (2020). Examining the Heterogeneous Impact of Ride-Hailing Services on Public Transit Use. *Information Systems Research*, 31(3), 820-834. doi:10.1287/isre.2019.0917
- Badr, H. S., Du, H., Marshall, M., Dong, E., Squire, M. M., & Gardner, L. M. (2020). Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. *The Lancet. Infectious diseases*, 20(11), 1247-1254. doi:10.1016/S1473-3099(20)30553-3
- Bahbouch, K., Wagner, J. R., Morency, C., & Berdier, C. (2017). Travel demand corridors: Modelling approach and relevance in the planning process. *Journal of Transport Geography*, 58, 196-208. doi:10.1016/j.jtrangeo.2016.12.007
- Bai, J., & Perron, P. (2003). Computation and Analysis of Multiple Structural Change Models. *Journal of Applied Econometrics*, 18(1), 1-22.
- Basu, R., & Ferreira, J. (2021). Sustainable mobility in auto-dominated Metro Boston: Challenges and opportunities post-COVID-19. *Transport Policy*, 103, 197-210. doi:10.1016/j.tranpol.2021.01.006
- Bauernschuster, S., Hener, T., & Rainer, H. (2017). When Labor Disputes Bring Cities to a Standstill: The Impact of Public Transit Strikes on Traffic, Accidents, Air Pollution, and Health. *American Economic Journal: Economic Policy*, 9(1), 1-37. doi:10.1257/pol.20150414
- 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* (p. 587-611): Emerald Group Publishing Limited.
- Beaudoin, J., Farzin, Y. H., & Lin Lawell, C. Y. C. (2015). Public transit investment and sustainable transportation: A review of studies of transit's impact on traffic congestion and air quality. *Research in Transportation Economics*, 52, 15-22. doi:10.1016/j.retrec.2015.10.004
- Beaudoin, J., & Lin Lawell, C. Y. C. (2018). The effects of public transit supply on the demand for automobile travel. *Journal of Environmental Economics and Management*, 88, 447-467. doi:10.1016/j.jeem.2018.01.007
- Beck, M. J., & Hensher, D. A. (2020). Insights into the impact of COVID-19 on household travel and activities in Australia - The early days under restrictions. *Transport Policy*, 96, 76-93. doi:10.1016/j.tranpol.2020.07.001

- Beck, M. J., Hensher, D. A., & Wei, E. (2020). Slowly coming out of COVID-19 restrictions in Australia: Implications for working from home and commuting trips by car and public transport. *Journal of Transport Geography*, 88, 102846. doi:10.1016/j.jtrangeo.2020.102846
- Becker, H., Ciari, F., & Axhausen, K. W. (2017). Comparing car-sharing schemes in Switzerland: User groups and usage patterns. *Transportation Research Part A: Policy and Practice*, 97, 17-29. doi:10.1016/j.tra.2017.01.004
- Becker, H., Ciari, F., & Axhausen, K. W. (2018). Measuring the car ownership impact of free-floating car-sharing – A case study in Basel, Switzerland. *Transportation Research Part D: Transport and Environment*, 65, 51-62. doi:10.1016/j.trd.2018.08.003
- Behrens, R., Freedman, M., & McGuckin, N. (2009). The Challenges of Surveying ‘Hard to Reach’ Groups: Synthesis of a Workshop. Dans P. Bonnel, M. Lee-Gosselin, J. Zmud, & J.-L. Madre (édit.), *Transport Survey Methods* (p. 145-152): Emerald Group Publishing Limited.
- Ben-Akiva, M., McFadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., . . . Kth. (2002). Hybrid Choice Models: Progress and Challenges. *Marketing Letters*, 13(3), 163-175. doi:10.1023/A:1020254301302
- Bergeron, U. (2022). Montréal pourrait donner une deuxième chance aux trottinettes en libre-service. *Le Devoir*. Tiré de <https://www.ledevoir.com/societe/transports-urbanisme/686646/transport-montreal-pourrait-donner-une-deuxieme-chance-aux-trottinettes-en-libre-service>
- Bergström, A., & Magnusson, R. (2003). Potential of transferring car trips to bicycle during winter. *Transportation Research Part A*, 37(8), 649-666. doi:10.1016/S0965-8564(03)00012-0
- Bernal, J. L., Cummins, S., & Gasparrini, A. (2017). Interrupted time series regression for the evaluation of public health interventions: a tutorial. *International journal of epidemiology*, 46(1), 348-355. doi:10.1093/ije/dyw098
- Bhat, C. R. (2015). A new generalized heterogeneous data model (GHDM) to jointly model mixed types of dependent variables. *Transportation Research Part B*, 79, 50-77. doi:10.1016/j.trb.2015.05.017
- Bhat, C. R., & Castelar, S. (2002). A unified mixed logit framework for modeling revealed and stated preferences: formulation and application to congestion pricing analysis in the San Francisco Bay area. *Transportation Research Part B*, 36(7), 593-616. doi:10.1016/S0191-2615(01)00020-0
- Bian, Z., Zuo, F., Gao, J., Chen, Y., Pavuluri Venkata, S. S. C., Duran Bernardes, S., . . . Wang, J. (2021). Time lag effects of COVID-19 policies on transportation systems: A comparative study of New York City and Seattle. *Transportation Research Part A: Policy and Practice*, 145, 269-283. doi:10.1016/j.tra.2021.01.019
- Bieliński, T., Kwapisz, A., & Ważna, A. (2021). Electric bike-sharing services mode substitution for driving, public transit, and cycling. *Transportation Research Part D: Transport and Environment*, 96, 102883. doi:10.1016/j.trd.2021.102883

- Bigazzi, A., & Wong, K. (2020). Electric bicycle mode substitution for driving, public transit, conventional cycling, and walking. *Transportation Research Part D: Transport and Environment*, 85, 102412. doi:10.1016/j.trd.2020.102412
- Bischoff, J., Rodier, C. J., Pourrahmani, E., Jaller, M., Pahwa, A., & Maciejewski, M. (2019). *Competition among Automated Taxis, Transit, and Conventional Passenger Vehicles: Traffic Effects in the San Francisco Bay Area*. Communication présentée à 98th Annual Meeting of the Transportation Research Board, Washington D.C, United States.
- BIXI. (2021). Open Data, Trip History, Year 2021. Tiré de <https://bixi.com/en/open-data>
- Boisjoly, G., Grisé, E., Maguire, M., Veillette, M. P., Deboosere, R., Berrebi, E., & El-Geneidy, A. (2018). Invest in the ride: A 14 year longitudinal analysis of the determinants of public transport ridership in 25 North American cities. *Transportation Research Part A: Policy and Practice*, 116, 434-445. doi:10.1016/j.tra.2018.07.005
- Bonnel, P. (2004). *Prévoir la demande de transport: Patrick Bonnel*. Paris: Presses de l'École nationale des ponts et chaussées.
- Bonnel, P., Bayart, C., & Smith, B. (2015). Workshop Synthesis: Comparing and Combining Survey Modes. *Transportation Research Procedia*, 11, 108-117. doi:10.1016/j.trpro.2015.12.010
- 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
- Bösch, P. M., Becker, F., Becker, H., & Axhausen, K. W. (2018). Cost-based analysis of autonomous mobility services. *Transport Policy*, 64, 76-91. doi:10.1016/j.tranpol.2017.09.005
- Boström, H., Andler, S. F., Brohede, M., Johansson, R., Karlsson, A., Van Laere, J., . . . Ziemke, T. (2007). On the definition of information fusion as a field of research. Skövde: Institutionen för kommunikation och information.
- Bourbonnais, P.-L. (2019). trRouting: transit routing engine. Tiré de <https://github.com/kaligrafy/trRouting>
- Bourbonnais, P.-L., Faucher, J., Morency, C., & Trépanier, M. (2019). Transition: transit network optimization and simulation platform. Tiré de <http://transition.city/>
- Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control* (5^e éd.). Hoboken, New Jersey: John Wiley & Sons.
- Bradley, M. A., & Daly, A. J. (1997). Estimation of logit choice models using mixed stated-preference and revealed-preference information. Dans P. Stopher & M. Lee-Gosselin (édit.), *Understanding travel behaviour in an era of change* (p. 209-232). Howard House, Wagon Lane, Bingley: Esmerald Group Publishing Limited.
- Briand, A. S., Côme, E., Trépanier, M., & Oukhellou, L. (2017). Analyzing year-to-year changes in public transport passenger behaviour using smart card data. *Transportation Research Part C: Emerging Technologies*, 79, 274 - 289. doi:10.1016/j.trc.2017.03.021
- Brockwell, P. J., & Davis, R. A. (2016). *Introduction to time series and forecasting* (3^e éd.): Springer.

- Broere, W. (2016). Urban underground space: Solving the problems of today's cities. *Tunnelling and Underground Space Technology*, 55, 245-248. doi:10.1016/j.tust.2015.11.012
- Bucsky, P. (2020). Modal share changes due to COVID-19: The case of Budapest. *Transportation Research Interdisciplinary Perspectives*, 8, 100141. doi:10.1016/j.trip.2020.100141
- Buehler, R., & Hamre, A. (2014). An examination of recent trends in multimodal travel behavior among American motorists. *International Journal of Sustainable Transportation*, 10(4), 354-364. doi:10.1080/15568318.2014.945672
- Buehler, R., & Hamre, A. (2015). The multimodal majority? Driving, walking, cycling, and public transportation use among American adults. *Transportation*, 42(6), 1081-1101. doi:10.1007/s11116-014-9556-z
- Buehler, R., & Pucher, J. (2021). COVID-19 Impacts on Cycling, 2019–2020. *Transport Reviews*, 41(4), 393-400. doi:10.1080/01441647.2021.1914900
- Bunn, A., Korpela, M., Biondi, F., Campelo, F., Mérian, P., Qeadan, F., & Zang, C. (2022). dplR: Dendrochronology Program Library in R (Version R package version 1.7.4). Tiré de <https://CRAN.R-project.org/package=dplR>
- Bwambale, A., Choudhury, C. F., Hess, S., & Iqbal, M. S. (2021). Getting the best of both worlds: a framework for combining disaggregate travel survey data and aggregate mobile phone data for trip generation modelling. *Transportation*, 48(5), 2287-2314. doi:10.1007/s11116-020-10129-5
- Cairns, S., Behrendt, F., Raffo, D., Beaumont, C., & Kiefer, C. (2017). Electrically-assisted bikes: Potential impacts on travel behaviour. *Transportation Research Part A: Policy and Practice*, 103, 327-342. doi:10.1016/j.tra.2017.03.007
- 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.
- Campbell, K. B., & Brakewood, C. (2017). Sharing riders: How bikesharing impacts bus ridership in New York City. *Transportation Research Part A: Policy and Practice*, 100, 264-282. doi:10.1016/j.tra.2017.04.017
- Cao, L., Fang, Q., Zhang, D., & Chen, T. (2018). Subway station construction using combined shield and shallow tunnelling method: Case study of Gaojiayuan station in Beijing. *Tunnelling and Underground Space Technology*, 82, 627-635. doi:10.1016/j.tust.2018.09.010
- Carrel, A., Vij, A., & Walker, J. L. (2011). *Understanding multimodal behavior: individual modality styles and their influence on travel behavior*. Communication présentée à 90th Transportation Research Board Annual Meeting, Washington, DC.
- Castanedo, F. (2013). A Review of Data Fusion Techniques. *Scientific World Journal*. doi:10.1155/2013/704504
- Ceccato, R., Chicco, A., & Diana, M. (2021). Evaluating car-sharing switching rates from traditional transport means through logit models and Random Forest classifiers. *Transportation Planning and Technology*, 44(2), 160-175. doi:10.1080/03081060.2020.1868084

- Ceccato, R., & Diana, M. (2021). Substitution and complementarity patterns between traditional transport means and car sharing: a person and trip level analysis. *Transportation*, 48, 1523–1540. doi:10.1007/s11116-018-9901-8
- Centre for Research on Impact Evaluation. (2015). *Quasi-experimental methods: Propensity Score Matching and Difference in Differences*. Competence Centre on Microeconomic Evaluation (CC-ME). Tiré de <https://profcjgodlewski.files.wordpress.com/2016/09/2amatchinganddiffdiffhavari.pdf>
- Cervero, R. (2002). Induced travel demand: Research design, empirical evidence, and normative policies. *Journal of Planning Literature*, 17(1), 3-20.
- Cervero, R. (2003). Road Expansion, Urban Growth, and Induced Travel: A Path Analysis. *Journal of the American Planning Association*, 69(2), 145-163. doi:10.1080/01944360308976303
- Cervero, R., Caldwell, B., & Cuellar, J. (2013). Bike-and-ride: build it and they will come. *Journal of Public Transportation*, 16(4), 5.
- Cervero, R., Golub, A., & Nee, B. (2007). City CarShare: Longer-term travel demand and car ownership impacts. *TRANSPORTATION RESEARCH RECORD*, (1992), 70-80. doi:10.3141/1992-09
- Cervero, R., & Hansen, M. (2002). Induced Travel Demand and Induced Road Investment: A Simultaneous Equation Analysis. *Journal of Transport Economics and Policy (JTEP)*, 36(3), 469-490.
- Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment*, 2(3), 199-219. doi:10.1016/S1361-9209(97)00009-6
- Chen, C., Feng, T., Gu, X., & Yao, B. (2022). Investigating the effectiveness of COVID-19 pandemic countermeasures on the use of public transport: A case study of The Netherlands. *Transport Policy*, 117, 98-107. doi:10.1016/j.tranpol.2022.01.005
- 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
- Cherchi, E., & Ortúzar, J. de D. (2002). Mixed RP/SP models incorporating interaction effects. *Transportation*, 29(4), 371-395. doi:10.1023/A:1016307308723
- Chiang, W.-C., Russell, R. A., & Urban, T. L. (2011). Forecasting ridership for a metropolitan transit authority. *Transportation Research Part A: Policy and Practice*, 45(7), 696-705. doi:10.1016/j.tra.2011.04.016
- Chikaraishi, M., Zhang, J., Fujiwara, A., & Axhausen, K. W. (2010). Exploring Variation Properties of Time Use Behavior on the Basis of a Multilevel Multiple Discrete–Continuous Extreme Value Model. *Transportation Research Record*, 2156(1), 101-110. doi:10.3141/2156-12
- Chiou, Y.-C., Jou, R.-C., & Yang, C.-H. (2015). Factors affecting public transportation usage rate: Geographically weighted regression. *Transportation Research Part A: Policy and Practice*, 78, 161-177. doi:10.1016/j.tra.2015.05.016

- Chlond, B., Wirtz, M., & Zumkeller, D. (2013). Data Quality and Completeness Issues in Multiday or Panel Surveys. Dans J. Zmud, M. Lee-Gosselin, M. Munizaga, & J. A. Carrasco (édit.), *Transport Survey Methods* (p. 373-392): Emerald Group Publishing Limited.
- Chow, J. Y. J., & Djavadian, S. (2015). Activity-based market equilibrium for capacitated multimodal transport systems. *Transportation Research Part C*, 59, 2-18. doi:10.1016/j.trc.2015.04.028
- Chu, X. (2004). *Ridership models at the stop level*. National Center of Transit Research, University of South Florida.
- 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.
- Clewlow, R. R. (2016). Carsharing and sustainable travel behavior: Results from the San Francisco Bay Area. *Transport Policy*, 51, 158-164. doi:10.1016/j.tranpol.2016.01.013
- Clewlow, R. R., & Mishra, G. S. (2017). *Disruptive transportation: The adoption, utilization, and impacts of ride-hailing in the United States*. UC Davis: Institute of Transportation Studies. Tiré de <https://escholarship.org/uc/item/82w2z91j>
- Clifton, K., & Muhs, C. D. (2012). Capturing and Representing Multimodal Trips in Travel Surveys: Review of the Practice. *Transportation Research Record*, 2285(1), 74-83. doi:10.3141/2285-09
- Commission européenne. (2006). *Pour une Europe en mouvement - Mobilité durable pour notre continent. Examen à mi-parcours du livre blanc sur les transports publié en 2001 par la Commission européenne*. Bruxelles: Tiré de [http://www.europarl.europa.eu/meetdocs/2004_2009/documents/com/com_com\(2006\)0314/com_com\(2006\)0314_fr.pdf](http://www.europarl.europa.eu/meetdocs/2004_2009/documents/com/com_com(2006)0314/com_com(2006)0314_fr.pdf)
- Contreras, S. D., & Paz, A. (2018). The effects of ride-hailing companies on the taxicab industry in Las Vegas, Nevada. *Transportation Research Part A: Policy and Practice*, 115, 63-70. doi:10.1016/j.tra.2017.11.008
- Cucinotta, D., & Vanelli, M. (2020). WHO Declares COVID-19 a Pandemic. *Acta bio-medica : Atenei Parmensis*, 91(1), 157-160. doi:10.23750/abm.v91i1.9397
- Cui, J., Broere, W., & Lin, D. (2021). Underground space utilisation for urban renewal. *Tunnelling and Underground Space Technology*, 108, 103726. doi:10.1016/j.tust.2020.103726
- Cui, J., & Lin, D. (2016). Utilisation of underground pedestrian systems for urban sustainability. *Tunnelling and Underground Space Technology*, 55, 194-204. doi:10.1016/j.tust.2015.11.004
- Cui, J., & Nelson, J. D. (2019). Underground transport: An overview. *Tunnelling and Underground Space Technology*, 87, 122-126. doi:10.1016/j.tust.2019.01.003
- D'Ambrosio, A., Aria, M., & Siciliano, R. (2007). *Robust Tree-Based Incremental Imputation Method for Data Fusion*. Communication présentée à 7th International Symposium on Intelligent Data Analysis, Ljubljana, Slovenia.
- D'Orazio, M., Di Zio, M., & Scanu, M. (2006). *Statistical matching: Theory and practice*. New York: John Wiley & Sons.

- Dagum, E. B., & Bianconcini, S. (2016). *Seasonal adjustment methods and real time trend-cycle estimation*: Springer.
- Das, S., Boruah, A., Banerjee, A., Raoniar, R., Nama, S., & Maurya, A. K. (2021). Impact of COVID-19: A radical modal shift from public to private transport mode. *Transport Policy*, 109, 1-11. doi:10.1016/j.tranpol.2021.05.005
- Dasarathy, B. V. (1997). Sensor fusion potential exploitation-innovative architectures and illustrative applications. *Proceedings of the IEEE*, 85(1), 24-38. doi:10.1109/5.554206
- Davidson, A., Peters, J., & Brakewood, C. (2017). *Interactive travel modes: Uber, transit, and mobility in New York City*. Communication présentée à Transportation Research Board 96th Annual Meeting, Washington, DC, USA.
- De Gruyter, C. (2019). Multimodal Trip Generation from Land Use Developments: International Synthesis and Future Directions. *Transportation Research Record*, 2673(3), 136-152.
- de Haas, M., Faber, R., & Hamersma, M. (2020). How COVID-19 and the Dutch ‘intelligent lockdown’ change activities, work and travel behaviour: Evidence from longitudinal data in the Netherlands. *Transportation Research Interdisciplinary Perspectives*, 6. doi:10.1016/j.trip.2020.100150
- De Haas, M., Scheepers, E., & Hoogendoorn-Lanser, S. (2017). *Identifying different types of observed immobility within longitudinal travel surveys*. Communication présentée à International Conference on Transport Survey Methods. Estérel, Canada.
- De Lorimier, A., & El-Geneidy, A. M. (2013). Understanding the Factors Affecting Vehicle Usage and Availability in Carsharing Networks: A Case Study of Communauto Carsharing System from Montréal, Canada. *International Journal of Sustainable Transportation*, 7(1), 35-51. doi:10.1080/15568318.2012.660104
- de Stasio, C., Fiorello, D., & Maffii, S. (2011). Public transport accessibility through co-modality: Are interconnectivity indicators good enough? *Research in Transportation Business & Management*, 2, 48-56. doi:10.1016/j.rtbm.2011.07.003
- De Witte, A., Hollevoet, J., Dobruszkes, F., Hubert, M., Macharis, C., Handelshögskolan, . . . Law. (2013). Linking modal choice to motility: A comprehensive review. *Transportation Research Part A*, 49, 329-341. doi:10.1016/j.tra.2013.01.009
- Deschaintres, E. (2018). *Analyse de la variabilité individuelle d'utilisation du transport en commun à l'aide de données de cartes à puce*. (École Polytechnique de Montréal, Montréal, Canada). Tiré de <https://publications.polymtl.ca/3284/>
- Deschaintres, E., Morency, C., & Trépanier, M. (2019). Analyzing transit user behavior with 51 weeks of smart card data. *Transportation Research Record*, 2673(6), 33-45. doi:10.1177/0361198119834917
- Deschaintres, E., Morency, C., & Trépanier, M. (2021). Measuring Changes in Multimodal Travel Behavior Resulting from Transport Supply Improvement. *Transportation Research Record*, 2675(9), 533-546. doi:10.1177/03611981211003104
- Deschaintres, E., Morency, C., & Trépanier, M. (2022). Cross-analysis of the variability of travel behaviors using one-day trip diaries and longitudinal data. *Transportation Research Part A: Policy and Practice*, 163, 228-246. doi:10.1016/j.tra.2022.07.013

- Di Ciommo, F., & Shiftan, Y. (2017). Transport equity analysis. *Transport Reviews*, 37(2), 139-151. doi:10.1080/01441647.2017.1278647
- Diab, E., DeWeese, J., Chaloux, N., & El-Geneidy, A. (2020). Adjusting the service? Understanding the factors affecting bus ridership over time at the route level in Montréal, Canada. *Transportation*. doi:10.1007/s11116-020-10147-3
- Diab, E., Kasraian, D., Miller, E. J., & Shalaby, A. (2020). The rise and fall of transit ridership across Canada: Understanding the determinants. *Transport Policy*, 96, 101-112. doi:10.1016/j.tranpol.2020.07.002
- Diab, E., & Shalaby, A. (2018). Subway Service Down Again? Assessing the Effects of Subway Service Interruptions on Local Surface Transit Performance. 2672(8), 443-454. doi:10.1177/0361198118791665
- Diana, M. (2010). From mode choice to modal diversion: A new behavioural paradigm and an application to the study of the demand for innovative transport services. *Technological Forecasting & Social Change*, 77(3), 429-441. doi:10.1016/j.techfore.2009.10.005
- Diana, M. (2012). Studying Patterns of Use of Transport Modes Through Data Mining Application to US National Household Travel Survey Data Set. *Transportation Research Record*, 2308(2308), 1-9. doi:10.3141/2308-01
- Diana, M., & Mokhtarian, P. L. (2009). Desire to change one's multimodality and its relationship to the use of different transport means. *Transportation Research Part F*, 12(2), 107. doi:10.1016/j.trf.2008.09.001
- Diana, M., & Pirra, M. (2016). A comparative assessment of synthetic indices to measure multimodality behaviours. *Transportmetrica A: Transport Science*, 12(9), 771-793. doi:10.1080/23249935.2016.1177133
- Diana, M., Pirra, M., Woodcock, A., & Martins, S. (2018). *Supporting Urban Integrated Transport Systems: Transferable tools for Local Authorities (SUITS)*. Communication présentée à 7th Transport Research Arena (TRA), Vienna, Austria.
- Diao, M., Kong, H., & Zhao, J. (2021). Impacts of transportation network companies on urban mobility. *Nature Sustainability*, 4(6), 494-500. doi:10.1038/s41893-020-00678-z
- Dibbelt, J., Pajor, T., Strasser, B., & Wagner, D. (2013// 2013). *Intriguingly Simple and Fast Transit Routing*. Communication présentée à International Symposium, SEA 2013, Rome, Italy (vol. 7933, p. 43-54). doi:10.1007/978-3-642-38527-8_6
- Ding, C., Cao, X., & Liu, C. (2019). How does the station-area built environment influence Metrorail ridership? Using gradient boosting decision trees to identify non-linear thresholds. *Journal of Transport Geography*, 77, 70-78. doi:10.1016/j.jtrangeo.2019.04.011
- Ding, C., Duan, J., Zhang, Y., Wu, X., & Yu, G. (2018). Using an ARIMA-GARCH Modeling Approach to Improve Subway Short-Term Ridership Forecasting Accounting for Dynamic Volatility. *IEEE Transactions on Intelligent Transportation Systems*, 19(4), 1054-1064. doi:10.1109/TITS.2017.2711046

- Ding, C., Wang, D., Ma, X., & Li, H. (2016). Predicting Short-Term Subway Ridership and Prioritizing Its Influential Factors Using Gradient Boosting Decision Trees. *Sustainability*, 8(11), 1100. doi:10.3390/su8111100
- Dong, X., Guerra, E., & Daziano, R. A. (2021). Impact of TNC on travel behavior and mode choice: a comparative analysis of Boston and Philadelphia. *Transportation*. doi:10.1007/s11116-021-10220-5
- Durrant-Whyte, H. F. (1988). Sensor models and multisensor integration. *International Journal of Robotics Research*, 7(6), 97–113.
- Egu, O., & Bonnel, P. (2020). How comparable are origin-destination matrices estimated from automatic fare collection, origin-destination surveys and household travel survey? An empirical investigation in Lyon. *Transportation Research Part A: Policy and Practice*, 138, 267-282. doi:10.1016/j.tra.2020.05.021
- Egu, O., & Bonnel, P. (2021). Medium-term public transit route ridership forecasting: What, how and why? A case study in Lyon. *Transport Policy*, 105, 124-133. doi:10.1016/j.tranpol.2021.03.002
- El-Assi, W., Morency, C., Miller, E. J., & Habib, K. N. (2020). Investigating the capacity of continuous household travel surveys in capturing the temporal rhythms of travel demand. *Transportation*, 47(4), 1787-1808. doi:10.1007/s11116-019-09981-x
- El-Assi, W., Salah Mahmoud, M., & Nurul Habib, K. (2017). Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto. *Transportation*, 44(3), 589-613. doi:10.1007/s11116-015-9669-z
- El-Geneidy, A. M., Kastelberger, L., & Abdelhamid, H. T. (2011). Montréal's Roots: Exploring the Growth of Montréal's Indoor City. *Journal of Transport and Land Use*, 4(2). doi:10.5198/jtlu.v4i2.176
- El-Geneidy, A. M., Tetreault, P., & Surprenant-Legault, J. (2010). *Pedestrian access to transit: Identifying redundancies and gaps using a variable service area analysis*. Communication présentée à 89th Annual Meeting of the Transportation Research Board, Washington D.C., United States.
- 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
- El Faouzi, N.-E., Leung, H., & Kurian, A. (2011). Data fusion in intelligent transportation systems: Progress and challenges - A survey. *Information Fusion*, 12(1), 4-10. doi:10.1016/j.inffus.2010.06.001
- Erdman, C., & Emerson, J. W. (2007). bcp: An R Package for Performing a Bayesian Analysis of Change Point Problems. *Journal of Statistical Software*, 23(3), 13. doi:10.18637/jss.v023.i03
- Erhardt, G. D., Hoque, J. M., Goyal, V., Berrebi, S., Brakewood, C., & Watkins, K. E. (2022). Why has public transit ridership declined in the United States? *Transportation Research Part A: Policy and Practice*, 161, 68-87. doi:10.1016/j.tra.2022.04.006

- Erhardt, G. D., Roy, S., Cooper, D., Sana, B., Chen, M., & Castiglione, J. (2019). Do transportation network companies decrease or increase congestion? *Science Advances*, 5(5), eaau2670. doi:doi:10.1126/sciadv.aau2670
- Esling, P., & Agon, C. (2012). Time-series data mining. *ACM Computing Surveys (CSUR)*, 45(1), 1-34. doi:10.1145/2379776.2379788
- Esteban, J., Starr, A., Willetts, R., Hannah, P., & Bryanston-Cross, P. (2005). A Review of data fusion models and architectures: towards engineering guidelines. *Neural Computing & Applications*, 14(4), 273-281. doi:10.1007/s00521-004-0463-7
- Ewing, R., & Cervero, R. (2010). Travel and the built environment: A meta-analysis. *Journal of the American planning association*, 76(3), 265-294.
- Fabre, L., & Morency, C. (2019). Enriching Travel Demand Forecasting Models with a Household Typology. *Transportation Research Record*, 2673(4), 975-987. doi:10.1177/0361198119838978
- Feigon, S., & Murphy, C. (2018). *Broadening Understanding of the Interplay Among Public Transit, Shared Mobility, and Personal Automobiles* (Rapport n° TCRP Research Report 195). Transportation Research Board.
- Ferraris, F. S. G. (2018). La mobilité de demain sera multimodale ou ne sera pas. *Le Devoir*. Tiré de <https://www.ledevoir.com/societe/transports-urbanisme/540068/mobility-as-a-service-l-avenir-sera-multimodal-ou-pas-du-tout?fbclid=IwAR3RZzsKefsVngEgrdo5VAc55sk9acZaElfCxMN19lh4j4lPUX5dyEhFY6Q>
- Fishman, E., Washington, S., & Haworth, N. (2014). Bike share's impact on car use: Evidence from the United States, Great Britain, and Australia. *Transportation Research Part D: Transport and Environment*, 31, 13-20. doi:10.1016/j.trd.2014.05.013
- Forsman, Å., Gustafsson, S., & Vadeby, A. (2007). Impact of Nonresponse and Weighting in Swedish Travel Survey. *Transportation Research Record*, 1993(1), 80-88. doi:10.3141/1993-11
- Fox, J., & Weisberg, S. (2019). Time-Series Regression and Generalized Least Squares in R - An Appendix to An R Companion to Applied Regression. Dans *An R Companion to Applied Regression, third edition* (3rd^e éd.). Thousand Oaks, CA: Sage.
- Fukushige, T., Fitch, D. T., & Handy, S. (2021). Factors influencing dock-less E-bike-share mode substitution: Evidence from Sacramento, California. *Transportation Research Part D: Transport and Environment*, 99, 102990. doi:10.1016/j.trd.2021.102990
- Fuller, D., Gauvin, L., Kestens, Y., Morency, P., & Drouin, L. (2013). The potential modal shift and health benefits of implementing a public bicycle share program in Montreal, Canada. *The international journal of behavioral nutrition and physical activity*, 10(1), 66-66. doi:10.1186/1479-5868-10-66
- Fuller, D., Luan, H., Buote, R., & Auchincloss, A. H. (2019). Impact of a public transit strike on public bicycle share use: An interrupted time series natural experiment study. *Journal of Transport & Health*, 13, 137-142. doi:10.1016/j.jth.2019.03.018

- Fyhri, A., & Beate Sundfør, H. (2020). Do people who buy e-bikes cycle more? *Transportation Research Part D: Transport and Environment*, 86, 102422. doi:10.1016/j.trd.2020.102422
- Garaus, M., & Garaus, C. (2021). The Impact of the Covid-19 Pandemic on Consumers' Intention to Use Shared-Mobility Services in German Cities. *Frontiers in psychology*, 12, 646593. doi:10.3389/fpsyg.2021.646593
- Gardner, E. S. (2006). Exponential smoothing: The state of the art—Part II. *International Journal of Forecasting*, 22(4), 637-666. doi:10.1016/j.ijforecast.2006.03.005
- Gebhardt, L., Krajzewicz, D., Oostendorp, R., Goletz, M., Greger, K., Klötzke, M., . . . Heinrichs, D. (2016). Intermodal Urban Mobility: Users, Uses, and Use Cases. *Transportation Research Procedia*, 14(C), 1183-1192. doi:10.1016/j.trpro.2016.05.189
- Gehrke, S. R., Felix, A., & Reardon, T. G. (2019). Substitution of Ride-Hailing Services for More Sustainable Travel Options in the Greater Boston Region. *Transportation Research Record*, 2673(1), 438-446. doi:10.1177/0361198118821903
- Geurs, K. T., La Paix, L., & Van Weperen, S. (2016). A multi-modal network approach to model public transport accessibility impacts of bicycle-train integration policies. *European Transport Research Review*, 8(4), 25. doi:10.1007/s12544-016-0212-x
- 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>
- Gkiotsalitis, K., & Cats, O. (2021). Public transport planning adaption under the COVID-19 pandemic crisis: literature review of research needs and directions. *Transport Reviews*, 41(3), 374-392. doi:10.1080/01441647.2020.1857886
- Goulias, K. G. (1997). *Surveys using multiple approaches*. Communication présentée à Transport Surveys: Raising the Standard, Grainau, Germany.
- Goulias, K. G., Pendyala, R. M., & Bhat, C. R. (2013). Keynote — Total Design Data Needs for the New Generation Large-Scale Activity Microsimulation Models. Dans J. Zmud, M. Lee-Gosselin, M. Munizaga, & J. A. Carrasco (édit.), *Transport Survey Methods* (p. 21-46): Emerald Group Publishing Limited.
- Graehler Jr, M., Mucci, R. A., & Erhardt, G. D. (2019). *Understanding the Recent Transit Ridership Decline in Major US Cities: Service Cuts or Emerging Modes?* Communication présentée à 98th Annual Meeting of the Transportation Research Board, Washington D.C, United States.
- 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>
- Guidon, S., Becker, H., Dediu, H., & Axhausen, K. W. (2019). Electric Bicycle-Sharing: A New Competitor in the Urban Transportation Market? An Empirical Analysis of Transaction Data. *Transportation Research Record*, 2673(4), 15-26. doi:10.1177/0361198119836762
- Gurumurthy, K. M., Kockelman, K. M., & Zuniga-Garcia, N. (2020). First-Mile-Last-Mile Collector-Distributor System using Shared Autonomous Mobility. *Transportation Research Record*, 2674(10), 638-647. doi:10.1177/0361198120936267

- Habib, K. M. N., & Miller, E. J. (2008). Modelling daily activity program generation considering within-day and day-to-day dynamics in activity-travel behaviour. *Transportation*, 35(4), 467-484. doi:10.1007/s11116-008-9166-8
- Hall, J. D., Palsson, C., & Price, J. (2018). Is Uber a substitute or complement for public transit? *Journal of Urban Economics*, 108, 36-50. doi:10.1016/j.jue.2018.09.003
- Hamilton, T. L., & Wichman, C. (2018). Bicycle infrastructure and traffic congestion: Evidence from DC's Capital Bikeshare. *Journal of Environmental Economics and Management*, 87, 72-93.
- Handy, S. (2005). Smart Growth and the Transportation-Land Use Connection: What Does the Research Tell Us? *International Regional Science Review*, 28(2), 146-167. doi:10.1177/0160017604273626
- He, L., Agard, B., & Trépanier, M. (2018). A classification of public transit users with smart card data based on time series distance metrics and a hierarchical clustering method. *Transportmetrica A: Transport Science*, 1-20. doi:10.1080/23249935.2018.1479722
- Heiler, G., Reisch, T., Hurt, J., Forghani, M., Omani, A., Hanbury, A., & Karimipour, F. (2020). *Country-wide mobility changes observed using mobile phone data during COVID-19 pandemic*. Communication présentée à 2020 IEEE International Conference on Big Data (Big Data) (p. 3123-3132).
- Heinen, E. (2018). Are multimodals more likely to change their travel behaviour? A cross-sectional analysis to explore the theoretical link between multimodality and the intention to change mode choice. *Transportation Research Part F*, 56, 200-214. doi:10.1016/j.trf.2018.04.010
- Heinen, E., & Chatterjee, K. (2015). The same mode again? An exploration of mode choice variability in Great Britain using the National Travel Survey. *Transportation Research Part A*, 78, 266-282. doi:10.1016/j.tra.2015.05.015
- Henao, A., & Marshall, W. E. (2019). The impact of ride-hailing on vehicle miles traveled. *Transportation*, 46(6), 2173-2194. doi:10.1007/s11116-018-9923-2
- 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
- Hensher, D. A. (2020). What might Covid-19 mean for mobility as a service (MaaS)? *Transport Reviews*, 40(5), 551-556. doi:10.1080/01441647.2020.1770487
- Hietanen, S. (2014). "Mobility as a Service" - The new transport model? ITS & Transport management supplement. *Eurotransport*, 12(2), 2-4.
- Hilgert, T., Behrena, S. v., Kistorza, N., Kagerbauera, M., & Vortischa, P. (2017). *Does Travel Behavior of People using Mobility-Apps differ? Findings from a Market Analysis in Germany*. Communication présentée à 11th International Conference on Transport Survey Methods, Estérel (Quebec), Canada.
- Hilgert, T., von Behren, S., Eisenmann, C., & Vortisch, P. (2018). Are Activity Patterns Stable or Variable? Analysis of Three-Year Panel Data. *Transportation Research Record*, 2672(47), 46-56. doi:10.1177/0361198118773557

- Hollingsworth, J., Copeland, B., & Johnson, J. X. (2019). Are e-scooters polluters? The environmental impacts of shared dockless electric scooters. *Environmental Research Letters*, 14(8), 084031. doi:10.1088/1748-9326/ab2da8
- Hoogendoorn, R. G., de Haas, M. C., Scheepers, C. E., Gelauff, G. M. M., & Hoogendoorn-Lanser, S. (2018). Using inverted relative entropy to determine the representativeness of samples in mobility panels. *Transportation Research Procedia*, 32, 253-259. doi:10.1016/j.trpro.2018.10.047
- Hörcher, D., Singh, R., & Graham, D. J. (2022). Social distancing in public transport: mobilising new technologies for demand management under the Covid-19 crisis. *Transportation*, 49(2), 735-764. doi:10.1007/s11116-021-10192-6
- Hou, Y., Garikapati, V., Weigl, D., Henao, A., Moniot, M., & Sperling, J. (2020). Factors Influencing Willingness to Pool in Ride-Hailing Trips. *Transportation Research Record*, 2674(5), 419-429. doi:10.1177/0361198120915886
- Huang, F. L. (2018). Multilevel Modeling and Ordinary Least Squares Regression: How Comparable Are They? *The Journal of Experimental Education*, 86(2), 265-281. doi:10.1080/00220973.2016.1277339
- Huang, H., Cheng, Y., & Weibel, R. (2019). Transport mode detection based on mobile phone network data: A systematic review. *Transportation Research Part C: Emerging Technologies*, 101, 297-312. doi:10.1016/j.trc.2019.02.008
- Hymel, K. M., Small, K. A., & Dender, K. V. (2010). Induced demand and rebound effects in road transport. *Transportation Research Part B*, 44(10), 1220-1241. doi:10.1016/j.trb.2010.02.007
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice* (2nd^e éd.). Melbourne, Australia: OTexts.
- Hyndman, R. J., Koehler, A., Ord, K., & Snyder, R. (2008). *Forecasting with Exponential Smoothing: The State Space Approach* (1. Aufl.^e éd.). Berlin, Heidelberg: Springer-Verlag.
- Iacus, S. M., Santamaria, C., Sermi, F., Spyrtos, S., Tarchi, D., & Vespe, M. (2021). Mobility functional areas and COVID-19 spread. *Transportation*. doi:10.1007/s11116-021-10234-z
- Ignacio Nazif-Muñoz, J., Nandi, A., & Ruiz-Casares, M. (2018). Impact of child restraint policies on child occupant fatalities and injuries in Chile and its regions: An interrupted time-series study. *Accident Analysis & Prevention*, 120, 38-45. doi:10.1016/j.aap.2018.07.028
- INSPQ. (2020). Ligne du temps COVID-19 au Québec. Tiré de <https://www.inspq.qc.ca/covid-19/donnees/ligne-du-temps?fbclid=IwAR03hqhQlvE1m4iaIYQ-r0ilMerd5BJ-PPN1XM05XHMK-Ow3zWZCNcTWUxU>
- Irawan, M. Z., Belgiawan, P. F., Joewono, T. B., Bastariant, F. F., Rizki, M., & Ilahi, A. (2022). Exploring activity-travel behavior changes during the beginning of COVID-19 pandemic in Indonesia. *Transportation*, 49, 529–553. doi:10.1007/s11116-021-10185-5
- Janzen, M. (2017). *Population synthesis for long-distance travel demand simulations*. Communication présentée à 6th symposium of the European association for research in transportation (hEART 2017).

- Jara-Díaz, S., & Rosales-Salas, J. (2015). Understanding time use: Daily or weekly data? *Transportation Research Part A: Policy and Practice*, 76, 38-57. doi:10.1016/j.tra.2014.07.009
- Jiang, S., Guan, W., He, Z., & Yang, L. (2018). Exploring the intermodal relationship between taxi and subway in Beijing, China. *Journal of Advanced Transportation*, 2018. doi:10.1155/2018/3981845
- Jin, H., Jin, F., Wang, J. e., Sun, W., & Dong, L. (2019). Competition and Cooperation between Shared Bicycles and Public Transit: A Case Study of Beijing. *Sustainability*, 11(5), 1323. doi:10.3390/su11051323
- Jittrapirom, P., Caiati, V., Feneri, A.-M., Ebrahimigharehbaghi, S., Alonso González, M. J., & Narayan, J. (2017). Mobility as a service: A critical review of definitions, assessments of schemes, and key challenges. *Urban Planning*, 2(2), 13–25. doi:0.17645/up.v2i2.931
- Jones, W. B., Cassady, C. R., & Bowden Jr, R. O. J. T. L. (2000). Developing a standard definition of intermodal transportation. 27, 345.
- Jun, M.-J., Choi, K., Jeong, J.-E., Kwon, K.-H., & Kim, H.-J. (2015). Land use characteristics of subway catchment areas and their influence on subway ridership in Seoul. *Journal of Transport Geography*, 48, 30-40. doi:10.1016/j.jtrangeo.2015.08.002
- Kager, R., Bertolini, L., & Te Brömmelstroet, M. (2016). Characterisation of and reflections on the synergy of bicycles and public transport. *Transportation Research Part A*, 85, 208-219. doi:10.1016/j.tra.2016.01.015
- Kagerbauer, M., Hilgert, T., Schroeder, O., & Vortisch, P. (2015). Household Travel Survey of Intermodal Trips – Approach, Challenges and Comparison. *Transportation Research Procedia*, 11(C), 330-339. doi:10.1016/j.trpro.2015.12.028
- Kamargianni, M., Li, W., Matyas, M., & Schäfer, A. (2016). A Critical Review of New Mobility Services for Urban Transport. *Transportation Research Procedia*, 14(C), 3294-3303. doi:10.1016/j.trpro.2016.05.277
- Kang, H., & Scott, D. M. (2010). Exploring day-to-day variability in time use for household members. *Transportation Research Part A: Policy and Practice*, 44(8), 609-619. doi:10.1016/j.tra.2010.04.002
- Khaleghi, B., Khamis, A., Karray, F. O., & Razavi, S. N. (2013). Multisensor data fusion: A review of the state-of-the-art. *Information Fusion*, 14(1), 28-44. doi:10.1016/j.inffus.2011.08.001
- Kim, J., Rasouli, S., & Timmermans, H. (2014). Hybrid Choice Models: Principles and Recent Progress Incorporating Social Influence and Nonlinear Utility Functions. *Procedia Environmental Sciences*, 22, 20-34. doi:10.1016/j.proenv.2014.11.003
- Kitamura, R., Yamamoto, T., Susilo, Y. O., & Axhausen, K. W. (2006). How routine is a routine? An analysis of the day-to-day variability in prism vertex location. *Transportation Research Part A: Policy and Practice*, 40(3), 259-279. doi:10.1016/j.tra.2005.07.002
- Kong, H., Jin, S. T., & Sui, D. Z. (2020). Deciphering the relationship between bikesharing and public transit: Modal substitution, integration, and complementation. *Transportation Research Part D: Transport and Environment*, 85, 102392. doi:10.1016/j.trd.2020.102392

- Kressner, J. D. (2017). *Synthetic Household Travel Data Using Consumer and Mobile Phone Data*. Final Report for NCHRP IDEA Project 184. Tiré de <http://www.trb.org/Main/Blurbs/176216.aspx>
- Krier, C., Chrétien, J., & Louvet, N. (2019). *Usages et usagers de services de trottinettes électriques en free-floating en France*. 6t. Tiré de <https://6-t.co/trottinettes-freefloating/>
- Kuhnimhof, T. (2009). *Measuring and modeling multimodal mode use in the longitudinal section*. Communication présentée à 88th Annual Meeting of the Transportation Research Board, Washington, D.C., USA.
- Kuhnimhof, T., Buehler, R., & Dargay, J. (2011). A New Generation: Travel Trends for Young Germans and Britons. *Transportation Research Record*, 2230(1), 58-67. doi:10.3141/2230-07
- Kuhnimhof, T., Buehler, R., Wirtz, M., & Kalinowska, D. (2012). Travel trends among young adults in Germany: increasing multimodality and declining car use for men. *Journal of Transport Geography*, 24, 443-450. doi:10.1016/j.jtrangeo.2012.04.018
- Kuhnimhof, T., Chlond, B., & Huang, P.-C. (2010). Multimodal Travel Choices of Bicyclists: Multiday Data Analysis of Bicycle Use in Germany. *Transportation Research Record*, 2190(1), 19-27. doi:10.3141/2190-03
- Kuhnimhof, T., Chlond, B., & Von Der Ruhren, S. (2006). Users of transport modes and multimodal travel behavior: Steps toward understanding travelers' options and choices. *Transportation Research Record*, 1985(1), 40-48.
- 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
- Labrecque, M. (1997). *Le cocktail transport: la solution pour freiner le déclin du transport en commun*. Montréal: Le Groupe vélo.
- Lahat, D., Adali, T., & Jutten, C. (2015). Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects. *Proceedings of the IEEE*, 103(9), 1449-1477. doi:10.1109/JPROC.2015.2460697
- Lane, C. (2005). PhillyCarShare: First-Year Social and Mobility Impacts of Carsharing in Philadelphia, Pennsylvania. *Transportation Research Record*, 1927(1), 158-166. doi:10.1177/0361198105192700118
- Lanzini, P., & Khan, S. A. (2017). Shedding light on the psychological and behavioral determinants of travel mode choice: A meta-analysis. *Transportation research part F: traffic psychology behaviour*, 48, 13-27.
- Lau, B. P. L., Marakkalage, S. H., Zhou, Y., Hassan, N. U., Yuen, C., Zhang, M., & Tan, U. X. (2019). A survey of data fusion in smart city applications. *Information Fusion*, 52, 357-374. doi:10.1016/j.inffus.2019.05.004
- Lavery, T. A., Páez, A., & Kanaroglou, P. S. (2013). Driving out of choices: An investigation of transport modality in a university sample. *Transportation Research Part A*, 57, 37-46. doi:10.1016/j.tra.2013.09.010

- Laviolette, J., Morency, C., & Waygood, E. O. D. (2020). Persistance de l'automobilité ? Analyse en trois perspectives. *Flux*, 119-120(1), 142-172. doi:10.3917/flux1.119.0142
- Le Vine, S., Lee-Gosselin, M., Sivakumar, A., & Polak, J. (2014). A new approach to predict the market and impacts of round-trip and point-to-point carsharing systems: Case study of London. *Transportation Research Part D*, 32, 218-229. doi:10.1016/j.trd.2014.07.005
- Le Vine, S., & Polak, J. (2019). The impact of free-floating carsharing on car ownership: Early-stage findings from London. *Transport Policy*, 75, 119-127. doi:10.1016/j.tranpol.2017.02.004
- Lechner, M. (2011). The estimation of causal effects by difference-in-difference methods. *Foundations and Trends® in Econometrics*, 4(3), 165-224. doi:10.1561/08000000014
- Lee, M. S., & McNally, M. G. (2003). On the structure of weekly activity/travel patterns. *Transportation Research Part A: Policy and Practice*, 37(10), 823-839. doi:10.1016/S0965-8564(03)00047-8
- Lefebvre-Ropars, G., & Morency, C. (2018). Walkability: which measure to choose, where to measure it, and how? *Transportation research record*, 2672(35), 139-150.
- 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: Journal of the Transportation Research Board*, 2675(7), 259-274. doi:10.1177/0361198121995196
- Lenth, R. V. (2016). Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software*, 69(1), 33. doi:10.18637/jss.v069.i01
- Lepage, S., & Morency, C. (2019). *Impact of weather, service disruptions and other events on transportation demand* Communication présentée à TRANSITDATA2019 5th International Workshop and Symposium, Paris, France.
- Lepage, S., & Morency, C. (2021). Impact of Weather, Activities, and Service Disruptions on Transportation Demand. *Transportation Research Record*, 2675(1), 294-304. doi:10.1177/0361198120966326
- Li, H., Zhang, Y., Zhu, M., & Ren, G. (2021). Impacts of COVID-19 on the usage of public bicycle share in London. *Transportation Research Part A: Policy and Practice*, 150, 140-155. doi:10.1016/j.tra.2021.06.010
- Li, M., Dong, L., Shen, Z., Lang, W., & Ye, X. (2017). Examining the Interaction of Taxi and Subway Ridership for Sustainable Urbanization. *Sustainability*, 9(2), 242. doi:10.3390/su9020242
- Li, Q., Liao, F., Timmermans, H. J. P., Huang, H., & Zhou, J. (2018). Incorporating free-floating car-sharing into an activity-based dynamic user equilibrium model: A demand-side model. *Transportation Research Part B: Methodological*, 107, 102-123. doi:10.1016/j.trb.2017.11.011
- Li, S., Lyu, D., Huang, G., Zhang, X., Gao, F., Chen, Y., & Liu, X. (2020). Spatially varying impacts of built environment factors on rail transit ridership at station level: A case study in Guangzhou, China. *Journal of Transport Geography*, 82, 102631. doi:10.1016/j.jtrangeo.2019.102631

- Li, W., Shalaby, A., & Habib, K. N. (2022). Exploring the correlation between ride-hailing and multimodal transit ridership in Toronto. *Transportation*, 49(3), 765-789. doi:10.1007/s11116-021-10193-5
- Li, X., Liu, Z., & Ma, X. (2022). Measuring Access and Egress Distance and Catchment Area of Multiple Feeding Modes for Metro Transferring Using Survey Data. *Sustainability*, 14(5), 2841. doi:10.3390/su14052841
- Li, X., Xu, H., Li, C., Sun, L., & Wang, R. (2016). Study on the demand and driving factors of urban underground space use. *Tunnelling and Underground Space Technology*, 55, 52-58. doi:10.1016/j.tust.2016.02.010
- Lin, D., Broere, W., & Cui, J. (2022). Underground space utilisation and new town development: Experiences, lessons and implications. *Tunnelling and Underground Space Technology*, 119, 104204. doi:10.1016/j.tust.2021.104204
- Lin, S., & Tian, H. (12-14 June 2020 2020). *Short-Term Metro Passenger Flow Prediction Based on Random Forest and LSTM*. Communication présentée à 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC) (vol. 1, p. 2520-2526). doi:10.1109/ITNEC48623.2020.9084974
- Lin, T. Y.-T. (2017). *Transit User Mode Choice Behaviour in Response to TTC Rapid Transit Service Disruption*. (University of Toronto). Tiré de <https://tspace.library.utoronto.ca/handle/1807/77838>
- Litman, T. (2004). Transit price elasticities and cross-elasticities. *Journal of Public Transportation*, 7(2), 3.
- Litman, T. (2017). *Introduction to multi-modal transportation planning*. Victoria Transport Policy Institute. Tiré de https://vtpi.org/multimodal_planning.pdf
- Litman, T. (2019). *Towards more comprehensive and multi-modal transport evaluation*. Victoria Transport Policy Institute. Tiré de https://www.vtpi.org/comp_evaluation.pdf
- Liu, B., Xu, W., Zhang, D., & Zhang, Q. (2022). Deformation behaviors and control indexes of metro-station deep excavations based on case histories. *Tunnelling and Underground Space Technology*, 122, 104400. doi:10.1016/j.tust.2022.104400
- Loa, P., Mashrur, S. M., & Nurul Habib, K. (2021). What influences the substitution of ride-sourcing for public transit and taxi services in Toronto? An exploratory structural equation model-based study. *International Journal of Sustainable Transportation*, 1-14. doi:10.1080/15568318.2021.1978018
- Long, Y., & Thill, J.-C. (2015). Combining smart card data and household travel survey to analyze jobs-housing relationships in Beijing. *Computers, Environment and Urban Systems*, 53, 19-35. doi:10.1016/j.compenvurbsys.2015.02.005
- 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
- Luo, H., Kou, Z., Zhao, F., & Cai, H. (2019). Comparative life cycle assessment of station-based and dock-less bike sharing systems. *Resources, Conservation and Recycling*, 146, 180-189. doi:10.1016/j.resconrec.2019.03.003

- Luo, H., Zhang, Z., Gkritza, K., & Cai, H. (2021). Are shared electric scooters competing with buses? a case study in Indianapolis. *Transportation Research Part D: Transport and Environment*, 97, 102877. doi:10.1016/j.trd.2021.102877
- Luo, R. C., Yih, C.-C., & Su, K. L. (2002). Multisensor fusion and integration: approaches, applications, and future research directions. *IEEE Sensors Journal*, 2(2), 107-119. doi:10.1109/JSEN.2002.1000251
- Luxen, D., & Vetter, C. (2011). *Real-time routing with OpenStreetMap data*. Communication présentée à 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Chicago, Illinois (p. 513–516). doi:10.1145/2093973.2094062
- Lyons, G., Hammond, P., & Mackay, K. (2019). The importance of user perspective in the evolution of MaaS. *Transportation Research Part A: Policy and Practice*, 121, 22-36. doi:10.1016/j.tra.2018.12.010
- Ma, T., & Knaap, G.-J. (2019). Estimating the Impacts of Capital Bikeshare on Metrorail Ridership in the Washington Metropolitan Area. *Transportation Research Record*, 2673(7), 371–379. doi:10.1177/0361198119849407
- Ma, T., Liu, C., & Erdoğan, S. (2015). Bicycle sharing and public transit: does Capital Bikeshare affect Metrorail ridership in Washington, DC? *Transportation Research Record*, 2534(1), 1-9.
- Ma, X., Ji, Y., Yang, M., Jin, Y., & Tan, X. (2018). Understanding bikeshare mode as a feeder to metro by isolating metro-bikeshare transfers from smart card data. *Transport Policy*, 71, 57-69. doi:10.1016/j.tranpol.2018.07.008
- Ma, X., Zhang, X., Li, X., Wang, X., & Zhao, X. (2019). Impacts of free-floating bikesharing system on public transit ridership. *Transportation Research Part D: Transport and Environment*, 76, 100-110. doi:10.1016/j.trd.2019.09.014
- Mageau-Béland, J. (2019). *Développement d'indicateurs de durabilité pour le transport en commun*. (École Polytechnique de Montréal). Tiré de <https://publications.polymtl.ca/4039/>
- Martel-Poliquin, É. (2012). *Mieux comprendre les déterminants du choix modal*. (École Polytechnique de Montréal). Tiré de <https://publications.polymtl.ca/904/>
- Martens, K. (2004). The bicycle as a feeding mode: experiences from three European countries. *Transportation Research Part D*, 9(4), 281-294. doi:10.1016/j.trd.2004.02.005
- Martens, K. (2007). Promoting bike-and-ride: The Dutch experience. *Transportation Research Part A: Policy and Practice*, 41(4), 326-338.
- Martin, E., & Shaheen, S. (2011). The impact of carsharing on public transit and non-motorized travel: An exploration of North American carsharing survey data. *Energies*, 4(11), 2094-2114. doi:10.3390/en4112094
- Martin, E., & Shaheen, S. (2014). Evaluating public transit modal shift dynamics in response to bikesharing: a tale of two U.S. cities. *Journal of Transport Geography*, 41, 315-324. doi:10.1016/j.jtrangeo.2014.06.026
- Martin, E., Shaheen, S., & Lidicker, J. (2010). Impact of Carsharing on Household Vehicle Holdings Results from North American Shared-Use Vehicle Survey. *Transportation Research Record*, 2143(2143), 150-158. doi:10.3141/2143-19

- McKenzie, G. (2019). Spatiotemporal comparative analysis of scooter-share and bike-share usage patterns in Washington, D.C. *Journal of Transport Geography*, 78, 19-28. doi:10.1016/j.jtrangeo.2019.05.007
- McNally, M. G. (2007). The Four Step Model. Dans D. A. Hensher & K. J. Button (édit.), *Handbook of transport modelling* (2^e éd., vol. 1). Amsterdam, The Netherlands: Emerald Group Publishing Limited.
- 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:10.1016/j.trpro.2018.10.067
- Ministère des Transports du Québec. (2018). *Transporter le Québec vers la modernité - Politique de mobilité durable 2030*.
- Minnen, J., Glorieux, I., & van Tienoven, T. P. (2015). Transportation habits: Evidence from time diary data. *Transportation Research Part A: Policy and Practice*, 76, 25-37. doi:10.1016/j.tra.2014.12.013
- Miramontes, M., Pfertner, M., Rayaprolu, H. S., Schreiner, M., & Wulforth, G. (2017). Impacts of a multimodal mobility service on travel behavior and preferences: user insights from Munich's first Mobility Station. *Transportation*, 44(6), 1325-1342. doi:10.1007/s11116-017-9806-y
- Mishra, G. S., Mokhtarian, P. L., Clewlow, R. R., & Widaman, K. F. (2019). Addressing the joint occurrence of self-selection and simultaneity biases in the estimation of program effects based on cross-sectional observational surveys: case study of travel behavior effects in carsharing. *Transportation*, 46(1), 95-123. doi:10.1007/s11116-017-9791-1
- Molin, E., Mokhtarian, P., & Kroesen, M. (2016). Multimodal travel groups and attitudes: A latent class cluster analysis of Dutch travelers. *Transportation Research Part A*, 83, 14-29. doi:10.1016/j.tra.2015.11.001
- Molloy, J., Schatzmann, T., Schoeman, B., Tchervenkova, C., Hintermann, B., & Axhausen, K. W. (2021). Observed impacts of the Covid-19 first wave on travel behaviour in Switzerland based on a large GPS panel. *Transport Policy*, 104, 43-51. doi:10.1016/j.tranpol.2021.01.009
- Moody, J., & Zhao, J. (2019). Car pride and its bidirectional relations with car ownership: Case studies in New York City and Houston. *Transportation Research Part A: Policy and Practice*, 124, 334-353. doi:10.1016/j.tra.2019.04.005
- Moorthy, K. S., & Png, I. P. (1992). Market segmentation, cannibalization, and the timing of product introductions. *Management science*, 38(3), 345-359.
- Morency, C. (2004). *Contributions à la modélisation totalement désagrégée des interactions entre mobilité urbaine et dynamiques spatiales*. (École Polytechnique de Montréal). Tiré de <https://publications.polymtl.ca/7282/>
- Morency, C., Bourdeau, J.-S., & Verreault, H. (2020). Modeling the interactions between mobility options in the surrounding of bikesharing stations. Dans *Mapping the Travel Behavior Genome* (p. 527-542): Elsevier.

- Morency, C., & Trépanier, M. (2013). *Enrichissement du processus de modélisation de la demande de transport à l'aide des modèles d'activités*. Ministère des Transports du Québec. Tiré de <http://www.bv.transports.gouv.qc.ca/mono/1147869.pdf>
- Morency, C., Trépanier, M., Faucher, J., Páez, A., & Verreault, H. (2017). *Modelling bikesharing usage in Montreal over 6 years*. CIRRELT, Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport. Tiré de <https://www.cirrelt.ca/documentstravail/cirrelt-2017-33.pdf>
- Morency, C., Trépanier, M., Saunier, N., Verreault, H., & Bourdeau, J.-S. (2017). *The challenges of using 5 parallel passive data streams to report on a wide range of mobility options*. Communication présentée à 11th International Conference on Travel Survey Methods, Estérel, Canada.
- 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. doi:10.1016/j.trpro.2018.10.014
- Morency, C., & Verreault, H. (2021). *Assessing the Metropolitan Impacts of Teleworking Scenarios*. Communication présentée à Transportation Research Board (TRB) 100th Annual Meeting, Washington, D.C.
- Mouratidis, K., & Papagiannakis, A. (2021). COVID-19, internet, and mobility: The rise of telework, telehealth, e-learning, and e-shopping. *Sustainable Cities and Society*, 74, 103182. doi:10.1016/j.scs.2021.103182
- 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
- Narayanan, S., Chaniotakis, E., & Antoniou, C. (2020). Shared autonomous vehicle services: A comprehensive review. *Transportation Research Part C: Emerging Technologies*, 111, 255-293. doi:10.1016/j.trc.2019.12.008
- Nguyen, M. H. (2021). Factors influencing home-based telework in Hanoi (Vietnam) during and after the COVID-19 era. *Transportation*, 48(6), 3207-3238. doi:10.1007/s11116-021-10169-5
- Ni, X., Lu, J., Wang, Y., Shi, J., Chen, W., & Tang, L. (2021). Field investigation of the influence of basement excavation and dewatering on ground and structure responses. *Tunnelling and Underground Space Technology*, 117, 104121. doi:10.1016/j.tust.2021.104121
- Nie, Y. (2017). How can the taxi industry survive the tide of ridesourcing? Evidence from Shenzhen, China. *Transportation Research Part C: Emerging Technologies*, 79, 242-256. doi:10.1016/j.trc.2017.03.017
- Nobis, C. (2007). Multimodality: facets and causes of sustainable mobility behavior. *Transportation Research Record*, 2010(1), 35-44.
- Noursalehi, P., Koutsopoulos, H. N., & Zhao, J. (2018). Real time transit demand prediction capturing station interactions and impact of special events. *Transportation Research Part C: Emerging Technologies*, 97, 277-300. doi:10.1016/j.trc.2018.10.023
- Odum, E., & Barrett, G. (2005). *Fundamentals of Ecology*. Belmont, CA: Thomson Brooks.

- Olafsson, A. S., Nielsen, T. S., & Carstensen, T. A. (2016). Cycling in multimodal transport behaviours: Exploring modality styles in the Danish population. *Journal of Transport Geography*, 52, 123-130. doi:10.1016/j.jtrangeo.2016.03.010
- Olaru, D., Smith, B., Xia, J., & Lin, T. (2014). Travellers' Attitudes Towards Park-and-Ride (PnR) and Choice of PnR Station: Evidence from Perth, Western Australia. *Procedia - Social and Behavioral Sciences*, 162, 101-110. doi:10.1016/j.sbspro.2014.12.190
- Oostendorp, R., & Gebhardt, L. (2018). Combining means of transport as a users' strategy to optimize traveling in an urban context: empirical results on intermodal travel behavior from a survey in Berlin. *Journal of Transport Geography*, 71, 72-83. doi:10.1016/j.jtrangeo.2018.07.006
- Oostendorp, R., Krajzewicz, D., Gebhardt, L., & Heinrichs, D. (2019). Intermodal mobility in cities and its contribution to accessibility. *Applied Mobilities*, 4(2), 183-199. doi:10.1080/23800127.2018.1554293
- 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
- Ortúzar, J. de D., & Willumsen, L. G. (2011). *Modelling Transport* (4. Aufl.^e éd.). Hoboken: Wiley.
- Ounoughi, C., & Ben Yahia, S. (2022). Data fusion for ITS: A systematic literature review. *Information Fusion*, 89, 267-291. doi:10.1016/j.inffus.2022.08.016
- Pan, H., Shen, Q., & Xue, S. (2010). Intermodal Transfer between Bicycles and Rail Transit in Shanghai, China. *Transportation Research Record: Journal of the Transportation Research Board*, 2144(1), 181-188. doi:10.3141/2144-20
- Papon, F., Beauvais, J. M., Midenet, S., Côme, E., Polombo, N., Abours, S., . . . Soulas, C. (2017). Evaluation of the bicycle as a feeder mode to regional train stations. *Transportation Research Procedia*, 25, 2717-2736. doi:10.1016/j.trpro.2017.05.211
- Papu Carrone, A., Hoening, V. M., Jensen, A. F., Mabit, S. E., & Rich, J. (2020). Understanding car sharing preferences and mode substitution patterns: A stated preference experiment. *Transport Policy*, 98, 139-147. doi:10.1016/j.tranpol.2020.03.010
- Pas, E. I. (1987). Intrapersonal variability and model goodness-of-fit. *Transportation Research Part A: General*, 21(6), 431-438. doi:10.1016/0191-2607(87)90032-X
- Paulley, N., Balcombe, R., Mackett, R., Titheridge, H., Preston, J., Wardman, M., . . . White, P. (2006). The demand for public transport: The effects of fares, quality of service, income and car ownership. *Transport Policy*, 13(4), 295-306. doi:10.1016/j.tranpol.2005.12.004
- Pawar, D. S., Yadav, A. K., Akolekar, N., & Velaga, N. R. (2020). Impact of physical distancing due to novel coronavirus (SARS-CoV-2) on daily travel for work during transition to lockdown. *Transportation Research Interdisciplinary Perspectives*, 7, 100203. doi:10.1016/j.trip.2020.100203
- Pfaff, B. (2008). *Analysis of integrated and cointegrated time series with R*: Springer Science & Business Media.

- Pirie, S., Trépanier, M., & Gendron, B. (2021). *Characterization of a COVID-Fired Urban Bike Delivery System*. CIRRELT. Tiré de <https://www.cirrelt.ca/documentstravail/cirrelt-2021-02.pdf>
- Pnevmatikou, A. M., Karlaftis, M. G., & Kepaptsoglou, K. (2015). Metro service disruptions: how do people choose to travel? *Transportation*, 42(6), 933-949. doi:10.1007/s11116-015-9656-4
- Polak, J. W., & Cornelis, E. (2009). Best Practices in Data Fusion: Synthesis of a Workshop. Dans P. Bonnel, M. Lee-Gosselin, J. Zmud, & J.-L. Madre (édit.), *Transport Survey Methods* (p. 613-619): Emerald Group Publishing Limited.
- Porter, S. R., Whitcomb, M. E., & Weitzer, W. H. (2004). Multiple surveys of students and survey fatigue. *New Directions for Institutional Research*, 2004(121), 63-73. doi:10.1002/ir.101
- Pucher, J., & Buehler, R. (2009). Integrating bicycling and public transport in North America. *Journal of Public Transportation*, 12(3), 5.
- Qian, X., Lei, T., Xue, J., Lei, Z., & Ukkusuri, S. V. (2020). Impact of transportation network companies on urban congestion: Evidence from large-scale trajectory data. *Sustainable Cities and Society*, 55, 102053. doi:10.1016/j.scs.2020.102053
- Qiao, Y.-K., Peng, F.-L., Sabri, S., & Rajabifard, A. (2019). Low carbon effects of urban underground space. *Sustainable Cities and Society*, 45, 451-459. doi:10.1016/j.scs.2018.12.015
- Qin, L., Li, W., & Li, S. (2019). Effective passenger flow forecasting using STL and ESN based on two improvement strategies. *Neurocomputing*, 356, 244-256. doi:10.1016/j.neucom.2019.04.061
- Quddus, M. A., Ochieng, W. Y., & Noland, R. B. (2007). Current map-matching algorithms for transport applications: State-of-the art and future research directions. *Transportation Research Part C*, 15(5), 312-328. doi:10.1016/j.trc.2007.05.002
- Rasouli, S., & Timmermans, H. (2014). Activity-based models of travel demand: promises, progress and prospects. *International Journal of Urban Sciences*, 18(1), 31-60. doi:10.1080/12265934.2013.835118
- Raux, C., Ma, T.-Y., & Cornelis, E. (2016). Variability in daily activity-travel patterns: the case of a one-week travel diary. *European Transport Research Review*, 8(4), 1-14. doi:10.1007/s12544-016-0213-9
- Rayle, L., Dai, D., Chan, N., Cervero, R., & Shaheen, S. (2016). Just a better taxi? A survey-based comparison of taxis, transit, and ridesourcing services in San Francisco. *Transport Policy*, 45, 168-178. doi:10.1016/j.tranpol.2015.10.004
- Rebecq, A. (2016). *Icarus: un package r pour le calage sur marges et ses variantes*. Communication présentée à 9e colloque francophone sur les sondages, Gatineau (Canada).
- Reck, D. J., Martin, H., & Axhausen, K. W. (2022). Mode choice, substitution patterns and environmental impacts of shared and personal micro-mobility. *Transportation Research Part D: Transport and Environment*, 102, 103134. doi:10.1016/j.trd.2021.103134
- Rodrigue, J.-P. (2020). *The geography of transport systems*: Routledge.

- Rodriguez Cote, J. D., & Diana, M. (2017). Exploring the benefits of a traveller clustering approach based on multimodality attitudes and behaviours. *Transportation Research Procedia*, 25, 2552-2565. doi:10.1016/j.trpro.2017.05.295
- Roy, S., Komanduri, A., & Proussaloglou, K. (2020). Evolution of Transportation Network Companies and Taxis through 2013–2018 in Chicago. *Transportation Research Record*, 2674(7), 385-397. doi:10.1177/0361198120922851
- Saberi, M., Ghamami, M., Gu, Y., Shojaei, M. H., & Fishman, E. (2018). Understanding the impacts of a public transit disruption on bicycle sharing mobility patterns: A case of Tube strike in London. *Journal of Transport Geography*, 66, 154-166. doi:10.1016/j.jtrangeo.2017.11.018
- Sabir, M. (2011). *Weather and travel behaviour*. (Vrije Universiteit Amsterdam). Tiré de <https://research.vu.nl/ws/portalfiles/portal/42203407/complete+dissertation.pdf>
- Salazar, M., Rossi, F., Schiffer, M., Onder, C. H., & Pavone, M. (2018 2018). *On the Interaction between Autonomous Mobility-on-Demand and Public Transportation Systems*. Communication présentée à 21st International Conference on Intelligent Transportation Systems, Maui, Hawaii, USA (p. 2262-2269). doi:10.1109/ITSC.2018.8569381
- Sallis, J. F., Frank, L. D., Saelens, B. E., & Kraft, M. K. (2004). Active transportation and physical activity: opportunities for collaboration on transportation and public health research. *Transportation Research Part A: Policy and Practice*, 38(4), 249-268. doi:10.1016/j.tra.2003.11.003
- Sandén, B., & Hillman, K. (2011). A framework for analysis of multi-mode interaction among technologies with examples from the history of alternative transport fuels in Sweden. *Research Policy*, 40(3), 403.
- Santamaria, C., Sermi, F., Spyrtos, S., Iacus, S. M., Annunziato, A., Tarchi, D., & Vespe, M. (2020). Measuring the impact of COVID-19 confinement measures on human mobility using mobile positioning data. A European regional analysis. *Safety science*, 132, 104925. doi:10.1016/j.ssci.2020.104925
- Santos, G., Maoh, H., Potoglou, D., & von Brunn, T. (2013). Factors influencing modal split of commuting journeys in medium-size European cities. *Journal of Transport Geography*, 30, 127-137. doi:10.1016/j.jtrangeo.2013.04.005
- Saxena, N., Hossein Rashidi, T., & Auld, J. (2019). Studying the tastes effecting mode choice behavior of travelers under transit service disruptions. *Travel Behaviour and Society*, 17, 86-95. doi:10.1016/j.tbs.2019.07.004
- Schaller, B. (2017). *Unsustainable? The growth of app-based ride services and traffic, travel and the future of New York City*. Schaller Consulting. Tiré de <http://www.schallerconsult.com/rideservices/unsustainable.pdf>
- Schaller, B. (2021). Can sharing a ride make for less traffic? Evidence from Uber and Lyft and implications for cities. *Transport Policy*, 102, 1-10. doi:10.1016/j.tranpol.2020.12.015
- Schauder, S. A., & Foley, M. C. (2015). The relationship between active transportation and health. *Journal of Transport & Health*, 2(3), 343-349. doi:10.1016/j.jth.2015.06.006

- Scheiner, J., Chatterjee, K., & Heinen, E. (2016). Key events and multimodality: A life course approach. *Transportation Research Part A*, 91, 148-165. doi:10.1016/j.tra.2016.06.028
- Schlich, R., & Axhausen, K. W. (2003). Habitual travel behaviour: Evidence from a six-week travel diary. *Transportation*, 30(1), 13-36. doi:10.1023/A:1021230507071
- Schnobb, P. (2018). Discours de Philippe Schnobb, président du conseil d'administration de la STM, devant la Chambre de commerce du Montréal métropolitain.
- Schönfelder, S., & Axhausen, K. W. (2010). *Urban Rhythms and Travel Behaviour: Spatial and Temporal Phenomena of Daily Travel*: Ashgate Publishing Ltd.
- Schuler, M. S., Chu, W., & Coffman, D. (2016). Propensity score weighting for a continuous exposure with multilevel data. *Health Services and Outcomes Research Methodology*, 16(4), 271-292. doi:10.1007/s10742-016-0157-5
- Schuppan, J., Kettner, S., Delatte, A., & Schwedes, O. (2014). Urban Multimodal Travel Behaviour: Towards Mobility without a Private Car. *Transportation Research Procedia*, 4, 553-556. doi:10.1016/j.trpro.2014.11.042
- Secrétariat à l'enquête Origine-Destination. (2020). *Enquête Origine-Destination 2018, version 18.2b*.
- Secrétariat à l'enquête Origine-Destination. (2010). *Mobilité des personnes dans la région de Montréal, Enquête Origine-Destination 2008, version 08.2a*. Tiré de <https://www.artm.quebec/wp-content/uploads/2020/01/Mobilit%C3%A9-des-personnes-dans-la-r%C3%A9gion-de-Montr%C3%A9al-1.pdf>
- Secrétariat à l'enquête Origine-Destination. (2015). *Mobilité des personnes dans la région de Montréal, Enquête Origine-Destination 2013, version 13.2a*. Tiré de <https://www.artm.quebec/wp-content/uploads/2020/01/Mobilit%C3%A9%20des%20personnes%20dans%20la%20r%C3%A9gion%20de%20Montr%C3%A9al%202013.pdf>
- Secrétariat à l'enquête Origine-Destination. (2020). *Mobilité des personnes dans la région de Montréal, Enquête Origine-Destination 2018, version 18.2b*. Tiré de https://www.artm.quebec/wp-content/uploads/2020/06/document-mobilite_EOD_2018.pdf
- Sehgal, V., Getoor, L., & Viechnicki, P. (2006 2006). *Entity resolution in geospatial data integration* (p. 83-90). doi:10.1145/1183471.1183486
- Shaheen, S., Martin, E., & Cohen, A. (2013). Public Bikeshearing and Modal Shift Behavior: A Comparative Study of Early Bikeshearing Systems in North America. *International Journal of Transportation*, 1(1), 35-54. doi:10.14257/ijt.2013.1.1.03
- Shaheen, S., Martin, E., Cohen, A., & Finson, R. (2012). *Public bikeshearing In North America: early operator and user understanding*. Mineta Transportation Institute. Tiré de <https://rosap.nrl.bts.gov/view/dot/24566>
- Shaheen, S., & Wong, S. (2020). *Public Transit and Shared Mobility COVID-19 Recovery: Policy Recommendations and Research Needs*. UC Office of the President: University of California Institute of Transportation Studies. Tiré de <https://escholarship.org/uc/item/9nh6w2gq>

- Shakibaei, S., de Jong, G. C., Alpkökin, P., & Rashidi, T. H. (2021). Impact of the COVID-19 pandemic on travel behavior in Istanbul: A panel data analysis. *Sustainable Cities and Society*, 65. doi:10.1016/j.scs.2020.102619
- Shen, Y., Zhang, H., & Zhao, J. (2018). Integrating shared autonomous vehicle in public transportation system: A supply-side simulation of the first-mile service in Singapore. *Transportation Research Part A: Policy and Practice*, 113, 125-136. doi:10.1016/j.tra.2018.04.004
- Shortall, R., Mouter, N., & Van Wee, B. (2021). COVID-19 passenger transport measures and their impacts. *Transport Reviews*, 1-26. doi:10.1080/01441647.2021.1976307
- Sijpkens, P., & Brown, D. (1997). *Montreal's indoor city—35 years of development*. Communication présentée à 7th International Conference on Underground Space, Montreal, Canada (p. 15).
- Silvestri, A., Foudi, S., Galarraga, I., & Ansuategi, A. (2021). The contribution of carsharing to low carbon mobility: Complementarity and substitution with other modes. *Research in Transportation Economics*, 85, 100968. doi:10.1016/j.retrec.2020.100968
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: modeling change and event occurrence* ; Judith D. Singer, John B. Willett. Oxford;Toronto: Oxford University Press.
- Singleton, P. A., & Clifton, K. J. (2014). Exploring Synergy in Bicycle and Transit Use: Empirical Evidence at Two Scales. *Transportation Research Record*, 2417(1), 92-102. doi:10.3141/2417-10
- Sioui, L., Morency, C., & Trépanier, M. (2013). How Carsharing Affects the Travel Behavior of Households: A Case Study of Montréal, Canada. *International Journal of Sustainable Transportation*, 7(1), 52-69. doi:10.1080/15568318.2012.660109
- Sochor, J., Arby, H., Karlsson, I. C. M., & Sarasini, S. (2018). A topological approach to Mobility as a Service: A proposed tool for understanding requirements and effects, and for aiding the integration of societal goals. *Research in Transportation Business & Management*, 27, 3-14. doi:10.1016/j.rtbm.2018.12.003
- Société de transport de Montréal. (2002). Rapport annuel 2021. Tiré de <https://www.stm.info/fr/a-propos/informations-entreprise-et-financieres/rapport-annuel-2021>
- Song, Y., & Huang, Y. (2020). Investigating Complementary and Competitive Relationships between Bikeshare Service and Public Transit: A Spatial-Temporal Framework. *Transportation Research Record*, 2674(1), 260-271. doi:10.1177/0361198119899389
- Song, Y., & Noyce, D. (2019). Effects of transit signal priority on traffic safety: Interrupted time series analysis of Portland, Oregon, implementations. *Accident Analysis & Prevention*, 123, 291-302. doi:10.1016/j.aap.2018.12.001
- Soteropoulos, A., Berger, M., & Ciari, F. (2019). Impacts of automated vehicles on travel behaviour and land use: an international review of modelling studies. *Transport Reviews*, 39(1), 29-49. doi:10.1080/01441647.2018.1523253
- Spickermann, A., Grienitz, V., & von der Gracht, H. A. (2014). Heading towards a multimodal city of the future? Multi-stakeholder scenarios for urban mobility. *Technological Forecasting And Social Change*, 89, 201-221. doi:10.1016/j.techfore.2013.08.036

- Spurr, T., Chapleau, R., & Piche, D. (2014). Use of Subway Smart Card Transactions for the Discovery and Partial Correction of Travel Survey Bias. *Transportation Research Record*, 2405(2405), 57-67. doi:10.3141/2405-08
- 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
- Stankutė, S., & Asche, H. (2009). *An integrative approach to geospatial data fusion*. Communication présentée à International Conference on Computational Science and Its Applications (p. 490-504).
- Stopher, P. R., & Greaves, S. P. (2007). Household travel surveys: Where are we going? *Transportation Research Part A: Policy and Practice*, 41(5), 367-381. doi:10.1016/j.tra.2006.09.005
- Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science : a review journal of the Institute of Mathematical Statistics*, 25(1), 1.
- Stuart, E. A., Huskamp, H. A., Duckworth, K., Simmons, J., Song, Z., Chernew, M., & Barry, C. L. (2014). Using propensity scores in difference-in-differences models to estimate the effects of a policy change. *Health services & outcomes research methodology*, 14(4), 166-182. doi:10.1007/s10742-014-0123-z
- Sun, Q., Feng, T., Kemperman, A., & Spahn, A. (2020). Modal shift implications of e-bike use in the Netherlands: Moving towards sustainability? *Transportation Research Part D: Transport and Environment*, 78, 102202. doi:10.1016/j.trd.2019.102202
- Susilo, Y. O., & Axhausen, K. W. (2014). Repetitions in individual daily activity–travel–location patterns: a study using the Herfindahl–Hirschman Index. *Transportation*, 41(5), 995-1011.
- Susilo, Y. O., & Kitamura, R. (2005). Analysis of Day-to-Day Variability in an Individual's Action Space: Exploration of 6-Week Mobidrive Travel Diary Data. *Transportation Research Record*, 1902(1), 124-133. doi:10.1177/0361198105190200115
- Tang, J., Gao, F., Han, C., Cen, X., & Li, Z. (2021). Uncovering the spatially heterogeneous effects of shared mobility on public transit and taxi. *Journal of Transport Geography*, 95, 103134. doi:10.1016/j.jtrangeo.2021.103134
- Tao, S., Corcoran, J., Rowe, F., & Hickman, M. (2018). To travel or not to travel: ‘Weather’ is the question. Modelling the effect of local weather conditions on bus ridership. *Transportation Research Part C*, 86, 147-167. doi:10.1016/j.trc.2017.11.005
- Tarpin-Pitre, L., & Morency, C. (2020). Typology of Bikeshare Users Combining Bikeshare and Transit. *Transportation Research Record*, 2674(10), 475-483. doi:10.1177/0361198120936262
- Teixeira, J. F., & Lopes, M. (2020). The link between bike sharing and subway use during the COVID-19 pandemic: The case-study of New York's Citi Bike. *Transportation Research Interdisciplinary Perspectives*, 6, 100166. doi:10.1016/j.trip.2020.100166
- Thombre, A., & Agarwal, A. (2021). A paradigm shift in urban mobility: Policy insights from travel before and after COVID-19 to seize the opportunity. *Transport Policy*, 110, 335-353. doi:10.1016/j.tranpol.2021.06.010

- Tingting, L., Li, L., & Wei, L. (2011). *Analysis of public bicycles' impact on residents' trip chain*. Communication présentée à International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE) (p. 182-185).
- Tirachini, A. (2020). Ride-hailing, travel behaviour and sustainable mobility: an international review. *Transportation*, 47(4), 2011-2047. doi:10.1007/s11116-019-10070-2
- Tirachini, A., & Cats, O. (2020). COVID-19 and public transportation: Current assessment, prospects, and research needs. *Journal of Public Transportation*, 22(1), 1.
- Tirachini, A., & Gomez-Lobo, A. (2019). Does ride-hailing increase or decrease vehicle kilometers traveled (VKT)? A simulation approach for Santiago de Chile. *International Journal of Sustainable Transportation*, 14(3), 187-204. doi:10.1080/15568318.2018.1539146
- Toque, F., Khouadjia, M., Come, E., Trepanier, M., & Oukhellou, L. (2017 2017). *Short & long term forecasting of multimodal transport passenger flows with machine learning methods*. Communication présentée à IEEE 20th International Conference on Intelligent Transportation Systems Yokohama, Japan (p. 560-566). doi:10.1109/ITSC.2017.8317939
- Torrence, C., & Compo, G. P. (1998). A Practical Guide to Wavelet Analysis. *Bulletin of the American Meteorological Society*, 79(1), 61-78. doi:10.1175/1520-0477(1998)079<0061:Apgtwa>2.0.Co;2
- Tourisme Montréal. (2021). *Bilan touristique annuel 2020 à Montréal*. Tiré de <https://toolkit.mtl.org/bynder/media/7A1132A6-7FF2-4BEC-A97C36DE5AABF53F/download?filename=Bilan-annuel-2020&extension=pdf>
- Tremblay, P. (2002). *Survol technique du modèle de transport de la région de Montréal*. Service de la modélisation des systèmes de transport. Direction du partenariat, de la modélisation et de la géomatique. Ministère des Transports du Québec.
- Trépanier, M., Morency, C., & Blanchette, C. (2009). *Enhancing household travel surveys using smart card data*. Communication présentée à 88th Annual Meeting of the Transportation Research Board
- Trépanier, M., Morency, C., Nouri, P., & Braham, A. (2013). *Impacts of carsharing on urban mobility: Environmental and behavioural evidences*. Communication présentée à 13th World Conference on Transport Research, Rio de Janeiro, Brésil (p. 15-18).
- Truong, L. T., De Gruyter, C., Currie, G., & Delbosc, A. (2017). Estimating the trip generation impacts of autonomous vehicles on car travel in Victoria, Australia. *Transportation*, 44(6), 1279-1292. doi:10.1007/s11116-017-9802-2
- Utriainen, R., & Pöllänen, M. (2018). Review on mobility as a service in scientific publications. *Research in Transportation Business & Management*, 27, 15-23. doi:10.1016/j.rtbm.2018.10.005
- Van Eck, G., Brands, T., Wismans, L. J. J., Pel, A. J., & Van Nes, R. (2014). Model complexities and requirements for multimodal transport network design: Assessment of classical, state-of-the-practice, and state-of-the-research models. *Transportation Research Record*, 2429(1), 178-187. doi:10.3141/2429-19

- Venigalla, M. (2004). *Household travel survey data fusion issues*. Communication présentée à 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
- Vij, A., Carrel, A., & Walker, J. L. (2011). *Capturing modality styles using behavioral mixture models and longitudinal data*. Communication présentée à 2nd international choice modelling conference, Leeds.
- Vij, A., Carrel, A., & Walker, J. L. (2013). Incorporating the influence of latent modal preferences on travel mode choice behavior. *Transportation Research Part A*, 54, 164-178. doi:10.1016/j.tra.2013.07.008
- Ville de Montréal. (2019a). *Défi des villes intelligentes du Canada - Candidature finale de la Ville de Montréal*. Tiré de https://laburbain.montreal.ca/sites/default/files/candidature_fr_defi_des_villes_intelligentes_vf.pdf
- Ville de Montréal. (2019b). Les trottinettes électriques LIME s'ajoutent au cocktail de mobilité à Montréal.
- Vivre en Ville. (2018). Dépendance à l'automobile *Collectivitesviables.org*, *Vivre en Ville*. Tiré de <http://collectivitesviables.org/articles/dependance-a-l-automobile.aspx>
- Wallsten, S. (2015). The competitive effects of the sharing economy: how is Uber changing taxis. *Technology Policy Institute*, 22, 1-21.
- Wang, H., & Noland, R. B. (2021). Bikeshare and subway ridership changes during the COVID-19 pandemic in New York City. *Transport Policy*, 106, 262-270. doi:10.1016/j.tranpol.2021.04.004
- 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, K., Qian, X., Fitch, D. T., Lee, Y., Malik, J., & Circella, G. (2022). What travel modes do shared e-scooters displace? A review of recent research findings. *Transport Reviews*, 1-27. doi:10.1080/01441647.2021.2015639
- Wang, M., Perera, C., Jayaraman, P. P., Zhang, M., Strazdins, P., Shyamsundar, R. K., & Ranjan, R. (2016). City Data Fusion: Sensor Data Fusion in the Internet of Things. *International Journal of Distributed Systems and Technologies (IJDST)*, 7(1), 15-36. doi:10.4018/IJDST.2016010102
- Wang, M., & Zhou, X. (2017). Bike-sharing systems and congestion: Evidence from US cities. *Journal of Transport Geography*, 65, 147-154. doi:10.1016/j.jtrangeo.2017.10.022

- Wang, X. (26-28 April 2019 2019). *The Short-Term Passenger Flow Forecasting of Urban Rail Transit Based on Holt-Winters' Seasonal Method*. Communication présentée à 2019 4th International Conference on Electromechanical Control Technology and Transportation (ICECTT) (p. 265-268). doi:10.1109/ICECTT.2019.00067
- 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
- Ward, J. W., Michalek, J. J., Samaras, C., Azevedo, I. L., Henao, A., Rames, C., & Wenzel, T. (2021). The impact of Uber and Lyft on vehicle ownership, fuel economy, and transit across U.S. cities. *iScience*, 24(1), 101933. doi:10.1016/j.isci.2020.101933
- Watanabe, H., Chikaraishi, M., & Maruyama, T. (2021). How different are daily fluctuations and weekly rhythms in time-use behavior across urban settings? A case in two Japanese cities. *Travel Behaviour and Society*, 22, 146-154. doi:10.1016/j.tbs.2020.09.004
- Weis, C., & Axhausen, K. W. (2009). Induced travel demand: Evidence from a pseudo panel data based structural equations model. *Research in Transportation Economics*, 25(1), 8-18. doi:10.1016/j.retrec.2009.08.007
- Weiss, A., & Habib, K. N. (2017). Examining the difference between park and ride and kiss and ride station choices using a spatially weighted error correlation (SWEC) discrete choice model. *Journal of Transport Geography*, 59, 111-119. doi:10.1016/j.jtrangeo.2017.01.010
- White, F. E. (1991). *Data fusion lexicon*. San Diego, CA: Data Fusion Panel of the Joint Directors of Laboratories, Technical Panel for C3.
- Wielinski, G. (2018). *Modélisation des membres et de leur comportement dans un écosystème de services d'autopartage à Montréal*. (Ph.D. thesis, École Polytechnique de Montréal). Tiré de <https://publications.polymtl.ca/3781/>
- Wu, S.-s., Zhuang, Y., Chen, J., Wang, W., Bai, Y., & Lo, S.-m. (2019). Rethinking bus-to-metro accessibility in new town development: Case studies in Shanghai. *Cities*, 94, 211-224. doi:10.1016/j.cities.2019.06.010
- Xiao, J., Su, R., McBride, E., & Goulias, K. (2020). Exploring the correlations between spatiotemporal daily activity-travel patterns and stated interest and perception of risk with self-driving cars. *AGILE: GIScience Series*, 1, 1-15. doi:10.5194/agile-giss-1-22-2020
- Yan, X., Levine, J., & Zhao, X. (2019). Integrating ridesourcing services with public transit: An evaluation of traveler responses combining revealed and stated preference data. *Transportation Research Part C: Emerging Technologies*, 105, 683-696. doi:10.1016/j.trc.2018.07.029
- Yan, X., Yang, W., Zhang, X., Xu, Y., Bejleri, I., & Zhao, X. (2021). A spatiotemporal analysis of e-scooters' relationships with transit and station-based bikeshare. *Transportation Research Part D: Transport and Environment*, 101, 103088. doi:10.1016/j.trd.2021.103088
- Yang, H., Huo, J., Bao, Y., Li, X., Yang, L., & Cherry, C. R. (2021). Impact of e-scooter sharing on bike sharing in Chicago. *Transportation Research Part A: Policy and Practice*, 154, 23-36. doi:10.1016/j.tra.2021.09.012

- Yang, J., Chen, S., Qin, P., Lu, F., & Liu, A. A. (2018). The effect of subway expansions on vehicle congestion: Evidence from Beijing. *Journal of Environmental Economics and Management*, 88, 114-133. doi:10.1016/j.jeem.2017.09.007
- Yang, M., Pan, Y., Darzi, A., Ghader, S., Xiong, C., & Zhang, L. (2022). A data-driven travel mode share estimation framework based on mobile device location data. *Transportation*, 49(5), 1339-1383. doi:10.1007/s11116-021-10214-3
- Yang, X., Day, J. E., Langford, B. C., Cherry, C. R., Jones, L. R., Han, S. S., & Sun, J. (2017). Commute responses to employment decentralization: Anticipated versus actual mode choice behaviors of new town employees in Kunming, China. *Transportation Research Part D: Transport and Environment*, 52, 454-470. doi:10.1016/j.trd.2016.11.012
- Ye, J., Wang, D., Jia, Y., & Zhang, H. (2022). Competition or cooperation: Relationship between carsharing and other travel modes. *International Journal of Sustainable Transportation*, 16(7), 610-626. doi:10.1080/15568318.2021.1914792
- Yeh, C.-F., & Lee, M.-T. (2019). Effects of Taichung bus policy on ridership according to structural change analysis. *Transportation*, 46(1), 1-16. doi:10.1007/s11116-017-9778-y
- Yin, B., & Leurent, F. (2022). What are the multimodal patterns of individual mobility at the day level in the Paris region? A two-stage data-driven approach based on the 2018 Household Travel Survey. *Transportation*. doi:10.1007/s11116-022-10285-w
- Younes, H., Nasri, A., Baiocchi, G., & Zhang, L. (2019). How transit service closures influence bikesharing demand; lessons learned from SafeTrack project in Washington, D.C. metropolitan area. *Journal of Transport Geography*, 76, 83-92. doi:10.1016/j.jtrangeo.2019.03.004
- Young, M., Allen, J., & Farber, S. (2020). Measuring when Uber behaves as a substitute or supplement to transit: An examination of travel-time differences in Toronto. *Journal of Transport Geography*, 82, 102629. doi:10.1016/j.jtrangeo.2019.102629
- Zacharias, J. (2015). Underground pedestrian trips—trip generation, spatial distribution and activities in the Montréal underground. *Tunnelling and Underground Space Technology*, 46, 46-51. doi:10.1016/j.tust.2014.10.006
- Zeileis, A., Kleiber, C., Krämer, W., & Hornik, K. (2003). Testing and dating of structural changes in practice. *Computational Statistics & Data Analysis*, 44(1), 109-123. doi:10.1016/S0167-9473(03)00030-6
- Zeng, C.-F., Song, W.-W., Xue, X.-L., Li, M.-K., Bai, N., & Mei, G.-X. (2021). Construction dewatering in a metro station incorporating buttress retaining wall to limit ground settlement: Insights from experimental modelling. *Tunnelling and Underground Space Technology*, 116, 104124. doi:10.1016/j.tust.2021.104124
- Zhang, Y., & Zhang, Y. (2018). Exploring the Relationship between Ridesharing and Public Transit Use in the United States. *International Journal of Environmental Research and Public Health*, 15(8), 1763. doi:10.3390/ijerph15081763
- Zhang, Z., Zaman, A., Xu, J., & Liu, X. (2022). Artificial intelligence-aided railroad trespassing detection and data analytics: Methodology and a case study. *Accident Analysis & Prevention*, 168, 106594. doi:10.1016/j.aap.2022.106594

- Zhao, J., Deng, W., Song, Y., & Zhu, Y. (2013). What influences Metro station ridership in China? Insights from Nanjing. *Cities*, 35, 114-124. doi:10.1016/j.cities.2013.07.002
- Zheng, Y. (2015). Methodologies for Cross-Domain Data Fusion: An Overview. *IEEE Transactions on Big Data*, 1(1), 16-34. doi:10.1109/TBDATA.2015.2465959
- Zhong, C., Huang, X., Müller Arisona, S., Schmitt, G., & Batty, M. (2014). Inferring building functions from a probabilistic model using public transportation data. *Computers, Environment and Urban Systems*, 48, 124-137. doi:10.1016/j.compenvurbsys.2014.07.004
- Zhou, M., Wang, D., Li, Q., Yue, Y., Tu, W., & Cao, R. (2017). Impacts of weather on public transport ridership: Results from mining data from different sources. *Transportation Research Part C: Emerging Technologies*, 75, 17-29. doi:10.1016/j.trc.2016.12.001
- Zhou, Z., Irizarry, J., & Zhou, J. (2021). Development of a database exclusively for subway construction accidents and corresponding analyses. *Tunnelling and Underground Space Technology*, 111, 103852. doi:10.1016/j.tust.2021.103852
- 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
- Ziedan, A., Shah, N. R., Wen, Y., Brakewood, C., Cherry, C. R., & Cole, J. (2021). Complement or compete? The effects of shared electric scooters on bus ridership. *Transportation Research Part D: Transport and Environment*, 101, 103098. doi:10.1016/j.trd.2021.103098
- Zmud, J. (2003). Designing Instruments to Improve Response. Dans P. Jones & P. R. Stopher (édit.), *Transport Survey Quality and Innovation* (p. 89-108): Emerald Group Publishing Limited.
- Zuniga-Garcia, N., Tec, M., Scott, J. G., & Machemehl, R. B. (2022). Evaluation of e-scooters as transit last-mile solution. *Transportation Research Part C: Emerging Technologies*, 139, 103660. doi:10.1016/j.trc.2022.103660

ANNEXE A SYNTHÈSE DES EXEMPLES D'APPLICATION DE TECHNIQUES DE FUSION DE DONNÉES DANS UN CONTEXTE DE TRANSPORT

Tableau A.1 Synthèse des exemples d'application de techniques de fusion de données dans un contexte de transport

Référence	Données fusionnées	Objectif de la fusion	Méthode de fusion	Type de méthode
Fusion inter-domaines				
(An et al., 2019)	Données d'achalandage et informations sur l'utilisation du sol	Évaluer l'impact de l'utilisation du sol sur l'achalandage du métro	Agrégation par station de métro	<i>Exact matching</i>
(Jun et al., 2015)	Données de cartes à puce et points d'intérêt	Évaluer l'impact de l'environnement bâti sur l'achalandage du métro	Agrégation par station de métro	<i>Exact matching</i>
(Mageau-Béland, 2019)	Enquête, recensement, informations sur l'élévation géographique et le niveau d'activité physique	Mesurer le potentiel de contact social et le niveau d'activité physique induits par l'utilisation du transport en commun	Attributs transférés par paire de coordonnées, aire de diffusion ou type d'activité réalisée	<i>Exact matching</i>
(Schauder & Foley, 2015)	Enquête, examen médical, données de laboratoire	Modéliser l'effet du transport actif sur différentes variables liées à la santé	Liaison des différentes données pas un identifiant unique de personne	<i>Exact matching</i>
Fusion intra-domaine entre plusieurs enquêtes				
(Bradley & Daly, 1997)	Préférences déclarées et révélées	Combiner des données de préférences déclarées et révélées pour modéliser le choix du mode	Modèle logit mixte	<i>Statistical matching</i> (modèle explicite)
(Bhat & Castelar, 2002)	Préférences déclarées et révélées	Combiner des données de préférences déclarées et révélées pour évaluer l'impact de la tarification de la congestion	Modèle logit mixte	<i>Statistical matching</i> (modèle explicite)
(Cherchi & Ortúzar, 2002)	Préférences déclarées et révélées	Combiner des données de préférences déclarées et révélées pour évaluer l'impact d'un nouveau service de train	Modèle logit mixte	<i>Statistical matching</i> (modèle explicite)
(Ben-Akiva et al., 2002)	Variables classiques et variables attitudinales	Inclure des variables latentes attitudinales dans la modélisation du choix du mode	Modèle de choix hybride	<i>Statistical matching</i> (modèle explicite)
(Kim et al., 2014)	Variables classiques et variables attitudinales	Inclure l'influence sociale et des fonctions d'utilité non linéaires dans la modélisation du choix du mode	Modèle de choix hybride	<i>Statistical matching</i> (modèle explicite)
(Bhat, 2015)	Variables classiques, attitudinales et contextuelles	Modéliser conjointement différents choix pour tous types de variables dépendantes multiples	Modèle généralisé de données hétérogènes	<i>Statistical matching</i> (modèle explicite)

Tableau A.1 Synthèse des exemples d'application de techniques de fusion de données dans un contexte de transport (suite)

Référence	Données fusionnées	Objectif de la fusion	Méthode de fusion	Type de méthode
(Morency, 2004)	Enquête ménages déplacements et recensement	Introduire certains attributs sociaux, économiques et démographiques dans l'étude des interactions urbaines	Transmission basée sur des distributions conjointes (tables croisées)	<i>Statistical matching</i> (modèle explicite)
(Venigalla, 2004)	Enquête ménages déplacements et recensement	Développer de modèles de génération qui peuvent inclure toutes les variables disponibles dans le recensement	Transmission basée sur la distribution d'une variable commune	<i>Statistical matching</i> (modèle explicite)
(Hilgert et al., 2017)	Enquête ménage nationale et enquête web spécifique	Analyser l'influence de l'utilisation des applications de mobilité sur les comportements de déplacement	Typologie et appariement avec le groupe le plus similaire	<i>Statistical matching</i> (modèle implicite)
(Verreault & Morency, 2018)	Enquête ménages téléphonique et enquête web ciblée (étudiants)	Obtenir un portrait plus représentatif des comportements typiques de déplacement	Méthode de repondération	<i>Statistical matching</i> (modèle implicite)
(Wang, Kaili et al., 2021)	Enquête ménages (mode mixte) et enquête web ciblée (étudiants)	Réduire à la fois les biais de proxy et de sous-représentation dans l'enquête ménages déplacements	Simulation (modèle d'activités) et insertion directe de carnets de déplacements	<i>Statistical matching</i> (modèle explicite)
Fusion intra-domaine entre plusieurs sources de données passives				
(Lovelace et al., 2016)	Données de téléphones cellulaires, enquête commerciale auprès des consommateurs et messages Twitter	Tester le potentiel de vérification mutuelle des grands ensembles de données	Comparaison de trois matrices OD entre lieux de domicile et centres commerciaux	Comparaisons au niveau agrégé
(Giraud et al., 2016)	Données de cartes à puce, données de comptage et données GTFS	Compléter l'information fournie par le système de cartes à puce (localisation de l'origine et de la destination)	Algorithme basé sur le numéro du bus, l'horodatage de la transaction, les habitudes de déplacement et le service programmé	<i>Exact matching</i>
Fusion intra-domaine entre enquêtes et données passives				
(Egu & Bonnel, 2020)	Enquête ménages déplacements, enquête OD face-à-face et données de validation	Comparer la demande de transport estimée à partir de différentes sources de données	Comparaison de statistiques descriptives et de matrices OD	Comparaisons au niveau agrégé
(Spurr et al., 2014)	Enquête ménages déplacements et données de cartes à puce	Confronter le jour moyen typique de l'enquête à une journée spécifique des données de cartes à puce	Comparaison de distributions spatio-temporelles + développement d'une méthode de correction de l'enquête	Comparaisons au niveau agrégé
(Trépanier et al., 2009)	Enquête ménages déplacements et données de cartes à puce	Comparer le jour typique moyen de l'enquête avec les informations longitudinales des cartes à puce	Calcul d'indicateurs de comparaison	Comparaisons au niveau agrégé

Tableau A.1 Synthèse des exemples d'application de techniques de fusion de données dans un contexte de transport (suite)

Référence	Données fusionnées	Objectif de la fusion	Méthode de fusion	Type de méthode
(Munizaga et al., 2014)	Données de cartes à puce, enquête OD face-à-face et données d'un groupe de volontaires	Présenter une méthode de validation de matrices OD obtenues pour le TC à partir de données de cartes à puce	Comparaison des résultats obtenus par la méthode et des observations	Comparaisons au niveau désagrégué
(Long & Thill, 2015)	Enquête ménages déplacements, données de cartes à puce, données d'utilisation du sol	Identifier les lieux de domicile et de travail dans les données de cartes à puce	Algorithme basé sur des règles	<i>Statistical matching</i> (modèle explicite)
(Alsger et al., 2018)	Enquête ménages déplacements, données de cartes à puce, enquête OD spécifique, données d'utilisation du sol	Estimer le motif des déplacements dans les données de cartes à puce	Modèle basé sur des règles et le calcul de probabilités	<i>Statistical matching</i> (modèle explicite)
(Kusakabe & Asakura, 2014)	Enquête déplacements (pour une station de métro) et données de cartes à puce	Estimer le motif de déplacement dans les données de cartes à puce et analyser la relation entre la fréquence de déplacement et le motif	Modèle probabiliste bayésien	<i>Statistical matching</i> (modèle explicite)
(Zhong et al., 2014)	Enquête ménages déplacements, données de cartes à puce, données géographiques	Déduire la fonction principale des bâtiments à partir du motif des déplacements se destinant à ce bâtiment	Modèle probabiliste bayésien	<i>Statistical matching</i> (modèle explicite)
(Alexander et al., 2015)	Enquête ménages déplacements, recensement, données de cartes à puce	Estimer des matrices OD de déplacements quotidiens moyens par motif et période de la journée à partir de données cellulaires	Modèle basé sur des règles et le calcul de fonctions de densité de probabilité	<i>Statistical matching</i> (modèle explicite)
(Spurr et al., 2015)	Enquête ménages déplacements et données de cartes à puce	Faire correspondre des numéros de carte à puce à des répondants de l'enquête pour comparer les deux sources de données	Algorithme de recherche des déplacements déclarés basé sur différents attributs spatio-temporels	<i>Statistical matching</i> (modèle implicite)
(Grapperon et al., 2016)	Enquête ménages déplacements, recensement et données de cartes à puce	Enrichir les données de cartes à puce en leur imputant des informations socio-démographiques	Création d'une population synthétique et distribution des cartes à puce via une méthode d'appariement de la théorie des graphes	<i>Statistical matching</i> (modèle implicite)
(Zhu et al., 2018)	Enquête (véhicules interceptés) et données GPS	Compléter les valeurs manquantes dans les données GPS grâce à des déplacements similaires dans l'enquête	Calcul d'un degré de similarité entre les déplacements des deux bases de données puis choix du meilleur candidat par catégorie d'attributs	<i>Statistical matching</i> (modèle implicite)

Tableau A.1 Synthèse des exemples d'application de techniques de fusion de données dans un contexte de transport (suite et fin)

Référence	Données fusionnées	Objectif de la fusion	Méthode de fusion	Type de méthode
(Bwambale et al., 2021)	Enquête ménages déplacements, recensement et données de téléphones cellulaires	Enrichir les données d'enquête pour la génération des déplacements	Ajustement d'un modèle de génération de déplacements	<i>Statistical matching</i> (modèle explicite)
(Janzen, 2017)	Enquête ménages déplacements, statistiques nationales et données de téléphones cellulaires	Corriger la sous-déclaration des déplacements de longue distance dans l'enquête pour la synthèse de population	Ajustement des distributions cumulatives (<i>histogram matching</i>)	<i>Statistical matching</i> (modèle explicite)

ANNEXE B SYNTHÈSE DE L'ÉVALUATION DE L'IMPACT DES MODES ÉMERGENTS SUR LES MODES CLASSIQUES

Tableau B.1 Synthèse de l'évaluation de l'impact des modes émergents sur les modes classiques

Référence	Mode(s) émergent(s)	Mode(s) classique(s)	Zone d'étude	Méthode
(Shaheen et al., 2013)	Vélopartage	Automobile privée, taxi, transport en commun, mode actifs	4 villes nord-américaines	Analyse descriptive de données d'enquête
(Shaheen et al., 2012)	Vélopartage	Automobile privée, taxi, transport en commun, mode actifs	4 villes nord-américaines	Analyse descriptive de données d'enquête
(Fuller et al., 2013)	Vélopartage	Automobile privée, taxi, transport en commun, mode actifs	Montréal, Canada	Analyse descriptive de données d'enquête + calcul d'un indicateur de transfert modal
(Fishman et al., 2014)	Vélopartage	Automobile privée, taxi, transport en commun, mode actifs	Londres, 2 villes australiennes, 2 villes américaines	Analyse descriptive de données d'enquête + estimation de la réduction du kilométrage automobile
(Hamilton & Wichman, 2018)	Vélopartage	Automobile privée	Washington D.C., États-Unis	Méthode d'appariement par score de propension + modèle multiniveau à effets aléatoires
(Wang, Mingshu & Zhou, 2017)	Vélopartage	Automobile privée	96 zones urbaines américaines	Méthode d'appariement par score de propension + modèle DID
(Jin et al., 2019)	Vélopartage	Transport en commun (métro et bus)	Beijing, Chine	Histogrammes de distances + analyse de corrélation spatiale
(Ma, Xiaolei et al., 2019)	Vélopartage sans ancrage	Bus	Chengdu, Chine	Méthode d'appariement par score de propension + modèle DID
(Bielinski et al., 2021)	Vélopartage (vélos électriques)	Bus, automobile privée	Tricité, Pologne	Modèle « double-hurdle » basé sur des données d'enquête
(Guidon et al., 2019)	Vélopartage (vélos électriques)	Transport en commun (plusieurs modes)	Zurich, Suisse	Analyse descriptive + modèle de régression spatial
(Tingting et al., 2011)	Vélopartage	Transport en commun (bus et métro)	Jiangyin, Chine	Analyse des chaînes de déplacements + modèle logit imbriqué de choix modal
(Kong et al., 2020)	Vélopartage	Transport en commun (bus et métro)	4 villes américaines	Classification des déplacements + modèle de régression linéaire
(Ma, T. et al., 2015)	Vélopartage	Métro	Washington D.C., États-Unis	Modèle de régression linéaire

Tableau B.1 Synthèse de l'évaluation de l'impact des modes émergents sur les modes classiques (suite)

Référence	Mode(s) émergent(s)	Mode(s) classique(s)	Zone d'étude	Méthode
(Campbell & Brakewood, 2017)	Vélopartage	Bus	New-York, États-Unis	Modèle DID
(Graehler Jr et al., 2019)	Vélopartage, TNC	Transport en commun (plusieurs modes)	22 villes des États-Unis	Modèle multiniveau à effets aléatoires
(Ma, T. & Knaap, 2019)	Vélopartage	Métro	Washington D.C., États-Unis	Modèles de régression linéaire
(Ma, Xinwei et al., 2018)	Vélopartage	Métro	Nanjing, Chine	Identification des déplacements vélopartage-métro + analyse descriptive
(Tarpin-Pitre & Morency, 2020)	Vélopartage	Métro	Montréal, Canada	Identification des déplacements vélopartage-métro-vélopartage + typologie
(Martin & Shaheen, 2014)	Vélopartage	Transport en commun (bus et métro)	Washington D.C. et Minneapolis, États-Unis	Analyse descriptive de données d'enquête + modèles de régression ordinale
(Tang et al., 2021)	Vélopartage	Transport en commun (bus et métro)	Shenzhen, Chine	Modèle de régression spatial
(Bigazzi & Wong, 2020)	Vélo personnel électrique	Automobile privée, transport en commun, mode actifs	International	Méta-régression basée sur 24 études
(Fyhri & Beate Sundfør, 2020)	Vélo personnel électrique	Vélo personnel traditionnelle	Oslo, Norvège	Tests ANOVA pour comparer deux groupes enquêtés (clients vs contrôle)
(Cairns et al., 2017)	Vélo personnel électrique	Automobile privée, modes actifs	Brighton, Angleterre	Analyse descriptive de données d'enquête
(Sun et al., 2020)	Vélo personnel électrique	Automobile privée, modes actifs	Pays-Bas	Analyse descriptive de données d'enquête (panel)
(Wang, Kailai et al., 2022)	Trottinettes électriques	Automobile privée, taxi, TNC, transport en commun, modes actifs	International	Revue de littérature
(Krier et al., 2019)	Trottinettes électriques	Tous les autres modes	3 villes françaises	Analyse descriptive de données d'enquête
(Yang, H. et al., 2021)	Trottinettes électriques	Vélopartage	Chicago, États-Unis	Méthode d'appariement par score de propension + modèle DID
(McKenzie, 2019)	Trottinettes électriques	Vélopartage	Washington D.C., États-Unis	Comparaison des patterns spatio-temporels des deux modes
(Ziedan et al., 2021)	Trottinettes électriques	Bus	Nashville, États-Unis	Modèle multiniveau à effets fixes
(Luo, H. et al., 2021)	Trottinettes électriques	Bus	Indianapolis, États-Unis	Classification des déplacements + modèle DID

Tableau B.1 Synthèse de l'évaluation de l'impact des modes émergents sur les modes classiques (suite)

Référence	Mode(s) émergent(s)	Mode(s) classique(s)	Zone d'étude	Méthode
(Yan et al., 2021)	Trottinettes électriques	Transport en commun, vélopartage	Washington D.C., États-Unis	Analyses descriptives spatio-temporelles
(Zuniga-Garcia et al., 2022)	Trottinettes électriques	Bus	Austin, États-Unis	Méthode de Gradient Boosting + modèle ZINB (zero-inflated negative binomial)
(Reck et al., 2022)	Trottinettes électriques	Automobile privée, transport en commun, modes actifs	Zurich, Suisse	Modèle de choix modal + estimation des modes substitués avec MATSim + calcul des émissions GES
(Lane, 2005)	Autopartage (basé stations)	Automobile privée	Philadelphie, États-Unis	Analyse descriptive de données d'enquête
(Le Vine & Polak, 2019)	Autopartage (VLS)	Automobile privée	Londres, Angleterre	Analyse descriptive de données d'enquête + régression logistique binaire
(Martin et al., 2010)	Autopartage	Automobile privée	11 organismes d'Amérique du Nord	Analyse descriptive de données d'enquête
(Becker et al., 2018)	Autopartage (VLS)	Automobile privée	Basel, Suisse	Modèle DID
(Cervero et al., 2007)	Autopartage (basé stations)	Automobile privée, transport en commun, modes actifs	San Francisco, États-Unis	Analyse descriptive de données d'enquête + modèles de régression (linéaire et logit)
(Clewlow, 2016)	Autopartage	Automobile privée, taxi, covoiturage, transport en commun, modes actifs	San Francisco, États-Unis	Analyse descriptive de données d'enquête
(Sioui et al., 2013)	Autopartage	Automobile privée	Montréal, Canada	Comparaison des données de deux enquêtes
(Trépanier et al., 2013)	Autopartage	Automobile privée, taxi, transport en commun, modes actifs	Montréal, Canada	Analyse descriptive de données d'enquête + calcul du kilométrage et des émissions GES à partir de données transactionnelles
(Ceccato et al., 2021)	Autopartage (VLS)	Automobile privée, transport en commun	Turin, Italie	Modélisation du taux de substitution par un modèle binomial logit et un modèle RF (random forest)
(Martin & Shaheen, 2011)	Autopartage	Transport en commun, modes actifs, covoiturage	11 organismes d'Amérique du Nord	Analyse descriptive de données d'enquête
(Becker et al., 2017)	Autopartage (basé stations et VLS)	Automobile privée, transport en commun, taxi, TNC, modes actifs	Basel, Suisse	Analyse descriptive de données d'enquête + modèle de sélection de Heckman
(Silvestri et al., 2021)	Autopartage (basé stations et VLS)	Automobile privée, transport en commun	Plusieurs villes en Espagne	Analyse descriptive de données d'enquête

Tableau B.1 Synthèse de l'évaluation de l'impact des modes émergents sur les modes classiques (suite)

Référence	Mode(s) émergent(s)	Mode(s) classique(s)	Zone d'étude	Méthode
(Ceccato & Diana, 2021)	Autopartage (unidirectionnel basé stations et VLS)	Automobile privée, taxi, transport en commun, vélopartage	Turin, Italie	Analyse descriptive de données d'enquête + modèles logit binaire
(Ye et al., 2022)	Autopartage (unidirectionnel basé stations)	Automobile privée, taxi, transport en commun, modes non motorisés	Shanghai, Chine	Analyse descriptive de données d'enquête + analyse coût-bénéfice pour les modes substitués
(Le Vine et al., 2014)	Autopartage (basé stations et VLS)	Automobile privée, taxi, transport en commun, vélo	Londres, Angleterre	Modélisation basée sur le cadre conceptuel PAS (Perceived Activity Set) considérant le choix des outils de mobilité et le choix modal
(Papu Carrone et al., 2020)	Autopartage (VLS)	Transport en commun, transport en commun, vélo	Copenhague, Danemark	Modèle de choix modal (logit mixte) basé sur une enquête de préférences déclarées
(Contreras & Paz, 2018)	TNC	Taxi	Las Vegas, États-Unis	Modèle de régression linéaire
(Nie, 2017)	TNC	Taxi	Shenzhen, Chine	Analyse descriptive de traces GPS converties en déplacements
(Roy et al., 2020)	TNC	Taxi	Chicago, États-Unis	Analyse descriptive de données transactionnelles
(Schaller, 2017)	TNC	Taxi	New York, États-Unis	Analyse descriptive de différentes sources de données passives (déplacements et kilométrage)
(Wallsten, 2015)	TNC	Taxi	New York et Chicago, États-Unis	Analyse descriptive de données transactionnelles et de plaintes + modèles de régression
(Tirachini, 2020)	TNC	Automobile privée, taxi, transport en commun, modes actifs	International	Revue de littérature
(Hall et al., 2018)	TNC	Transport en commun (bus et métro)	339 villes américaines	Modèle DID
(Rayle et al., 2016)	TNC	Automobile privée, taxi, transport en commun, modes actifs	San Francisco, États-Unis	Analyse descriptive de données d'enquêtes
(Clewlow & Mishra, 2017)	TNC	Automobile privée, taxi, covoiturage, transport en commun, modes actifs	7 villes américaines	Analyse descriptive de données d'enquêtes
(Feigon & Murphy, 2018)	TNC	Automobile privée, transport en commun	Plusieurs régions urbaines américaines	Analyse descriptive de données d'enquêtes

Tableau B.1 Synthèse de l'évaluation de l'impact des modes émergents sur les modes classiques (suite)

Référence	Mode(s) émergent(s)	Mode(s) classique(s)	Zone d'étude	Méthode
(Ward et al., 2021)	TNC	Automobile privée, transport en commun	Plusieurs régions urbaines américaines	Modèle DID avec méthode de pondération par score de propension inverse
(Diao et al., 2021)	TNC	Automobile privée, transport en commun	Plusieurs régions urbaines américaines	Modèles multiniveaux à effets fixes
(Erhardt et al., 2019)	TNC	Automobile privée	San Francisco, États-Unis	Modèles multiniveaux à effets fixes
(Henao & Marshall, 2019)	TNC	Automobile privée, taxi, covoiturage, transport en commun, modes actifs	Denver, États-Unis	Analyse descriptive de données d'enquête (quasi-expérimentale)
(Tirachini & Gomez-Lobo, 2019)	TNC	Automobile privée, taxi, transport en commun, modes actifs	Santiago, Chili	Analyse descriptive de données d'enquête + simulations Monte Carlo
(Gehrke et al., 2019)	TNC	Transport en commun, modes actifs	Boston, États-Unis	Analyse descriptive de données d'enquête + régressions logistiques binaires
(Schaller, 2021)	TNC	Automobile privée	4 villes américaines + régions suburbaines de Californie, États-Unis	Calcul du kilométrage parcouru en se basant sur des hypothèses tirées de la littérature
(Qian et al., 2020)	TNC	Automobile privée	New York, États-Unis	Analyse descriptive de traces GPS et de différentes variables dérivées
(Zhang, Y. & Zhang, 2018)	TNC	Transport en commun (bus, métro, train)	États-Unis	Modèle ZINB (zero-inflated negative binomial)
(Diab, Kasraian, et al., 2020)	TNC, vélopartage	Bus	Montréal, Canada	Modèle multiniveau à effets aléatoires
(Erhardt et al., 2022)	TNC, vélopartage, trottinettes électriques	Transport en commun (bus et rail)	215 zones statistiques métropolitaines, États-Unis	Modèle multiniveau à effets fixes
(Babar & Burtch, 2020)	TNC	Transport en commun (bus et rail)	Plusieurs villes américaines	Modèle multiniveau à effets fixes
(Li, W. et al., 2022)	TNC	Transport en commun (bus et métro)	Toronto, Canada	Modèle de régression ordinaire et modèle multiniveau à effets aléatoires
(Dong et al., 2021)	TNC	Transport en commun (bus et métro)	Boston et Philadelphie, États-Unis	Analyse descriptive de données d'enquête + modèle logit mixte

Tableau B.1 Synthèse de l'évaluation de l'impact des modes émergents sur les modes classiques (suite et fin)

Référence	Mode(s) émergent(s)	Mode(s) classique(s)	Zone d'étude	Méthode
(Young et al., 2020)	TNC	Transport en commun (bus et métro)	Toronto, Canada	Analyse descriptive de données d'enquête + comparaison des temps de parcours + modèles de régression (ordinaire et logit)
(Yan et al., 2019)	TNC	Bus	Ann Arbor, États-Unis	Modèle logit mixte combinant des données de préférences déclarées et révélées + scénarios
(Davidson et al., 2017)	TNC	Métro	New York, États-Unis	Comparaison de deux sources de données (tous les usagers d'Uber vs les usagers d'Uber avec utilisation de l'application Transit)
(Soteropoulos et al., 2019)	Véhicules autonomes privés	Automobile privée, transport en commun, modes actifs	International	Revue de littérature
(Truong et al., 2017)	Véhicules autonomes privés	Automobile privée	Victoria, Australie	Estimation de la demande induite par groupe d'âges + analyse de scénarios du taux d'occupation
(Narayanan et al., 2020)	Véhicules autonomes partagés	Automobile privée, transport en commun, modes actifs	International	Revue de littérature
(Bischoff et al., 2019)	Véhicules autonomes partagés	Automobile privée, transport en commun (bus)	San Francisco, États-Unis	Modèle basé agents (MATSim)
(Bösch et al., 2018)	Véhicules autonomes partagés	Automobile privée, taxi, transport en commun	Zurich, Suisse	Analyse de coûts
(Salazar et al., 2018)	Véhicules autonomes partagés	Transport en commun (métro), marche	New York, États-Unis	Modèle d'optimisation des flux (multi-commodity flow optimization model)
(Shen et al., 2018)	Véhicules autonomes partagés	Transport en commun (bus et métro)	Singapour	Modèle basé agents (AnyLogic)
(Gurumurthy et al., 2020)	Véhicules autonomes partagés	Transport en commun (bus et train)	Austin, États-Unis	Modèle basé agents (MATSim)