

**Titre:** Modélisation de l'impact de la performance des modes actifs et collectifs sur le choix modal  
Title: **Modélisation de l'impact de la performance des modes actifs et collectifs sur le choix modal**

**Auteur:** Patrice Pelletier-Grenier  
Author: **Patrice Pelletier-Grenier**

**Date:** 2018

**Type:** Mémoire ou thèse / Dissertation or Thesis

**Référence:** Pelletier-Grenier, P. (2018). Modélisation de l'impact de la performance des modes actifs et collectifs sur le choix modal [Mémoire de maîtrise, École Polytechnique de Montréal]. PolyPublie. <https://publications.polymtl.ca/3047/>  
Citation: **Pelletier-Grenier, P. (2018). Modélisation de l'impact de la performance des modes actifs et collectifs sur le choix modal [Mémoire de maîtrise, École Polytechnique de Montréal]. PolyPublie. <https://publications.polymtl.ca/3047/>**

## Document en libre accès dans PolyPublie

Open Access document in PolyPublie

**URL de PolyPublie:** <https://publications.polymtl.ca/3047/>  
PolyPublie URL: **https://publications.polymtl.ca/3047/**

**Directeurs de recherche:** Catherine Morency  
Advisors: **Catherine Morency**

**Programme:** Génie civil  
Program: **Génie civil**

UNIVERSITÉ DE MONTRÉAL

MODÉLISATION DE L'IMPACT DE LA PERFORMANCE DES MODES ACTIFS ET  
COLLECTIFS SUR LE CHOIX MODAL

PATRICE PELLETIER-GRENIER

DÉPARTEMENT DES GÉNIES CIVIL, GÉOLOGIQUE ET DES MINES  
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION  
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES  
(GÉNIE CIVIL)

MARS 2018

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

MODÉLISATION DE L'IMPACT DE LA PERFORMANCE DES MODES ACTIFS ET  
COLLECTIFS SUR LE CHOIX MODAL

présenté par : PELLETIER-GRENIER Patrice

en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées

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

M. SAUNIER Nicolas, Ph. D., président

Mme MORENCY Catherine, Ph. D., membre et directrice de recherche

M. TRÉPANIER Martin, Ph. D., membre

## DÉDICACE

*À tous les éléments du terreau ayant permis à ce germe d'idée de devenir mémoire.*

*"The reality about transportation is that it's future-oriented."*

*"If we're planning for what we have, we're behind the curve."*

*-Anthony Foxx*

## REMERCIEMENTS

Je tiens à remercier premièrement ma directrice de recherche, la professeure Catherine Morency, pour son encadrement rigoureux dans le développement de la méthode proposée, ses commentaires constructifs, son support financier ainsi que son énergie contagieuse.

Également à la source de ce projet, j'aimerais remercier toute l'équipe de chercheurs impliquée et leurs partenaires institutionnels, le Dr. Patrick Morency, les professeurs Audrey Smargiassi, Marianne Hatzopoulou, Naveen Eluru, ainsi que la Direction de la santé publique (DSP) de Montréal, l'Institut National de santé Publique Québec et le Ministère des transports, de la mobilité durable et de l'électrification des transports (MTMDET). Un merci tout spécial à M. Pierre Tremblay, du MTMDET pour l'entente de transmission des données. Également impliqués à titre de financeurs, merci aux Instituts de Recherche en Santé du Canada, au Conseil de Recherche en Sciences Naturelles et en Génie du Canada ainsi qu'aux Fonds Vert.

Je suis également reconnaissant des collègues, autant de Polytechnique Montréal que d'ailleurs qui ont contribué de près ou de loin à mon parcours de maîtrise. Merci spécialement à Pierre-Léo Bourbonnais, Jean-Simon Bourdeau, Gabriel Lefebvre-Ropars, Astrid Sibiril, Amine Boussoffara, Jérôme Laviolette, Anaïs Moulin, Quentin Chevalier, Antoine Leroux, Charles Kehl, Léa Fabre, Élodie Deschaintres, Stephen Woodall, Laurence Namysl, Mikaël Chibok, Nicolas Pelé, Gregory Swielinski, Marc-Olivier Thibaut, Samuel Dubé, Daryosh Arsalan, Marc-André Tessier, Arnaud Geneste et Élisabeth Bouchard.

Merci à mes parents, ma famille et mes amis pour leur support depuis le début et plus particulièrement, merci à Olivia Pia Audet et Charles Farmer pour votre présence et votre support continu au fil des années.

## RÉSUMÉ

La mobilité urbaine étant de plus en plus un enjeu, de nombreux plans et projets sont proposés afin d'améliorer la qualité de l'offre de transport collectif et d'y attirer une quantité croissante d'utilisateurs. Le présent mémoire s'inscrit dans un plus vaste projet d'étude visant à identifier les impacts qu'aura le développement urbain de la grande région métropolitaine de Montréal à l'horizon 2031. Dans cette optique, ce projet de recherche vise à mettre sur pied des modèles de choix modal qui mettront en évidence le niveau de performance des divers modes, tant motorisés (auto, transport collectif) qu'actifs (marche, vélo). De tels modèles permettront une comparaison aisée de divers scénarios d'offre de transport et de répartition de la population. Pour y parvenir, ce projet compte deux objectifs principaux :

- Développer une méthode d'analyse spatiale permettant de caractériser la demande actuelle de transport et de corrélérer cette demande avec les niveaux de service actuels des différents modes de transport (scénario de référence).
- Développer un modèle de choix modal basé principalement sur les caractéristiques des déplacements et sur le niveau de service des différents modes, exprimé sous forme de ratios de compétitivité.

La première partie du mémoire est une revue de littérature mettant en lumière l'ensemble des variables explicatives du choix modal les plus couramment utilisées. Certaines de ces variables seront sélectionnées afin d'être utilisées dans deux méthodes de modélisation également présentées dans cette section, l'arbre de décision et le modèle de choix discret « logit ». Finalement, une méthode d'analyse spatiale ayant le potentiel d'identifier les zones privilégiées ou négligées de la desserte de transport collectif est également présentée.

La seconde partie présente la structure de la méthodologie proposée et compare l'ensemble des données utilisées dans ce mémoire. Toutes basées sur l'enquête Origine-Destination 2008 de la Grande région de Montréal, la première source d'information provient de simulations effectuées au ministère des Transports, de la Mobilité durable et de l'Électrification des transports tandis que la seconde provient de simulations de la Chaire Mobilité. Présentant des temps de parcours simulés pour chacun des modes et ce, pour l'ensemble des déplacements, la comparaison de ces données offre un aperçu des forces et des faiblesses de chacun des modèles de simulation. À partir de ces

temps de parcours, le concept de « ratio de compétitivité » est développé et sera à la base des analyses subséquentes.

La troisième section dresse le portrait des habitudes de mobilité de la population du Grand-Montréal de 2008. L'analyse porte principalement sur trois facteurs clés : l'heure de départ, la distance du déplacement et le niveau de performance relatif du mode choisi. À cette fin, une segmentation de la journée basée sur le test statistique Kolmogorov-Smirnov est proposée. L'analyse de la performance de chacun des modes par rapport à la distance à parcourir révèle que pour les courts déplacements, le transport collectif ne semble pas être une alternative alors que la marche est particulièrement attractive. Le vélo, bien que jouissant d'une très faible part modale, est particulièrement performant, et ce, peu importe la distance du déplacement. Un regroupement des déplacements selon leurs vecteurs de ratios de compétitivité est présenté tentant de mettre en lumière des corrélations entre la part modale, la distance et les ratios de compétitivité. Finalement, une analyse spatiale identifie certains secteurs privilégiés ou négligés par le transport collectif.

La quatrième section est celle qui présente l'ensemble des étapes de modélisation du choix modal. Débutant par l'analyse d'indépendance des variables, plusieurs arbres de décisions et modèles logit sont présentés. Différents arbres sont présentés pour représenter les critères de choix modal propres aux différents motifs de transports, qu'ils s'agissent des étudiants, des travailleurs ou de ceux qui vont magasiner.

La cinquième section vise à évaluer, à l'aide des modèles développés à la section précédente, l'évolution des habitudes de transport de la population en fonction des divers scénarios d'offre de transport et de répartition de la population. La codification des différents réseaux de transport n'ayant été complétée à temps pour être étudiée dans ce mémoire, cette section agit en tant que la démonstration méthodologique en analysant le scénario de référence à l'horizon 2031.

La sixième et dernière section regroupe les conclusions, recommandation, limitations et contributions de ce mémoire. La contribution principale de ce projet est sans doute les divers modèles mettant en lumière l'impact des ratios de compétitivité dans la mobilité des montréalais. Les diverses hypothèses simplificatrices, tel que l'exclusion de certains déplacements, et diverses limitations, telles que l'exclusions de données potentiellement pertinentes, y sont résumées et discutées.

## ABSTRACT

Urban mobility being an increasingly challenging issue, numerous plans and projects are proposed to improve the quality of the public transit so as to attract more riders. This research is conducted as part of a wider research project aiming to identify the impacts of transportation and urban scenarios on public health at the 2031 horizon. This specific research aims to model the modal choice of the population with respect to relative performance of motorised (car, public transit) and active (walk, bicycle) modes. Such models aim to facilitate the comparison between different scenarios of public transit development and population distribution. This project has two main objectives:

- Develop a spatial analysis method able to characterize the current transit demand and to link this demand to estimated levels of service (baseline scenario).
- Develop a mode choice model mainly based on the trips characteristics and on the level of service of different modes, expressed as competitiveness ratios.

The first section of the essay is a literature review enlightening a wide array of the most commonly found variables explaining mode choice. Some of these variables are selected to feed two modeling approaches also presented in this section, the decision tree and the discrete choice model “logit”. Finally, a spatial analysis method able to identify privileged or neglected area of public transit service is also presented.

The second section presents the structure of the proposed methodology and compares the different sources of data used in the analysis. Both based on the 2008 Origin-Destination survey, the first source of data comes from simulations made by the ministère des Transports, de la Mobilité durable et de l'Électrification des transports while the second comes from simulations made by Chaire Mobilité. Accounting for simulated transit time for each trip observed in the OD survey, the comparison of those data shed a light on the strength and weaknesses of each simulation platform. Based on the estimated trip durations, the “competitiveness ratio” concept is developed and used in the following analysis.

The third section describes the travel behaviours of Montrealers as observed in the 2008 survey. This analysis is mostly based on three key factors: the starting time, the distance to be travelled and the relative performance of the chosen mode. The time segmentation of the day has been made

based on statistical difference on the Kolmogorov-Smirnov test. The performance of each mode with respect to distance reveals that public transit is not an option for short trips and walking is preferred. Cycling, although barely used, seems very performant regardless of the distance. Based on vectors of competitiveness ratio between modes, a classification of trips is developed to analyse the correlation between mode choice, travel distance and mode ratios. Finally, a spatial analysis identifies some areas where public transit shows good performances and others where it needs improvement.

The fourth section presents the steps of the mode choice modeling. Starting with the independence of irrelevant alternative analysis required for the logit model, multiple decision trees and logit models are presented. The different trees allow to analyse important variables for various trip purposes (school, work or shopping).

The fifth section aims to evaluate, based on the models developed in the previous section, the evolution in travel behaviours throughout different scenarios of transportation supply and population distribution. The coding of these scenarios has not been done as part of this research, therefore this section acts as a proof of concept for the global method projecting the 2031 baseline scenario.

The sixth and final section groups the conclusions, the recommendations, the limitations and the contributions of this essay. The main contributions are the different models enlightening the impact of the competitiveness ratios in the Montrealers' daily travel. The various simplifying assumptions, such as the exclusion of some trips, and diverse limitations, such as the removal of potentially pertinent data, are presented and discussed.

## TABLE DES MATIÈRES

DÉDICACE.....	III
REMERCIEMENTS .....	IV
RÉSUMÉ.....	V
ABSTRACT .....	VII
TABLE DES MATIÈRES .....	IX
LISTE DES TABLEAUX.....	XIII
LISTE DES FIGURES.....	XIV
LISTE DES SIGLES ET ABRÉVIATIONS .....	XVIII
LISTE DES ANNEXES.....	XX
CHAPITRE 1 INTRODUCTION.....	1
1.1    Mise en contexte.....	1
1.2    Objectifs .....	2
1.3    Structure du document .....	3
CHAPITRE 2 REVUE DE LITTÉRATURE .....	4
2.1    Description de l'approche de prévision de la demande au Québec.....	4
2.1.1    Prévision de la demande.....	5
2.1.2    Prévision de l'offre.....	7
2.2    Choix modal et mobilité.....	8
2.2.1    Définition .....	8
2.3    Variables explicatives du choix modal.....	10
2.3.1    Variables liées aux caractéristiques sociodémographiques.....	12
2.3.2    Variables liées aux indicateurs spatiaux.....	15
2.3.3    Variables liées aux caractéristiques du déplacement .....	17

2.3.4	Autres variables.....	20
2.3.5	Modèles logit ou de régressions logistiques.....	20
2.3.6	Arbre de décision .....	22
2.4	Méthodes d'analyse spatiale.....	23
2.4.1	Définition .....	23
2.4.2	Notions de voisinage .....	24
2.4.3	Indicateurs locaux d'autocorrélation spatiale.....	25
2.5	Méthodes de classification des déplacements .....	27
2.5.1	Classification par k-moyennes .....	27
2.5.2	Clustering hiérarchique .....	29
2.6	Synthèse générale .....	31
	CHAPITRE 3 MÉTHODOLOGIE GÉNÉRALE.....	32
3.1	Structure générale de la méthode .....	32
3.2	Définition/Justification des ratios de compétitivité.....	33
3.3	Sources et description des données .....	35
3.3.1	Enquête OD .....	35
3.3.2	Simulations issues du modèle régional .....	37
3.3.3	Simulation Chaire Mobilité.....	39
3.4	Comparaison MOTREM vs Chaire Mobilité .....	43
3.4.1	Comparaison des temps auto.....	44
3.4.2	Analyse spatiale des écarts dans les résultats.....	47
3.4.3	Comparaison des temps TC.....	49
3.4.4	Analyse spatiale des erreurs .....	51
3.5	Préparation des données .....	55

3.5.1	Filtre #1 – Observations des non-mobiles.....	55
3.5.2	Filtre #2 – Départs avant 5h et après 24h.....	56
3.5.3	Filtre #3 – Personnes de 4 ans et moins .....	56
3.5.4	Filtre #4 – Déplacements hors zone .....	56
3.5.5	Filtre #5 – Impossibilité de simuler les déplacements avec certains modes .....	57
<b>CHAPITRE 4 SITUATION DE RÉFÉRENCE 2008 .....</b>		<b>59</b>
4.1	Statistiques descriptives .....	59
4.2	Compétitivité et heure de départ .....	60
4.3	Segmentation horaire des déplacements .....	61
4.4	Compétitivité et distance.....	66
4.4.1	Distances seuils .....	67
4.5	Regroupement des vecteurs de compétitivité.....	72
4.5.1	Définitions des vecteurs .....	73
4.5.2	Résultat du regroupement.....	73
4.6	Portrait global des ratios de compétitivité.....	76
4.7	Analyse d'autocorrélation spatiale par période horaire.....	79
<b>CHAPITRE 5 MODÈLES DE CHOIX MODAL .....</b>		<b>83</b>
5.1	Arbre de décision .....	83
5.1.1	Non-détenteurs de permis de conduire.....	85
5.1.2	Détenteurs de permis de conduire .....	86
5.1.3	Travailleurs adultes en heure de pointe am.....	89
5.1.4	Déplacements travail .....	90
5.1.5	Déplacements école .....	91
5.1.6	Déplacements magasinage .....	94

5.2	Logit multinomial.....	95
5.2.1	Indépendance des variables explicatives.....	96
5.2.2	Résultats du modèle .....	98
5.2.3	Modèle excluant la possession du permis de conduire et du titre TC .....	107
CHAPITRE 6 CONCLUSION ET RECOMMANDATIONS .....		109
6.1	Synthèse des travaux .....	109
6.2	Rappel des contributions .....	110
6.3	Limitations .....	111
6.4	Perspectives .....	112
BIBLIOGRAPHIE .....		114
ANNEXES .....		117

## LISTE DES TABLEAUX

Tableau 3-1 : Structure du vecteur de ratios de compétitivité.....	35
Tableau 3-2 : Statistiques descriptives de DT et RT, comparaison temps auto CM et MO .....	44
Tableau 3-3 : Indices de Moran des écarts de temps de parcours auto .....	49
Tableau 3-4 : Statistiques descriptives, comparaisons Chaire Mobilité et MOTREM .....	50
Tableau 3-5 : Indicateurs de Moran .....	51
Tableau 3-6 : Étapes de filtration des données.....	55
Tableau 3-7 : Données disponibles par mode .....	57
Tableau 4-1 : Statistiques descriptives du scénario de référence .....	59
Tableau 4-2 : Groupes horaires issus du test KS.....	64
Tableau 4-3 : Définition des vecteurs pour clustering .....	73
Tableau 4-4 : Caractéristiques des 9 clusters .....	75
Tableau 4-5 : Coefficients de corrélation linéaire .....	76
Tableau 4-6 : Valeur seuils des quintiles de ratios de compétitivité .....	77
Tableau 5-1 : Variables utilisées (noir) et retirées (rouge) dans le modèle logit .....	95
Tableau 5-2 : Présentation détaillée des variables explicatives .....	96
Tableau 5-3 : Rapports de cotes .....	100
Tableau 5-4 : Variables explicatives utilisées .....	107
Tableau 5-5 : Rapports de cotes .....	108

## LISTE DES FIGURES

Figure 2-1 : Schéma du processus prévisionnel de la demande au Québec (PB Consult Inc, 2002)	5
.....	.....
Figure 2-2: Schéma relationnel du scénario de demande (PB Consult Inc, 2002).....	6
Figure 2-3: Schéma relationnel du scénario d'offre (PB Consult Inc, 2002) .....	7
Figure 2-4 : Cycle de rétroaction transport - utilisation du sol (Wegener, 2004).....	9
Figure 2-5 : Détails du cycle de rétroaction transport-utilisation du sol (Wegener, 2004).....	9
Figure 2-6 : Cadre structurant des facteurs déterminants du choix modal (traduction libre de : De Witte et al. 2013).....	10
Figure 2-7 : Distribution des variables explicatives étudiées (De Witte et al. 2013).....	11
Figure 2-8 : Distribution de la population et des membres de services de vélopartage en fonction du revenu (CityLab, 2015) .....	14
Figure 2-9 : Forme générique des courbes de partage modal (Bonnel 2001) .....	21
Figure 2-10 : Types de contiguïté (Sawada, 2009) .....	24
Figure 2-11 : Valeur p et valeur seuil z (Desktop.ArcGIS.com – Spatial Statistics Toolbox) .....	26
Figure 2-12 : Type des zones des clusters de LISA .....	26
Figure 2-13 : Processus d'identification des clusters de LISA (Desktop.ArcGIS.com – Anselin's Local Moran's I).....	27
Figure 2-14 : Exemple de l'algorithme des k-moyennes.....	29
Figure 2-15 : Exemple de graphique WS .....	30
Figure 3-1 : Schéma méthodologique global .....	33
Figure 3-2 : Ménages enquêtés dans l'enquête OD 2008, par zone hexagonale de 1 km <sup>2</sup> .....	37
Figure 3-3 : Zones analyse transport (ZAT) utilisées dans le Motrem08 .....	38
Figure 3-4 : Distribution des temps d'attente initiaux moyens TC.....	41
Figure 3-5 : Temps de parcours moyens des déplacements en TC selon l'heure de départ.....	42

Figure 3-6 : Distribution des heures de départ des déplacements en TC* .....	42
Figure 3-7 : Distribution des différences de temps de parcours.....	45
Figure 3-8 : Distribution des différences de temps de parcours Chaire-MOTREM (mins) en fonction de la distance parcourue.....	46
Figure 3-9 : Distribution des ratios de temps de parcours.....	46
Figure 3-10 : Distribution des ratios de temps de parcours en voiture sous congestion fonction de la distance parcourue .....	47
Figure 3-11 : Différences de temps de parcours moyens, auto sous congestion, en minutes .....	48
Figure 3-12 : Ratios moyens par zone, temps de parcours en auto sous congestion .....	48
Figure 3-13 : Distribution des différences de temps de parcours TC*-CM vs TC-MO.....	50
Figure 3-14 : Distribution des ratios de temps de parcours TC*-CM vs TC-MO .....	51
Figure 3-15 : Clusters de Moran pour différences de temps de parcours TC* .....	53
Figure 3-16 : Répartition spatiale des temps d'attentes initiaux en TC, légende en secondes .....	54
Figure 3-17 : Cluster de Moran des temps d'attentes initiaux .....	54
Figure 3-18 : Déplacements avant filtre #4.....	57
Figure 4-1 : Ratios moyens par période horaire .....	61
Figure 4-2 : Illustration de la statistique D du test KS comparant les distributions de 5h-6h59 et 7h-9h59 .....	63
Figure 4-3 : Extrait de la matrice des statistiques D issues du test KS pour chaque paire de FDC .....	63
Figure 4-4 : Paires d'heures correspondant au critère $D < 7\%$ .....	64
Figure 4-5 : Distribution des déplacements durant la journée .....	65
Figure 4-6 : Distribution des Ratios TC* par groupe horaire .....	66
Figure 4-7 : Ratios moyens en fonction de la distance .....	67
Figure 4-8 : Déplacements cumulés par mode en fonction de la distance .....	68

Figure 4-9 : Part modale en fonction de la distance .....	68
Figure 4-10 : Distances seuil pour la marche par groupe démographique .....	69
Figure 4-11: Distances seuil pour le vélo par groupe démographique .....	70
Figure 4-12 : Ratios moyens marche/TC* en fonction de la distance réseau .....	71
Figure 4-13 : Distribution des déplacements TC en fonction du ratio marche/TC* .....	72
Figure 4-14 : Sommes des carrés résiduels par groupe .....	74
Figure 4-15 : Dendrogramme complet avec 9 clusters .....	74
Figure 4-16 : Distribution des ratios de compétitivité par mode .....	77
Figure 4-17 : Temps de trajet domicile-trajet (Le Devoir, 4 décembre 2017) .....	78
Figure 4-18 : Indices de Moran par période horaire .....	80
Figure 4-19 : Clusters de Moran TC*, 7h am à 9h59 .....	81
Figure 4-20 : Clusters de Moran Vélo, 7h am à 9h59 .....	82
Figure 4-21 : Clusters de Moran Marche, 7h am à 9h59 .....	82
Figure 5-1 : Contenu des feuilles des arbres de décision .....	83
Figure 5-2 : Arbre de décision, tous les déplacements .....	84
Figure 5-3 : Arbre de décision, non détenteurs de permis de conduire .....	86
Figure 5-4 : Arbre de décision, détenteurs de permis de conduire .....	87
Figure 5-5 : Arbre de décision excluant les facteurs de possession du permis de conduire, de titre mensuel TC et de motorisation du ménage .....	88
Figure 5-6 : Arbre des travailleurs en heure de pointe am .....	90
Figure 5-7 : Arbre de décision, adultes motif travail .....	91
Figure 5-8 : Arbre de décision, déplacements école, tous les âges .....	92
Figure 5-9 : Arbre de décision, déplacements école adultes .....	93
Figure 5-10 : Arbre de décision, déplacements motif magasinage .....	94
Figure 5-11 : Facteurs de corrélation dépassant le seuil .....	97

Figure 5-12 : Distribution des résidus du modèle logit.....	103
Figure 5-13 : Distribution des parts modales en fonction des probabilités de choisir le mode auto conducteur .....	104
Figure 5-14 : Distribution des parts modales en fonction des probabilités de choisir le mode auto passager .....	104
Figure 5-15 : Distribution des parts modales en fonction des probabilités de choisir le mode TC .....	105
Figure 5-16 : Distribution des parts modales en fonction des probabilités de choisir le mode vélo .....	105
Figure 5-17 : Distribution des parts modales en fonction des probabilités de choisir le mode marche .....	106
Figure 5-18 : Distribution des parts modales en fonction des probabilités de choisir un autre mode .....	106

## LISTE DES SIGLES ET ABRÉVIATIONS

AMT	Agence métropolitaine de transports, aujourd’hui ARTM et RTM
ARTM	Autorité régionale du transport métropolitain
CM	Chaire Mobilité
CMM	Communauté métropolitaine de Montréal
DT	Différence des temps de parcours
FDC	Fonction de distribution cumulative
GTFS	General Transit Feed Specification
IRSPUM	Institut de recherche en santé publique de l’Université de Montréal
ISQ	Institut de la statistique du Québec
LISA	Local indicators of spatial association – Indicateur local d’auto-corrélation spatiale
MO	MOTREM08
MRC	Municipalités régionales de comté
MTQ	Ministère des Transports du Québec, aujourd’hui MTMDET
MTMDET	Ministère des Transports, de la Mobilité durable et de l’Électrification des transports
KS	Kolmogorov-Smirnov
OD	Origine-Destination
OSRM	Open Source Routing Machine
PMAD	Plan Métropolitain d’Aménagement et de Développement
R_TC	Ratio de compétitivité TC/Auto
R_TC*	Ratio de compétitivité TC*/Auto
R_Vel	Ratio de compétitivité Vélo/Auto
R_Mar	Ratio de compétitivité Marche/Auto
RT	Ratio des temps de parcours
REM	Réseau électrique métropolitaine
RTM	Réseau de transport métropolitain
SIG / GIS	Systèmes d’information géographique
SLR	Système léger sur rail
SRB	Service rapide par bus

TC	Transport collectif ou transport en commun
TP	Temps de parcours
WSS	Within-cluster sum-of-square – Somme des erreurs quadratiques
ZAT	Zone analyse transport

## **LISTE DES ANNEXES**

ANNEXE A – Matrice complète – Test ks – ref08.....	117
ANNEXE B – Carte des ratios TC*, 7h à 9h59 .....	118
ANNEXE C – Arbre de décision, non-détenteurs de permis de conduire .....	119
ANNEXE D – Arbre de décision, détenteurs de permis de conduire .....	120

## CHAPITRE 1 INTRODUCTION

Aujourd’hui plus que jamais, l’enjeu de la mobilité est au cœur des débats sociétaux. Que l’on pense au débat du 3<sup>e</sup> lien à Québec, au REM ou au prolongement des lignes de métro à Montréal, plusieurs projets structurants sont en phase d’étude ou de réalisation. Dans un souci de développement stratégique, la Communauté métropolitaine de Montréal (CMM) a développé un Plan métropolitain d’aménagement et de développement (PMAD) (CMM 2011). Les orientations, objectifs et critères de ce plan ont pour mission d’encadrer les outils dont disposent les agglomérations et les municipalités régionales de comté (MRC) au sujet de l’offre de transport collectif, de l’aménagement du territoire et de la protection de l’environnement. Parmi ces objectifs figure le développement de nouvelles infrastructures de santé, d’éducation, sportive et culturelle à proximité des services de transport collectif (TC) et le souhait de diriger au moins 40% des nouveaux ménages dans les aires « Transit-Oriented Development » (TOD). Le PMAD prévoit, en effet, la croissance de la population de 530 000 personnes à l’horizon 2031. Le plan souhaite finalement favoriser le développement du réseau de transport en commun métropolitain afin de faire passer la part modale du TC de 25% à 30% en 2021 et à 35% en 2031, le tout en respectant les espaces et les ressources financières disponibles.

Afin d’atteindre ses objectifs, la CMM a proposé divers projets et politiques visant principalement le développement du réseau de transport collectif. Parmi ceux-ci, notons des projets de prolongement de métro, de développement de lignes de trains de banlieue, de construction de Système Léger sur Rail (SLR) et/ou de mise en place de Services Rapides par Bus (SRB). Le PMAD met ainsi de l’avant une série de recommandations dans une perspective de développement durable « afin d’assurer la compétitivité et l’attractivité du Grand-Montréal » (PMAD, 2012), le tout se chiffrant à 12,6 milliards de dollars à l’horizon 2031. Pour des projets de cet ordre de grandeur, il est important de développer des modèles prédictifs des impacts potentiels.

### 1.1 Mise en contexte

S’inscrivant dans la volonté de planification stratégique du Grand-Montréal de demain, un collectif de scientifiques cherche à quantifier les impacts sur la santé publique que peuvent avoir les divers scénarios de développement de la région proposés par le PMAD. Ce collectif est composé de chercheurs en santé publique de l’Institut de recherche en santé publique (IRSPUM) et de la Chaire

Mobilité, chaire de recherche sur l'évaluation et la mise en œuvre de la durabilité en transport de Polytechnique Montréal.

Les scénarios du PMAD ayant des composantes de répartition de la population et d'offres de transport, l'équipe de santé publique a procédé au recensement de l'ensemble des projets majeurs de développement immobilier. Riche de cette base de données, ils ont pu proposer des scénarios de répartition de la population contrastés qui serviront de base pour les simulations de transport.

*« Il y a des projets de transport et d'aménagement de milliards de dollars qui nous léguent une forme de ville. Ce qui importe, ce ne sont pas les petits chantiers, ce sont les gros projets qui structurent la ville, avec la place qu'on laisse ou non pour l'automobile. »*

*Patrick Morency (19 juillet 2016, Le Devoir)*

L'équipe transport a comme objectif de développer une méthode d'analyse des impacts de ces scénarios de répartition de la population en fonction des différents niveaux d'offre TC.

Ce mémoire vise donc à proposer des modèles illustrant l'impact qu'a la performance des modes actifs et collectifs dans les comportements de mobilité de la population. Se basant sur l'état de l'art en termes de modélisation des transports, des modèles d'estimation du mode de transport sur la base du statu quo seront élaborés en fonction de divers indicateurs et variables.

## 1.2 Objectifs

L'objectif principal est de développer une méthode d'analyse qui pourra évaluer les impacts de modifications de l'offre de transport et de la localisation résidentielle sur le choix modal. Les scénarios envisagés proviennent du PMAD et sont à l'horizon 2031. L'objectif ultime est de développer une méthode pouvant quantifier les habitudes de transport d'une population projetée en fonction d'une offre de transport donnée.

Pour y parvenir, ce projet compte deux sous-objectifs :

- Développer une méthode d'analyse spatiale permettant de caractériser la demande actuelle de transport et de corrélérer cette demande avec les niveaux de service actuels des différents modes de transport (scénario de référence).

- Développer un modèle de choix modal basé principalement sur les caractéristiques des déplacements et sur le niveau de service des différents modes, exprimé sous forme de ratios de compétitivité.

### **1.3 Structure du document**

Ce mémoire est structuré en six chapitres. Outre l'introduction, le chapitre 2 rassemble une présentation de l'état de l'art ainsi que l'essentiel du contenu théorique nécessaire à la compréhension du contenu du mémoire sous forme de revue de littérature. Le chapitre 3 présente exhaustivement les étapes de la méthode d'analyse proposée. Les chapitres 4 et 5 sont liés au processus de modélisation. On retrouve dans le chapitre 4 le scénario de référence alors que les modèles développés sont présentés au chapitre 5. Finalement, le chapitre 6 conclut le tout en présentant les limitations et les perspectives de recherche et/ou d'application.

## **CHAPITRE 2 REVUE DE LITTÉRATURE**

Ce chapitre offre une revue des méthodes en place présentement dans le contexte québécois en termes de prévision de la demande en transport ainsi que de la littérature pertinente à la problématique de recherche. Cette revue devrait donc fournir l'ensemble du contenu théorique nécessaire à l'appréciation de ce mémoire. Les sujets abordés débuteront par la description de l'approche prévisionnelle de la demande de transport au Québec suivie d'une définition de la mobilité et des variables explicatives du choix modal avant de terminer avec des méthodes d'analyse spatiale.

### **2.1 Description de l'approche de prévision de la demande au Québec**

Développé à la fin des années 1970 par le Ministère des Transports du Québec (maintenant MTMDET), le modèle de transport urbain pour la région de Montréal (MOTREM) est le modèle utilisé par le MTMDET depuis pour effectuer ses analyses prévisionnelles de la demande de transport et de l'utilisation des réseaux de transport. Ce modèle s'appuie sur deux composants indépendants : l'offre de transport, caractérisée par le réseau de transport collectif, le réseau routier ainsi que le niveau de service, et la demande de transport, caractérisée par la population et ses habitudes de mobilité. Présentés respectivement en bleu et en vert, sur la Figure 2-1, les scénarios générés sont définis dans les différents logiciels de simulation afin de projeter les impacts sur les réseaux de transport et de développer des plans de développement stratégique. Les différentes composantes de cette démarche sont expliquées ci-dessous.

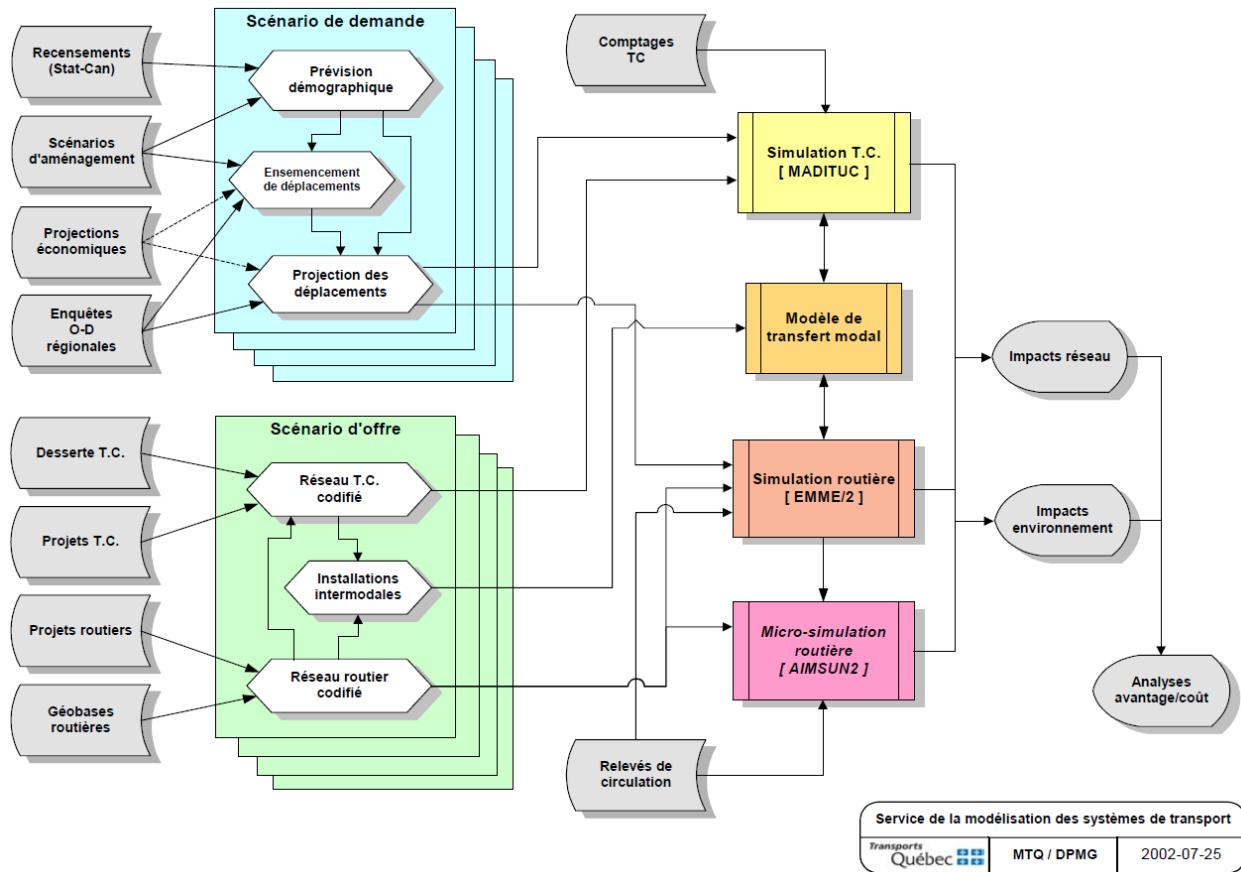


Figure 2-1 : Schéma du processus prévisionnel de la demande au Québec (PB Consult Inc, 2002)

### 2.1.1 Prévision de la demande

Les scénarios de la demande de transport, tel qu'illustré en Figure 2-1 et mis en évidence à la Figure 2-2, sont élaborés à partir des enquêtes Origine-Destination réalisées par un consortium d'autorités régionales de transport, dont l'Autorité régionale de transport métropolitain (ARTM) (anciennement l'Agence métropolitaine de transport (AMT)), ainsi que des prévisions démographiques basées sur les recensements canadiens.

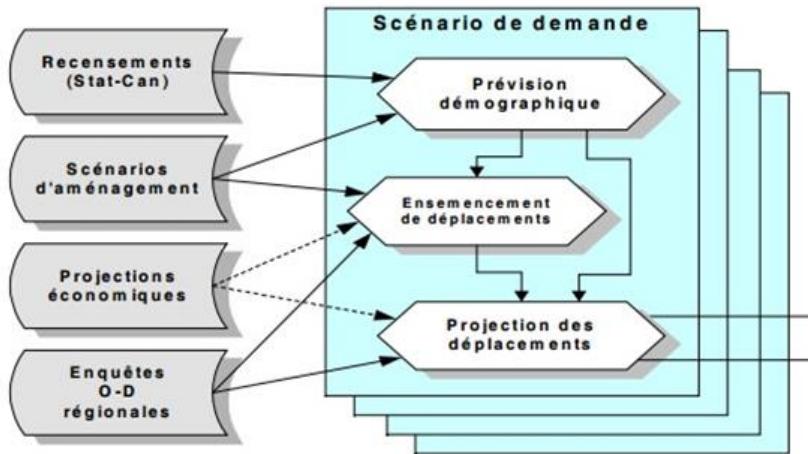


Figure 2-2: Schéma relationnel du scénario de demande (PB Consult Inc, 2002)

L'enquête Origine-Destination est la source d'informations principale. Tous les cinq ans, cette enquête sonde environ 5% de la population afin de créer un portrait des habitudes de déplacements des montréalais. En plus de recueillir les informations sur la personne, le ménage et les déplacements, un facteur de pondération est assigné à chaque ménage et à chaque personne afin de reconstituer le portrait démographique et de la mobilité de l'ensemble des résidents de la région.

En fonction du scénario à l'étude, des facteurs d'expansion liés à chaque déplacement sont définis en prenant en compte l'évolution prévue de la démographie, des statuts d'activités ainsi que du niveau de motorisation. (MTQ, 2002).

Basés sur les données de recensement de Statistique Canada, des scénarios d'évolution démographique sont réalisés avec le modèle ES-3. Développé au MTQ (devenu le MTMDT) en 1995 et basé sur les données de recensement de Statistiques Canada, le modèle ES-3 consiste en un modèle de projection démographique à l'échelle microscopique, c'est-à-dire au niveau des individus, intégrant les tendances observées aux données de recensements afin de générer des scénarios tendanciels de l'évolution de la population. Ainsi, à partir des populations dénombrées, les facteurs suivants sont considérés pour la réalisation des calculs : la fécondité, la migration, le vieillissement et la mortalité.

## 2.1.2 Prévision de l'offre

Reflétant l'ensemble des lignes de TC offertes à la population à une date donnée, les scénarios d'offre sont développés en trois étapes, également illustré à la Figure 2-1 et mis en évidence à la Figure 2-3. Premièrement, en codifiant un réseau routier à partir des géobases routières et des projets de développement routier. Les géobases sont des bases de données géoréférencées contenant des définitions précises de la géométrie des rues. Une fois le réseau routier codifié, les offres de transport collectif sont ensuite intégrées dans la codification d'un réseau TC. La plateforme de simulation TC utilisée au MTMDET se nomme MADITUC (Modèle d'Analyse Désagrégée des Itinéraires de Transport Urbain Collectif). Développée durant les années 90s à l'École Polytechnique de Montréal, cette plateforme a la particularité de faire ses simulations en régime permanent. Le réseau TC est défini par l'ensemble des lignes de transports, ainsi que les stations d'arrêts et les possibilités de correspondances. Il doit également inclure la définition du réseau piétonnier pour assurer une représentation adéquate de l'accès aux arrêts, à l'origine et à la destination.

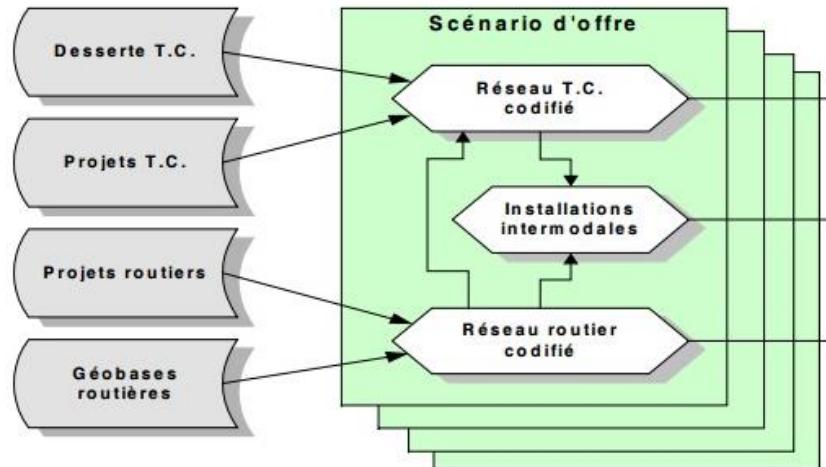


Figure 2-3: Schéma relationnel du scénario d'offre (PB Consult Inc, 2002)

Les deux réseaux étant jusqu'ici indépendants, la dernière étape est d'ajouter aux modèles des points de jonction intermodaux.

## 2.2 Choix modal et mobilité

Le choix modal est un composant majeur du portrait de la mobilité, tant à l'échelle de l'individu qu'à celle d'une communauté. Ce concept est au cœur de ce mémoire, il est important de bien le définir. Cette section présente donc une définition détaillée du choix modal et des facteurs qui l'influencent.

### 2.2.1 Définition

Le choix modal est ce que fait une personne lorsqu'elle considère l'ensemble des modes de transport disponibles et qu'elle en sélectionne un pour effectuer son déplacement. Ainsi, le processus de choix consiste à identifier quel mode de transport (auto, vélo, bus, marche, etc.) elle empruntera pour effectuer un déplacement donné. De nombreux facteurs peuvent influencer ce choix : en plus des caractéristiques intrinsèques au mode emprunté tel que le confort, les modalités d'usage, la disponibilité au domicile, le coût ou la capacité de chargement, d'autres facteurs comme le motif, la destination ou le fait d'effectuer une chaîne de déplacement peuvent avoir une incidence sur le choix d'un mode plutôt qu'un autre.

Wegener (2004) présente un cadre théorique mettant en relation les divers composants de ce qu'il nomme le cycle de rétroaction transport-utilisation du sol (« land-use transport feedback cycle »). Les Figure 2-4 et Figure 2-5 présentent respectivement ce cycle de manière sommaire et détaillée. La nature circulaire de ce modèle met en lumière l'impact du milieu bâti sur la mobilité et inversement, l'impact de la mobilité sur le milieu bâti.

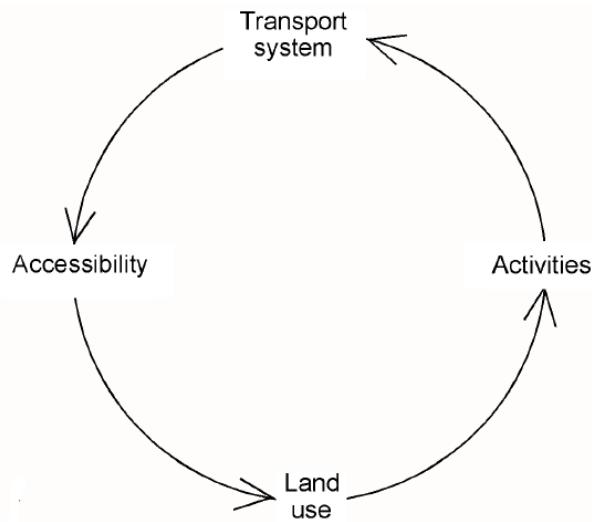


Figure 2-4 : Cycle de rétroaction transport - utilisation du sol (Wegener, 2004)

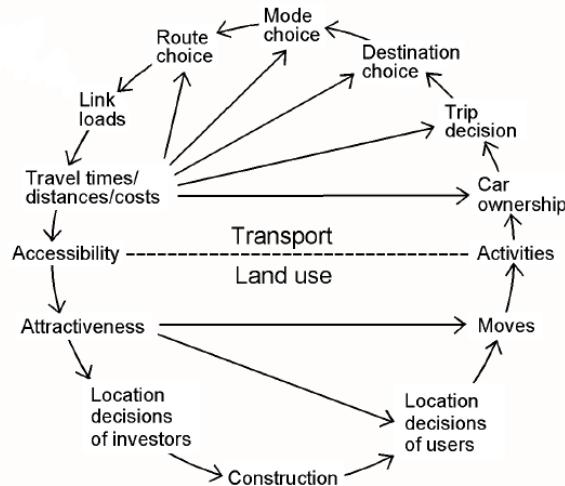


Figure 2-5 : Détails du cycle de rétroaction transport-utilisation du sol (Wegener, 2004)

Plusieurs facteurs expliquent les interactions transport-milieu bâti. Les facteurs tels que la densité résidentielle, la densité d'emplois et la mixité des usages peuvent influencer la durée des déplacements, leur fréquence ainsi que le mode emprunté pour les effectuer. De manière réciproque, l'accessibilité des transports va influencer le choix de localisation des unités résidentielles, industrielles, commerciales et de service.

## 2.3 Variables explicatives du choix modal

Les modèles de choix modaux s'appuient sur des variables explicatives. Étant variées et de natures diverses, ces variables peuvent ou non être corrélées au choix modal. La première étape est donc de faire un choix éclairé sur quelles variables serviront aux modèles.

Afin de classifier les différents facteurs pouvant influencer le mode, De Witte, Hollevoet, Dobruszkes, Hubert et Macharis (2013) proposent un schéma, présenté à la Figure 2-6, mettant en relation les diverses sphères d'indicateurs expliquant le choix modal tel que les facteurs socio-démographiques, les indicateurs spatiaux, les indicateurs des caractéristiques des déplacements ainsi que les indicateurs socio-psychologiques. Ces derniers émergent de travaux de sociologie et peuvent être liés au concept de motilité de Kaufmann (2004). Ce concept, emprunté au domaine de la biologie, réfère à la capacité de se déplacer. Bien que de plus en plus cité dans la littérature pour comprendre le processus de choix, il ne sera pas traité dans ce mémoire puisqu'il n'influera pas sur la forme du modèle développé dans ce projet.

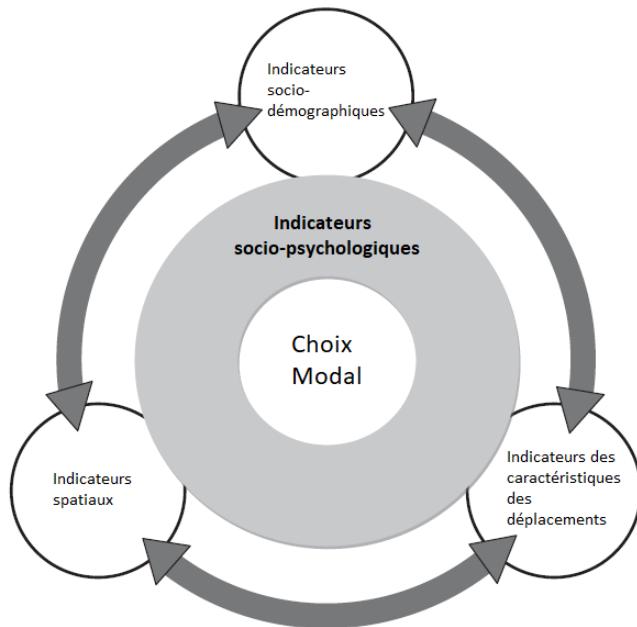


Figure 2-6 : Cadre structurant des facteurs déterminants du choix modal (traduction libre de : De Witte et al. 2013)

Ce cadre correspond également à la classification classique présentée par Bonnel (2001). Bonnel parle des trois catégories de variables relatives respectivement à l'offre de transport, aux

caractéristiques du déplacement et aux caractéristiques de l'individu ou du ménage, qui correspondent respectivement aux indicateurs spatiaux du déplacement et aux propriétés sociodémographiques.

La Figure 2-7 proposée par De Witte et al. (2013) présente les indicateurs étudiés dans les articles traitant le choix modal recensés dans leur article. L'axe horizontal représente la fréquence à laquelle l'indicateur est présent dans la littérature étudiée, c'est-à-dire que le temps de parcours est étudié dans près de 60% des papiers recensés par De Witte et al (2013). L'axe vertical représente la fréquence dans les articles à laquelle les variables ont été démontrées représentatives. Par exemple, la notion de chaînage n'est pas beaucoup étudiée, étant présente dans moins de 20% des études en dépit d'être trouvée significative dans près de 4 études sur 5. Cette notion réfère aux possibles contraintes inhérentes à des déplacements effectués en chaîne, sans retour au domicile. Cette notion sera expliquée plus en détail en section 2.3.3.7.

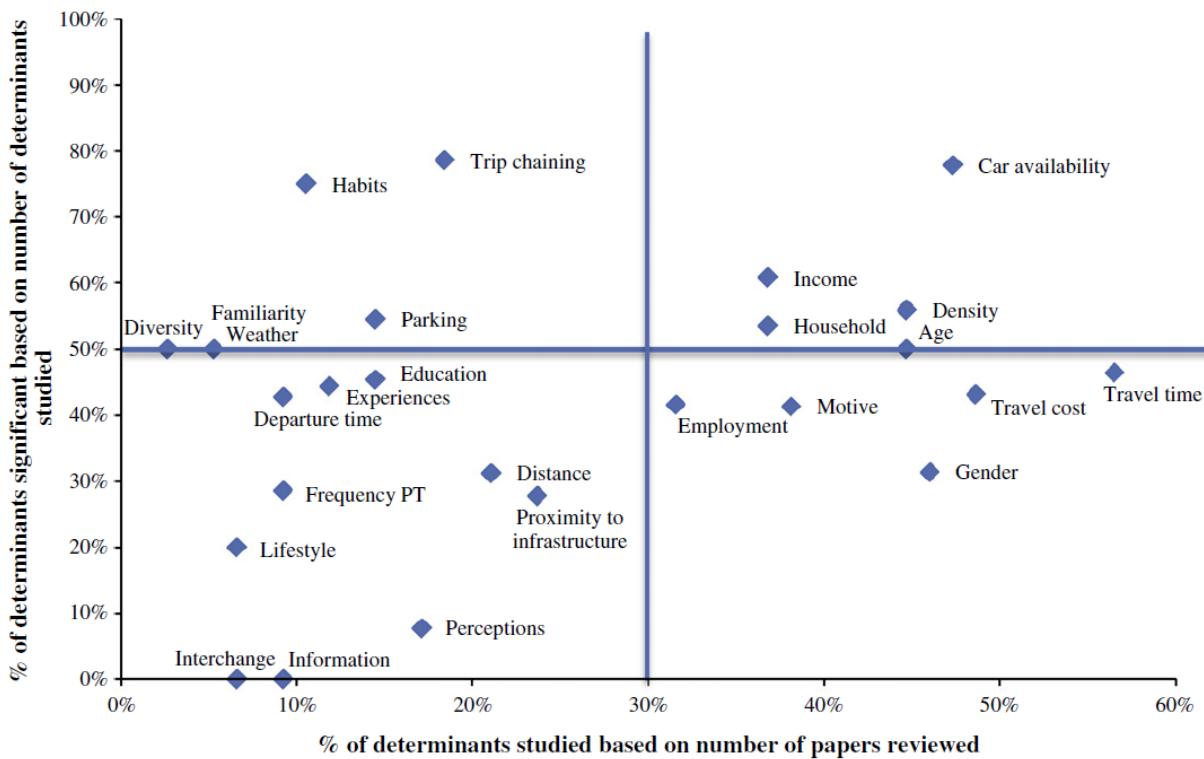


Figure 2-7 : Distribution des variables explicatives étudiées (De Witte et al. 2013)

Avec la multitude de facteurs pouvant jouer un rôle dans la modélisation du choix modal, il n'est pas rare de rencontrer d'autres indicateurs ou encore des indicateurs construits qui présentent une nouvelle orientation sur un concept connu. Parmi ceux-ci, Polquin (2012) propose une série

d'indicateurs spatiaux novateurs représentant les caractéristiques du milieu bâti de manière géo-centrée sur chacun des ménages alors que Frappier (2015) propose des indicateurs basés sur la qualité et la diversité des itinéraires de transport en commun dans son analyse des trajets alternatifs.

Les sections suivantes présentent une vaste sélection de variables explicatives, débutant par les caractéristiques socio-démographiques, suivies des caractéristiques spatiales et des caractéristiques des déplacements.

### **2.3.1 Variables liées aux caractéristiques sociodémographiques**

Cette première section présente les caractéristiques sociodémographiques pouvant être intrinsèques à la personne (âge, genre) ou extrinsèques (le revenu, la structure du ménage, le niveau de motorisation, etc.).

#### **2.3.1.1 Âge**

L'âge d'une personne reflète sa position dans son cycle de vie. Cette caractéristique informe sur deux aspects distincts, soit le statut social et les capacités physiques plausibles. De manière générale, on observe que le statut social augmente avec le temps et les réalisations des gens, alors que les capacités physiques tendent à diminuer avec le vieillissement. La littérature ne semble pas unanime sur l'impact qu'a l'âge sur le choix modal (De Witte et al. 2013); néanmoins, elle peut agir comme facteur limitant par rapport à certains modes, tel que la conduite automobile, qui exige de posséder un permis de conduire (Polquin 2012).

#### **2.3.1.2 Genre**

L'implication du genre sur le choix modal est fortement liée au rôle de la femme dans la société. Alors que depuis des générations, les hommes étaient les principaux pourvoyeurs du domicile familial et les femmes avaient un rôle domestique les menant à parcourir de moins grandes distances tout en étant plus indépendantes de l'automobile, on a assisté depuis l'après-guerre à une émancipation des femmes menant à leur entrée sur le marché du travail qui a provoqué une petite révolution au niveau de la mobilité des ménages. Parmi les conséquences de cette révolution, l'écart entre les habitudes de mobilité des hommes et des femmes s'est amoindri (Rosenbloom 2004, 2006).

Cette uniformisation des habitudes de mobilité entre les genres peut s'expliquer par divers facteurs. Ces facteurs sont présentés comme des facteurs de convergence. Parmi ceux-ci, nommons l'émancipation professionnelle des femmes, les rôles changeant dans les ménages, le fait de vivre seul et les ménages monoparentaux.

Il y a toutefois des incertitudes dans la communauté scientifique envers la pertinence de cette variable, puisque d'autre facteurs induisent des comportements de mobilité différents entre les genres (De Witte et al. 2013). Entre autres, l'équité au travail n'est pas atteinte, ce qui implique une différenciation au niveau des revenus, des types de postes occupés et du niveau de stabilité de l'emploi. Malgré l'évolution des rôles dans le ménage, ceux plus traditionnels sont toujours présents et contribuent à la différenciation des comportements de mobilité (Rosenbloom 2006).

### 2.3.1.3 Revenu

Le revenu est un des facteurs les plus significatifs dans la modélisation des choix modaux, selon la Figure 2-7. De Witte et al. (2013) présentent une liste d'auteurs qui s'entendent sur les affirmations suivantes :

- Le revenu est un facteur significatif dans le processus de modélisation du choix modal;
- Le revenu est positivement corrélé à l'utilisation de la voiture et inversement relié à l'utilisation du transport collectif;
- L'utilisation de l'auto en solo au lieu de l'autopartage et du TC est plus probable chez les personnes à hauts revenus;
- Les personnes à faibles revenus sont plus touchées par la variation des coûts de transport et peuvent être limitées dans leur mobilité.

Malgré ces observations, certains articles identifient une tendance inverse où une corrélation positive entre le revenu et l'utilisation du vélo peut être mesurée. C'est en autre le cas dans une étude sur le vélopartage présentée par le Mineta Transportation Institute en collaboration avec BIXI Montréal, BIXI Toronto, Nice Ride Minnesota, GreenBIKE SLC (Salt Lake City) et EcoBici, Mexico City (Shaheen, Martin, Chan, Cohen et Pogodzinski, 2014). Tel que rapporté par CityLab, journal électronique spécialisé en transport, dans son article du 30 octobre 2015 « The Poor Bike,

the Rich Bike-Share », la Figure 2-8 illustre que la proportion d'utilisateurs de services de vélopartage dans la population augmente avec le revenu du ménage.



Figure 2-8 : Distribution de la population et des membres de services de vélopartage en fonction du revenu (CityLab, 2015)

Un phénomène similaire survient lorsqu'on étudie le cas de l'enquête OD de la région de Montréal. Les ménages ayant le plus grand revenu disponible (revenu total du ménage divisé par la taille de celui-ci) sont aussi ceux qui marchent le plus. Ceci pourrait s'expliquer par la présence de nombreux ménages de professionnels célibataires vivant au centre-ville, qui sont donc ménages à haut revenu profitant de la plus grande densité et diversité de services de la région.

#### 2.3.1.4 Structure du ménage

La taille du ménage peut être liée à des comportements de mobilité particuliers. L'impact est généralement ressenti en termes de niveau d'utilisation de l'automobile et d'une plus grande complexité dans les chaînes de déplacements. De manière générale, plus la taille du ménage augmente, plus la probabilité d'utiliser l'automobile augmente (De Witte et al. 2013). La présence d'enfants a pour effet d'augmenter le niveau de complexité des chaînes de déplacements, ce qui est particulièrement ressenti lorsque ces enfants sont en bas âge (Poliquin 2012).

#### 2.3.1.5 Niveau de motorisation

Le niveau de motorisation, ou le niveau d'accès à une voiture, est un critère majeur dans le processus de choix modal, tel qu'illustré à la Figure 2-7. Ce paramètre a été trouvé significatif dans 78% des articles étudiés par De Witte et al. (2013).

Au niveau du ménage, la motorisation s'exprime par la quantité de voitures qui y est associée. Trois catégories sont alors utilisées, la non-motorisation, la mono-motorisation et la multi-motorisation, représentant respectivement l'absence de voiture, la présence d'une seule voiture et la présence de plus d'une voiture (Collet 2008).

Du point de vue de l'individu, le niveau de motorisation s'exprime par la quantité de véhicules disponibles par conducteur potentiel. Ce niveau s'exprime comme le ratio entre le nombre de véhicules possédés par le ménage et le nombre de personnes possédant un permis de conduire. On peut classifier le niveau de motorisation en trois niveaux : la motorisation nulle, partielle et complète. Le niveau de motorisation nulle (ratio = 0) correspond à l'absence totale de véhicule dans le ménage. Ce niveau de motorisation est fortement lié au phénomène de dépendance au TC (Cervero 2002). Le niveau de motorisation complète est atteint lorsque chaque membre du ménage qui possède un permis de conduire a accès à une voiture ou plus (ratio  $\geq 1$ ). Finalement, le niveau de motorisation partielle correspond à une utilisation partagée de la ou des voitures du ménage entre les conducteurs potentiels.

De façon générale, le niveau de motorisation d'un ménage croît en fonction de sa taille et de la distance séparant son domicile du centre-ville (Désilets 2012).

### **2.3.2 Variables liées aux indicateurs spatiaux**

Les variables liées aux indicateurs spatiaux comprennent les caractéristiques qui dépendent de l'environnement spatial dans lequel le déplacement, donc le choix modal, se déroule (De Witte et al. 2013). Ces variables incluent les indicateurs du milieu bâti tel que les « 5D » (Cervero Kockelman 1997; Ewing et Cervero 2001; Ewing et al. 2009), définis ci-dessous, et les indicateurs liés à l'offre de transport.

#### **2.3.2.1 Les 5D de Cervero**

De nombreuses études ont démontré l'influence du milieu bâti sur la mobilité, menant à de nombreuses explications, autant économiques que comportementales (Ewing & Cervero 2010). La classification des indicateurs spatiaux en « D » a débuté avec les « trois D » de Cervero et Kockelman (1997) : la densité, la diversité et le design. Plus tard, deux autres « D » se sont ajoutés, la distance d'accès au transport en commun et l'accessibilité à la destination (Ewing et Cervero

2001; Ewing et al. 2009). Depuis, un 6<sup>e</sup> et un 7<sup>e</sup> « D » ont vu le jour, mais ne sont pas des variables du milieu bâti : la gestion de la demande et l'aspect démographique (Ewing et Cervero, 2010).

La densité s'exprime en nombre d'éléments par unité de surface. Ces unités peuvent être des unités de population, de ménages, d'emplois ou autre. En jumelant la densité de la population et d'emploi, on obtient la densité d'activités (Cervero et Kockelman, 1997).

La diversité mesure les divers usages d'un secteur donné. Parmi les divers usages possibles, notons l'usage résidentiel, commercial et industriel. Cette variable s'exprime généralement comme un ratio emplois/ménages ou emplois/personnes (Cervero et Kockelman, 1997).

Le concept de design renvoi aux caractéristiques du réseau routier dans une région donnée. Le design a été principalement mesuré en densité d'intersections et en taille moyenne des quartiers. Toutefois, notamment dans le cas d'études de la mobilité des piétons, d'autres mesures peuvent être utilisées, telles que la densité de trottoirs, la largeur des rues, la quantité de traverses piétonnes et la présence d'arbres (Cervero et Kockelman, 1997).

Ajouté quelques années plus tard, l'accès à la destination mesure la facilité qu'ont les usagers de se rendre à la destination. Défini sur une échelle locale ou régionale, l'accès à la destination peut s'exprimer en unité de temps ou de distance. Handy (1993) définit l'accessibilité locale comme la distance entre le domicile et le commerce le plus près. D'un point de vue régional, l'accessibilité à la destination peut être la distance au centre-ville, ou encore la quantité d'emplois ou de lieux d'intérêt atteignables en un temps donné (Ewing et Cervero, 2010).

Le 5e « D » de Cervero est la distance d'accès au TC. Généralement exprimée comme la plus petite distance réseau séparant le domicile et le plus proche arrêt d'autobus ou station de métro, la distance d'accès peut aussi être exprimée en distance moyenne entre les arrêts, ou encore par densité d'arrêts dans une zone donnée.

### **2.3.2.2 Fréquence de transport collectif**

La fréquence de passage du service de transport collectif est un déterminant majeur de la disponibilité du service. Dans les zones urbaines denses, on retrouve généralement une plus grande densité de service, augmentant ainsi la compétitivité de ce mode par rapport aux autres. Toutefois, cette variable ne semble pas soulever beaucoup d'intérêt puisqu'elle est rarement trouvée significative (De Witte et al. 2013).

### 2.3.2.3 Disponibilité du stationnement

La disponibilité du stationnement devient un facteur important en milieu urbain dense. En effet, la rareté des places de stationnement peut devenir un facteur limitant pour l'utilisation de l'automobile (Kajita, Toi, Chishaki, Matsuoka, 2004). À l'inverse, on remarque que les travailleurs disposant d'une place de stationnement garantie à leur milieu de travail ont deux fois plus de chance de prendre la voiture que le TC comparativement à ceux qui n'en n'ont pas (Bonnel 2001).

Ce facteur déterminant mériterait d'être étudié davantage puisque, tel qu'illustré sur la Figure 2-7, il est trouvé significatif dans plus de 50% des cas mais n'est étudié que dans 13% des articles recensés par De Witte et al. (2013).

## 2.3.3 Variables liées aux caractéristiques du déplacement

Les caractéristiques liées aux déplacements sont celles qui définissent la nature du déplacement à effectuer. Étant généralement quantitatives, ces caractéristiques incluent entre autres la position des extrémités du déplacement, la distance, le motif, le temps de parcours, le coût, etc.

### 2.3.3.1 Extrémités

Les extrémités du déplacement sont évidemment des facteurs importants du choix modal puisque l'offre de transport n'est pas uniforme à travers le territoire. Un déplacement ayant son origine et sa destination dans un milieu urbain dense est plus enclin à avoir de bons niveaux d'accessibilité au transport collectif (Bonnel 2001).

### 2.3.3.2 Motif

Le motif du déplacement définit la raison même d'être du déplacement. Ces motifs peuvent être divers et se regroupent généralement en six catégories : travail, étude, magasinage, loisir, retour au domicile et autres. Certains modes de transport peuvent être plus adaptés pour des déplacements d'un motif donné alors que ce même mode peut ne pas être viable pour un autre motif. Par exemple, le transport en commun peut être très approprié pour se rendre à l'école, mais est souvent très peu pratique pour les déplacements professionnels (De Witte et al. 2013).

### 2.3.3.3 Distance

La distance à parcourir est un facteur déterminant dans le choix modal. En effet, les modes actifs ne peuvent tout simplement pas être utilisés pour parcourir de grandes distances comme le sont les modes motorisés. Dans la littérature, la notion de distance seuil réfère à la distance « qui cumule 80 % des déplacements faits à pied par un segment démographique particulier » (Polquin, Morency et Demers 2013). Au Québec, la distance seuil de la marche se situe généralement aux environs de 1 km et varie selon les tranches sociodémographiques.

De plus, pour les mêmes extrémités de déplacement, la distance à parcourir peut varier selon les modes. Les véhicules routiers sont contraints à circuler dans les rues en respectant le sens de la circulation, alors que le piéton peut emprunter des chemins alternatifs, tel que des chemins de parcs, des escaliers ou autres qui pourraient raccourcir ou allonger la distance à parcourir.

### 2.3.3.4 Temps de parcours et perception du temps

Que l'on pense à l'itinéraire choisi ou au niveau de congestion routière pour la voiture, à la fréquence de passage ou aux transferts requis pour un trajet en TC, ces variables impactent directement le temps de parcours. Toutefois, la perception du temps varie d'une personne à l'autre (Bonnel 2001) et les divers composants d'un trajet de transport en commun sont perçus de manière différente (Hunt 1990).

Il est effectivement commun d'associer des coefficients de pénibilité aux divers composants du trajet de transport en commun. L'équation (2.1) présente les indices de pénibilité proposés par Hunt (1990). Les variables présentées sont les suivantes :

- OWALK – temps d'accès à l'origine;
- NTRANS – nombre de transfert;
- RIDE – temps passé en véhicule;
- HDWAY – fréquence de passage;
- DWALK – temps d'accès à destination.

Ces indices sont utilisés dans le calcul de la fonction d'utilité de ce mode pour l'ensemble des déplacements *i*.

$$(2.1) \quad U(i) = -0.00634 * OWALK(i) - 2.83 * NTRANS(i) - 0.1580 * RIDE(i) \\ -0.600 * HDWAY(i) - 0.00335 * DWALK(i)$$

On y voit que le temps d'accès à l'origine est perçu comme deux fois plus pénible que le temps d'accès à la destination et on y voit également une grande pénibilité au niveau de la présence de transfert. Ainsi, la desserte devrait être plus dense où il y a le plus d'origines de déplacement et les gens seraient même prêts à allonger leur parcours de 15 minutes pour éviter un transfert (Hunt 1990). Effectivement, la présence de transferts est reconnue pour être particulièrement pénible et augmente le temps de déplacement perçu (Kaufmann 2007 ; Texas A&M 2013; Eboli et Mazzula 2011).

### 2.3.3.5 Coûts et perception des coûts

Lorsqu'une personne considère le prix relié à l'usage d'un mode, il évalue ce qu'il devra payer au moment de l'utilisation du mode, et non le coût réel engendré par son utilisation (Bonnel 2001) ni le coût subi par la société, tels que les coûts engendrés par les accidents de la route ou les polluants émis. Le conducteur moyen considérera donc uniquement les coûts directs, soient l'essence et les postes de péage éventuels. Les coûts d'achat, d'assurance, d'immatriculation et autres ne seront donc pas considérés. De manière similaire, le détenteur d'un titre mensuel de TC ne percevra aucun coût monétaire direct à l'utilisation de ce mode, ce qui favorisera grandement son utilisation (De Witte et al. 2014; Kim et Ulfarsson 2008).

### 2.3.3.6 Heure de départ

L'heure de départ est étudiée puisqu'elle correspond à un niveau de demande à un moment de la journée puisque de manière typique, la journée est marquée d'une pointe de demande le matin et le soir correspondant au déplacement des travailleurs. Le milieu de la journée est caractérisé par une plus faible demande de transport. On y trouve donc une plus faible densité d'offre TC et moins de congestion routière. Ces deux phénomènes ont comme effet de diminuer l'attractivité du TC durant cette période (Habib, Day et Miller 2009). Toutefois, l'heure de départ n'est trouvée significative que moins d'une fois sur deux (De Witte et al. 2014).

### 2.3.3.7 Chaînage

Valiquette (2010) définit la chaîne de déplacements comme étant « une série de déplacements interreliés effectués par les individus dans le but de consolider leurs activités. » La notion de chaînage devient d'autant plus critique dans l'étude du choix modal lors de l'utilisation d'un mode avec ancrage. Un mode avec ancrage est dit d'un mode que l'on stationne à destination et que l'on doit reprendre pour le ramener à la maison tel que l'automobile ou le vélo. À ce moment, la personne devient dépendante du moyen initialement choisi (Sicotte 2014). Le chaînage est, avec le niveau de motorisation, la variable qui est la plus souvent trouvée significative dans les articles recensés par De Witte et al. (2014), mais également une des variables les moins souvent étudiées.

### 2.3.4 Autres variables

D'autres familles de variables sont trouvées dans la littérature. Souvent des concepts plus qualitatifs, tels que le confort, l'esthétique et l'efficacité, sont plus difficiles à mesurer mais tout de même pertinents (Ampt, Swanson et Pearmain 1995). Outre les trois nommées précédemment, on retrouve fréquemment la sécurité, la fiabilité, les conditions météorologiques, l'accessibilité universelle, l'information au voyageur et la facilité d'utilisation (Bonnel 2001; Litman 2013; Leving et McGehee 2008). L'ensemble des variables présentées a donc une influence dans le processus de choix modal.

La prédiction du choix modal passe par la création de modèles prédictifs basés sur des données observées. Selon le niveau de précision désiré et/ou encore en fonction des données disponibles, plusieurs familles de modèles peuvent être utilisées.

Cette section présente les deux approches principales, soit l'approche agrégée et l'approche désagrégée, ainsi qu'au moins un type d'application pour chacune.

### 2.3.5 Modèles logit ou de régressions logistiques

Les modèles de choix discrets utilisés pour la modélisation du choix modal sont les modèles de régressions logistiques, dits modèles logit binaire dans le cas d'un choix binaire ou multinomial lorsqu'il y a plusieurs alternatives. Ils sont généralement utilisés puisqu'ils modélisent la probabilité qu'un événement se produise parmi un bassin d'alternatives. Dans le cas qui nous intéresse, l'événement correspond au choix d'un mode.

Ces modèles sont basés sur la notion de *fonction d'utilité*. Cette fonction regroupe un sous-ensemble des variables susceptibles d'affecter le choix entre une ou plusieurs alternatives et est définie sous la forme de l'équation (2.2) où  $V_{ik}$  représente l'utilité du mode  $i$  pour l'individu  $k$ ,  $X_{ik}$  correspond aux différentes variables explicatives du choix modal,  $\beta_{ik}$  le coefficient relatif associé à chacune de ces variables et  $\varepsilon_{ik}$  l'erreur (Bonnel 2001).

$$(2.2) \quad V_{ik} + \varepsilon_{ik} = \sum_k \beta_k X_{ik} + \varepsilon_{ik}$$

L'équation (2.3) présente l'expression générale des modèles logit multinomiaux. Ces modèles donnent la probabilité de choisir une alternative parmi un ensemble fini d'alternatives où  $Pr(i)$  est la probabilité de choisir l'alternative  $i$ ,  $V_i$  la fonction d'utilité de l'alternative  $i$  et  $V_j$  la fonction d'utilité de l'ensemble des alternatives.

$$(2.3) \quad Pr(i) = \frac{\exp(V_i)}{\sum_{j=1}^J \exp(V_j)}$$

Le calage de données empiriques de choix modal permet de calculer les coefficients  $\beta_k$  de l'équation 2.2 afin de prédire les modes empruntés pour un nouveau jeu de données d'intérêt. Les courbes de partage modal obtenues ont toujours une forme en sigmoïde avec l'équation (2.3), c'est-à-dire que, comme l'illustre la Figure 2-9 où  $V_A$  est l'utilité de l'alternative A et  $V_B$  celle de l'alternative B, l'impact de la différence d'utilité est amoindri pour les valeurs extrêmes jusqu'à tendre vers zéro.

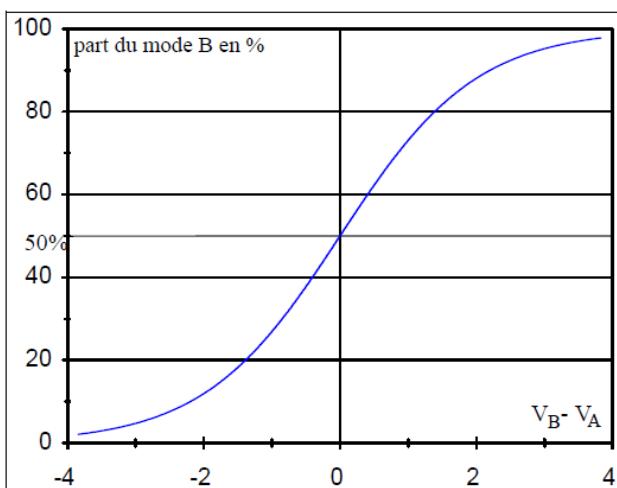


Figure 2-9 : Forme générique des courbes de partage modal (Bonnel 2001)

« Cette structure de modèle nécessitera toutefois le respect de l'hypothèse d'indépendance entre les alternatives disponibles » (Bonnel 2001) sinon la situation paradoxale du bus rouge et du bus bleu risque de survenir. Ce paradoxe est souvent utilisé dans la littérature pour représenter ce problème. Imaginons une personne qui doit choisir entre deux modes, l'auto et le bus, dont les utilités sont identiques. Cette personne effectuant un choix rationnel en comparant les utilités aura 50% de probabilité de choisir l'auto et tout autant de choisir le bus. Supposons maintenant qu'on décide de peindre la moitié de la flotte de bus en bleu et l'autre moitié en rouge. Il y a maintenant trois options ayant des utilités identiques. La personne aura maintenant une probabilité de 33% de choisir l'automobile, 33% de choisir le bus bleu et 33% de choisir le bus rouge; la probabilité de choisir le bus, sans porter attention à la couleur, est devenue 67%.

Cette situation illustre que l'utilisation d'alternatives non-indépendantes aura un impact sur la probabilité de choisir les autres modes. Nommée IIA (Independence of Irrelevant Alternatives), en français l'indépendance vis-à-vis des autres alternatives (Bonnel 2001), cette propriété est vitale à la précision des modèles.

### 2.3.6 Arbre de décision

L'arbre de décision est un modèle de classification dont la popularité est grandement due à sa simplicité. L'objectif est de produire des groupes d'individus les plus homogènes possibles du point de vue de la variable à prédire (Rakotomalala, 2005). Différentes méthodes de classification peuvent être utilisées. Le choix de la méthode dépend du type de variable à prédire : un arbre de décision pour les variables qualitatives ou un arbre de régression pour les variables quantitatives (Gonzalez, 2010).

L'arbre représente graphiquement les critères de classification identifiés comme étant les plus significatifs. Dans le cas des variables qualitatives, tel que le choix modal, le niveau de pertinence des variables choisies peut être défini par l'indice de Gini, tel que dans le cas présent.

La méthode CART, proposé par Breiman, Friedman, Ohlsen et Stone (1984), permet de construire un arbre de décision par division binaire à chaque nœud. Appropriée pour l'analyse de variables catégorielles, cette méthode n'impose pas de règle d'arrêt et fournit l'arbre le plus complet possible, tendant vers des feuilles pures.

L’arbre de décision est composé de branches et de feuilles. À partir du sommet, les branches mènent aux feuilles, qui sont le niveau terminal de chacune des branches. Le niveau de pureté d’une feuille représente la mixité du choix modal issu des déplacements classifiés dans cette branche. Une feuille où 100% des déplacements auraient été faits par le même mode serait une feuille pure. Chaque nouvelle branche tend à augmenter la pureté des feuilles.

Le taux d’erreur associé à un arbre de décision est défini comme la proportion d’individus appartenant à une classe donnée ayant été affectés à une classe autre.

Cette méthode comporte plusieurs avantages mais aussi des limitations majeures. Il peut aisément être représenté graphiquement, ce qui aide à sa compréhension et il peut être utilisé comme un puissant outil de communication. C’est une méthode flexible qui accepte une mixité de variables; l’arbre peut classer un mélange de variables nominales et continues. C’est également un algorithme très robuste vis-à-vis des valeurs extrêmes. Toutefois, l’aspect séquentiel est sa majeure limitation. Bien que rapide, le modèle d’investigue qu’un nombre restreint de possibilités à chaque étape, sans remise en cause des choix précédents (Decaestecker et Saerens, 2017). Aussi, cette méthode nécessite des échantillons de grandes tailles et la sélection des variables présente un léger biais quant à la multitude d’options puisque, comme le mentionne (Gonzalez 2010), les variables avec le plus grand nombre d’options ont tendance à être plus souvent sélectionnées.

## 2.4 Méthodes d’analyse spatiale

L’analyse spatiale des données géoréférencées est devenue incontournable dans de nombreux domaines et la principale méthode utilisée est l’évaluation de l’autocorrélation spatiale (Sawada, 2009). Cette section présente la définition de l’autocorrélation spatiale, la notion de continuité ainsi que divers indicateurs d’autocorrélation spatiale.

### 2.4.1 Définition

Plusieurs définitions de l’autocorrélation spatiale sont présentes dans la littérature. Sawada (2009) en présente quelques-unes.

La première, proposée par Upton et Fingleton (1985) tel que rapporté par Sawada, définit l’autocorrélation spatiale comme une propriété des données géoréférencées qui se manifeste par la présence de motifs distincts dans les données. Ces motifs peuvent être au niveau de la localisation

géographique des données et/ou dans les valeurs des données elles-mêmes. Lorsque des valeurs élevées ont un voisinage de valeurs élevées, l'indice est positif, tandis que dans le cas où les valeurs alternent entre élevées et faibles, la corrélation est négative.

La définition de Sokal et Oden (1978a) dit que les tests d'autocorrélation spatiale évaluent si la valeur observée des variables est localement indépendante des valeurs de ces mêmes variables dans son voisinage (« Spatial autocorrelation quantifies the dependence of the values of a geographically distributed variable on values of the same variable at other localities »). Une autre définition citée est celle de Goodchild (1987) qui dit que, de manière globale, l'autocorrélation spatiale reflète la première loi de géographie de Tobler selon laquelle « toute chose est liée à toute autre, mais que les choses les plus proches le sont encore plus. »

#### 2.4.2 Notions de voisinage

La notion de voisinage est primordiale à l'analyse d'autocorrélation spatiale. Afin de définir les zones voisines d'une zone d'analyse, trois critères de contiguïté ont été développés. Parmi ceux-ci, le critère du fou (bishop's case) est très peu utilisé et c'est généralement le critère de la tour (rook's case) et celui de la reine (queen's case) qui sont utilisés. Ces critères sont illustrés à la Figure 2-10.

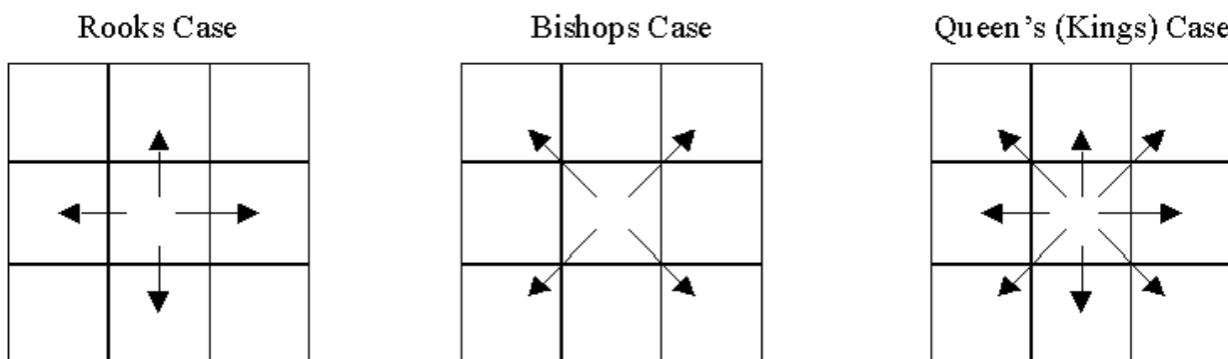


Figure 2-10 : Types de contiguïté (Sawada, 2009)

Selon le critère de la reine, deux polygones sont voisins si au moins un point d'une frontière d'un polygone coïncide avec au moins un point d'une frontière d'un autre polygone. Selon le critère de la tour, seuls les polygones partageant plus d'un point situé sur les faces sont considérés comme voisins.

Le critère utilisé pour les analyses spatiales contenues dans ce mémoire est le « critère de la reine ». Toutefois, dans ce cas précis, c'est-à-dire avec un quadrillage fait de zones hexagonales, les résultats des critères de la reine et de la tour sont identiques.

### 2.4.3 Indicateurs locaux d'autocorrélation spatiale

L'arrivée des technologies numériques dont les systèmes d'information géographique (SIG ou GIS en anglais) a repoussé les limites de l'analyse spatiale en permettant la création de nouvelles méthodes exploratoires d'analyses spatiales. Sur la base de l'indicateur I de Moran (Anselin 1950), Anselin (1995) a développé un indicateur local d'autocorrélation spatiale, communément référencé par l'acronyme LISA, pour *Local Indicators of Spatial Association*. Il le définit en deux points :

1. Le LISA de toute observation donne une indication de l'importance du regroupement spatial localisé autour de cette observation.
2. La somme des LISA de l'ensemble des observations est proportionnelle à l'indicateur global d'autocorrélation.

Le LISA est en fait la décomposition locale du Moran's I et s'exprime de manière générale sous la forme de l'équation 2.7 où  $I_i$  est le Moran's I local pour une caractéristique  $i$ ,  $x_i$  est la valeur de la caractéristique  $i$ ,  $\bar{X}$  la moyenne de ces valeurs,  $w_{i,j}$  les attributs de voisinage entre les secteurs  $w_i$  et  $w_j$ .  $S_i^2$  est défini par l'équation (2.8) dans laquelle  $n$  est le nombre total de zones d'analyse (Desktop.ArcGIS.com – Anselin's Local Moran's I).

$$(2.7) \quad I_i = \frac{x_i - \bar{X}}{S_i^2} \sum_{j=1, j \neq i}^n w_{i,j} (x_j - \bar{X})$$

$$(2.8) \quad S_i^2 = \frac{\sum_{j=1, j \neq i}^n (x_j - \bar{X})^2}{n - 1}$$

L'hypothèse nulle associée à l'utilisation du LISA est définie comme l'absence d'autocorrélation spatiale. Les seuils de significativité peuvent être définis selon les besoins, généralement à  $p < 0.05$ , pouvant aller de  $p < 0.1$  à  $p < 0.001$ .

L'identification de groupes ou de clusters significativement différents de la moyenne passe par l'étude du seuil de significativité  $p$  (p-value), tel qu'illustré sur la Figure 2-11. Les p-values permettant d'identifier les seuils de significativité peuvent être représentés dans un plan cartésien

de quatre quadrants identifiant les types d'autocorrélation spatiales observés (voir Figure 2-12). Les quadrants 1 et 3 réfèrent aux situations d'autocorrélation positive, où des zones ayant des valeurs élevées (respectivement faibles) sont à proximité d'autres zones ayant aussi des valeurs élevées (respectivement faibles). Les quadrants 2 et 4 réfèrent quant à eux plutôt à la présence d'autocorrélation négative i.e. les zones de valeurs élevées cernées de valeurs faibles et vice-versa

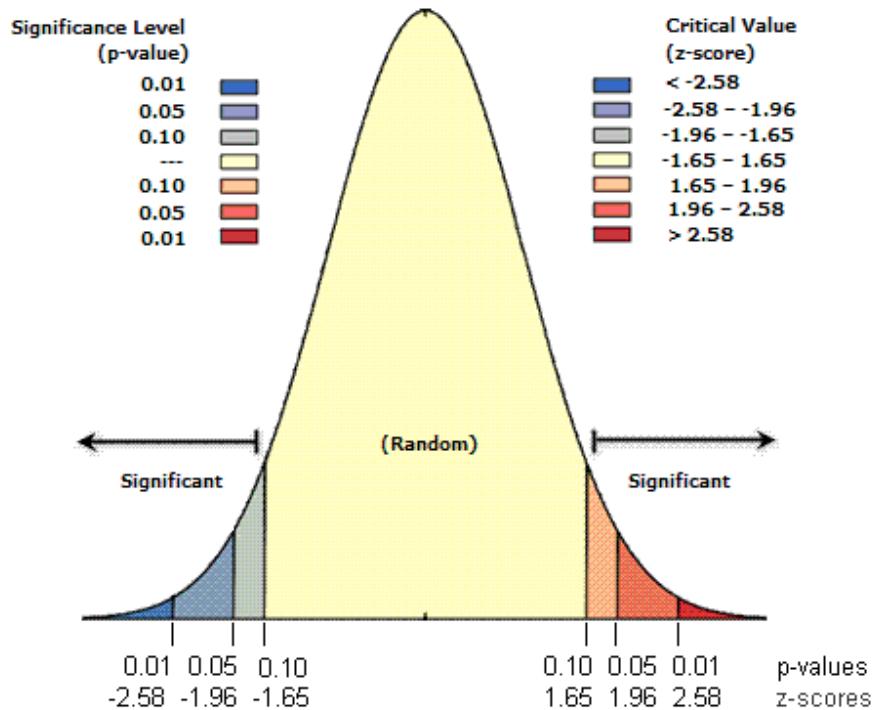


Figure 2-11 : Valeur p et valeur seuil z (Desktop.ArcGIS.com – Spatial Statistics Toolbox)



Figure 2-12 : Type des zones des clusters de LISA

La Figure 2-13 résume les étapes du processus d'identification des clusters de LISA telles qu'effectuées avec un logiciel SIG moderne tel que ArcGIS ou GeoDa. À partir de données zonales géoréférencées, les indicateurs locaux de Moran sont calculés. Les écarts-types et les niveaux de significativité sont calculés et finalement, les zones sont identifiées en couleur.

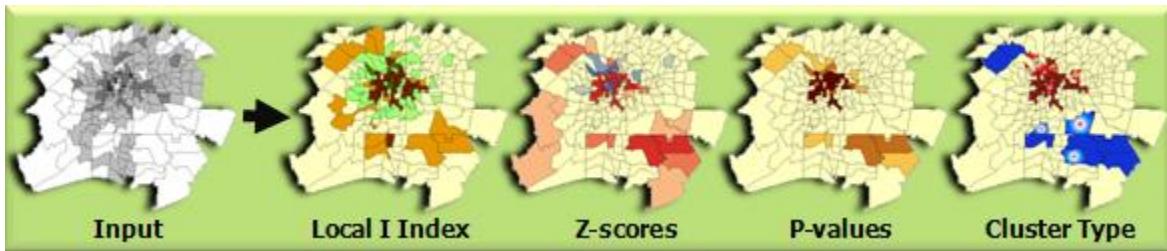


Figure 2-13 : Processus d'identification des clusters de LISA (Desktop.ArcGIS.com – Anselin's Local Moran's I)

Notons que, tel que l'affirme Anselin (1995), cette analyse spatiale n'explique rien en elle-même mais peut servir, lors de démarches exploratoires, à analyser et identifier des motifs dans des jeux de données géoréférencées.

## 2.5 Méthodes de classification des déplacements

L'analyse des déplacements contenus dans une base de données d'une taille telle que celle de l'enquête OD se prête bien à l'utilisation de méthodes de regroupement (clustering) d'observations similaires. À cette fin, une méthode de clustering en deux étapes est proposée. En premier lieu, la méthode des k-moyennes est utilisée pour créer cinquante sous-ensembles de déplacements similaires. Ensuite, les sous-ensembles sont regroupés par similarité, jusqu'à avoir un arbre complet. Cette méthode a pour avantage d'alléger la base de données à regrouper pour pouvoir appliquer la méthode de clustering hiérarchique.

### 2.5.1 Classification par k-moyennes

La méthode k-moyennes vise à répartir  $n$  observations décrites par des vecteurs de dimension  $d$  en  $k$  sous-ensembles de données  $S_k$  de manière à réduire la somme du carré des distances euclidiennes dans chaque sous-ensemble. En définissant un espace normé de  $d$  dimensions correspondant à chacune des variables à analyser, il est possible de calculer la distance entre les différents points. Dans ce mémoire, la méthode des k-moyennes a été utilisée avec la distance euclidienne, soit la généralisation du théorème de Pythagore qui définit la distance entre deux points comme étant la

racine de la somme des carrés des composants dimensionnels. La méthode des k-moyennes cherche donc à créer les ensembles dans lesquels les points seraient les plus rapprochés possibles.

L'équation (4.1) présente l'expression que le processus de regroupement cherche à réduire où k est le nombre de groupes,  $x_j$  est la position de chacun des points,  $\bar{x}_i$  la position du centroïde,  $S_i$  chacun des sous-ensembles et  $\bar{x}$  la position du centroïde du sous-ensemble S.

$$(4.1) \quad \sum_{i=1}^k \sum_{x_j \in S_i} \|x_j - \bar{x}_i\|^2$$

Cette méthode n'identifiant pas forcément le minimum global, il ne serait pas judicieux de l'utiliser à cette fin sauf si l'on ne désire que deux clusters pour un petit jeu de données. Plutôt, cette méthode identifie des minimums locaux jusqu'à convergence, c'est-à-dire jusqu'à ce que la permutation d'un point d'un cluster à un autre ne diminue plus la somme des distances quadratiques des clusters (Hartigan 1979).

Le processus est itératif et comprend deux étapes. Initialement, « k » centroïdes sont définis aléatoirement au travers de la plage de données. La **Erreur ! Source du renvoi introuvable.** illustre deux itérations du processus décrit dans le prochain paragraphe.

La première étape est l'assignation. À cette étape, chaque point est assigné au centroïde le plus près formant des sous-ensembles de points. La seconde étape consiste au calcul des points moyens. Durant cette étape, la position des centroïdes est redéfinie au centre de ces sous-ensembles, c'est-à-dire à la position qui minimise la somme des carrés. Suite à cette étape, une nouvelle assignation est effectuée. Le processus continue ainsi jusqu'à l'atteinte du maximum d'itération fixé préalablement, ou jusqu'à la convergence des résultats, c'est-à-dire qu'aucune réassignation de données dans un autre groupe ne diminuerait la somme du carré des distances. Dans cette étude, ce nombre a été fixé à 2000.

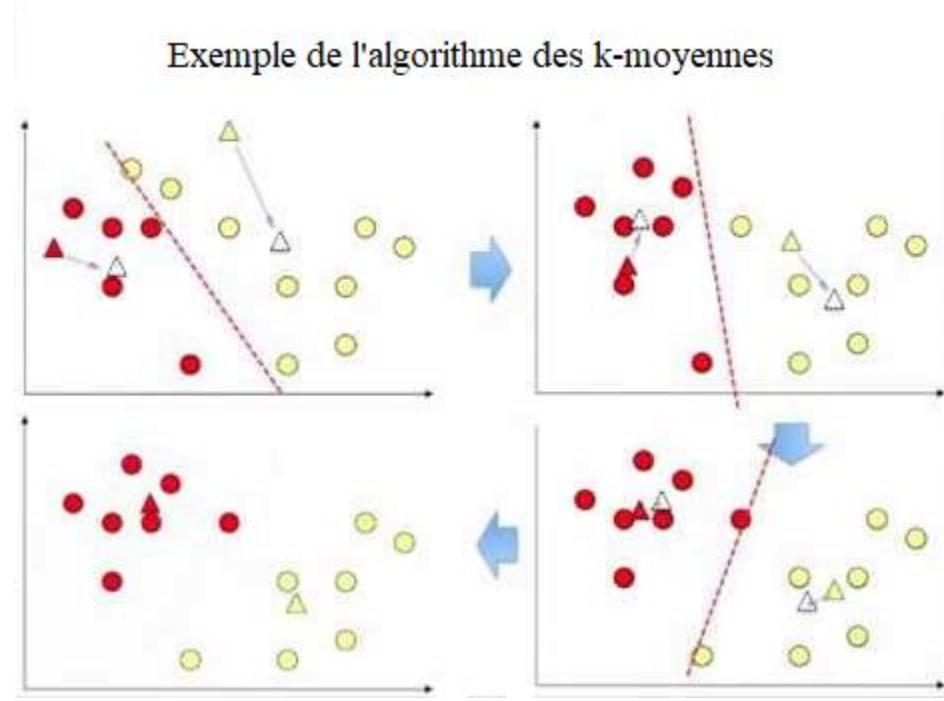


Figure 2-14 : Exemple de l'algorithme des k-moyennes

La méthode de clustering K-moyennes peut être utilisée pour de grands jeux de données mais le coût de calcul ( $O$ ) s'amplifie de manière exponentielle de la façon suivante :  $O = i * k * n * d$  où  $i$  est le nombre d'itérations,  $k$  le nombre de groupes,  $n$  la taille de l'ensemble de données et  $d$  le nombre de dimensions.

Il est important de savoir que la méthode k-moyennes ne supporte que des variables numériques. L'algorithme calculant des distances euclidiennes et cherchant à minimiser les différences quadratiques, il est impossible de calculer directement des distances entre les variables catégorielles. Pour les traiter, des variables binaires (« dummy ») doivent être créées afin de représenter numériquement l'ensemble des états que peuvent emprunter les variables catégorielles. Également, cette méthode est très sensible aux données extrêmes puisque leurs positions euclidiennes sont généralement très éloignées.

## 2.5.2 Clustering hiérarchique

Dans cette méthode à deux étapes, la méthode de clustering hiérarchique cherche à regrouper les sous-ensembles créés à l'étape précédente. Cette méthode regroupe les données qui sont les plus similaires jusqu'à ce que l'ensemble des données aient été regroupées, regroupement ayant la forme

d'un arbre lorsque représenté graphiquement. Cette méthode nécessite le calcul d'une matrice de distance entre tous les points. Le coût des calculs de cette méthode est  $O = (n^2 \log(n))$ , où  $n$  est la taille de l'ensemble de données à trier. Pour un grand jeu de données, le calcul de la matrice de distance rend cette méthode inutilisable.

Pour réaliser cette méthode, il peut être important de débuter en normalisant les données. Dans cette étude, la cote Z a été utilisée, soit une normalisation par écart à la moyenne divisé par l'écart-type. Cette étape assure une pondération uniforme de l'ensemble des caractéristiques des vecteurs dans le processus de clustering. Ensuite vient le calcul de la matrice de proximité à partir de laquelle les clusters peuvent être calculés. La quantité appropriée de clusters à créer dépend toujours des données à trier elles-mêmes mais il existe une méthode permettant d'évaluer le niveau de similitude des données dans chaque groupe. Cette méthode calcule le carré des différences entre la position des membres du groupe et leur valeur moyenne. L'objectif étant de diminuer cette différence (équation 4.1), le graphique des sommes des erreurs quadratiques (WSS ou within-cluster sum-of-square en anglais) illustre l'évolution de la somme des erreurs et aide à choisir la quantité appropriée de groupes à analyser. L'idée est de sélectionner un nombre de clusters minimisant la somme des erreurs quadratiques et qui se trouve immédiatement suite à un point d'inflexion du graphique WSS. Sur l'exemple présenté à la **Erreur ! Source du renvoi introuvable.**, le point d'inflexion étant situé à 3 groupes, quatre clusters devraient être une bonne division. La somme des carrés résiduels ne diminue d'ailleurs presque plus dépassé ce seuil.

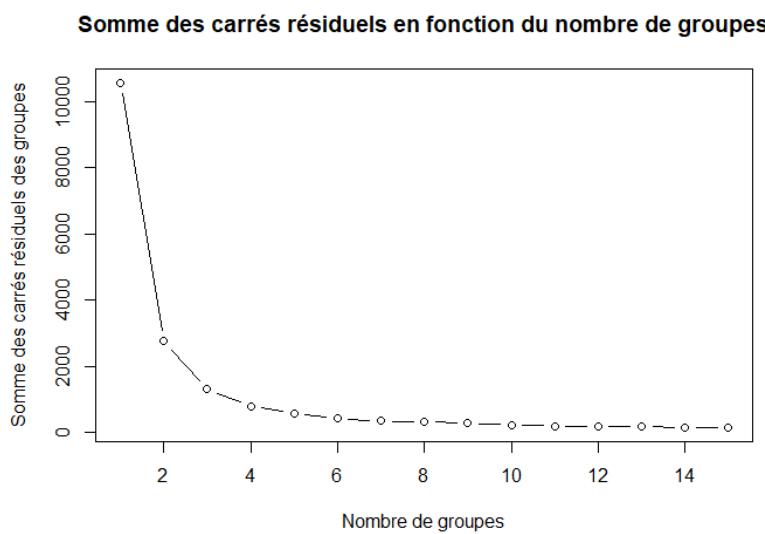


Figure 2-15 : Exemple de graphique WS

## 2.6 Synthèse générale

Ce chapitre a présenté une revue de littérature pertinente aux objectifs de recherche. Un vaste éventail de variables explicatives du choix modal présentes dans la littérature y a été présenté. Classés en trois catégories principales, soit les variables liées aux caractéristiques sociodémographiques, aux indicateurs spatiaux et aux caractéristiques du déplacement, on y voit que certaines de ces variables sont étudiées de manière plus ou moins exhaustive et que le choix des variables utilisées diffère d'une étude à l'autre. Les plus étudiées sont le temps de parcours, le coût, le niveau de motorisation, le genre, la densité de population et l'âge que l'on retrouve dans plus de 45% des articles étudiés par De Witte et al. De manière inverse, le chaînage, la présence de stationnement et les habitudes personnelles se sont retrouvées dans moins de 20% de ces articles bien qu'ils aient été significatifs dans la majorité des études où ils étaient présents.

Deux types de modèles ont été présentés, les modèles de choix discrets (logit) et les arbres de décisions. Basés sur une fonction d'utilité, les modèles logit calculent la probabilité qu'un évènement se produise en assignant à chacune des variables explicatives un coefficient d'importance relative. L'utilisation de ce modèle nécessite une sélection de variables indépendantes les unes des autres. L'arbre de décision peut, quand-à-lui, traiter l'ensemble des variables qu'elles soient indépendantes ou non, tant que celles-ci sont numériques. Advenant des variables catégorielles, des variables binaires (ou dummy) devront potentiellement être créés pour représenter numériquement les divers états des variables catégorielles.

La méthode d'analyse spatiale développée par Anselin a été présentée. Cette méthode consiste à calculer le coefficient Local de Moran (LISA) pour l'ensemble des zones d'analyse afin d'en identifier des regroupements. Le logiciel GeoDa est particulièrement utile pour ce genre d'analyse.

Finalement, une méthode de regroupement en deux étapes, soit regroupement par la méthode des k-moyennes suivi d'un regroupement hiérarchique, est présenté. Ces regroupements sont effectués à l'aide de la bibliothèque R du logiciel RStudio.

Pour les fins d'analyse et de modélisation de ce mémoire, certaines variables doivent être sélectionnées. Ce choix sera fait à la lumière des informations présentées dans ce chapitre, des données disponibles dans l'enquête OD ainsi qu'à la lumière des objectifs poursuivis par le modèle de choix modal.

## CHAPITRE 3 MÉTHODOLOGIE GÉNÉRALE

Ce chapitre présente la méthodologie développée pour l’analyse de la compétitivité du transport en commun et des impacts de scénarios projetés. Il présente le cadre méthodologique, les données utilisées, précise les étapes des manipulations de données et met en relation les divers concepts traités dans ce mémoire.

### 3.1 Structure générale de la méthode

La méthodologie menant aux divers résultats et modèles est résumée à la Figure 3-1. À partir des données disponibles dans l’enquête OD 2008, des simulations de temps de parcours ont été effectuées par la Chaire Mobilité (CM) sur la plateforme Open Source Routing Machine (OSRM) et par le MMTDET grâce au MOTREM-08 (MO). À partir de ces temps, des indicateurs de compétitivité modale nommés « ratios de compétitivité » ont été développés. Détaillés en section 3.2, ces ratios servent à l’analyse spatiale de la performance intermodale ainsi qu’à identifier des familles de déplacements à l’aide de méthodes de clustering.

Les données simulées à la CM sont comparées à celles issues du MO dans une fin de validation des simulations. Ensuite, les temps simulés de la CM sont utilisés pour caler des modèles de choix modal qui sont, à leur tour, pourront être utilisés pour la projection des scénarios développés dans le cadre du projet de recherche global.

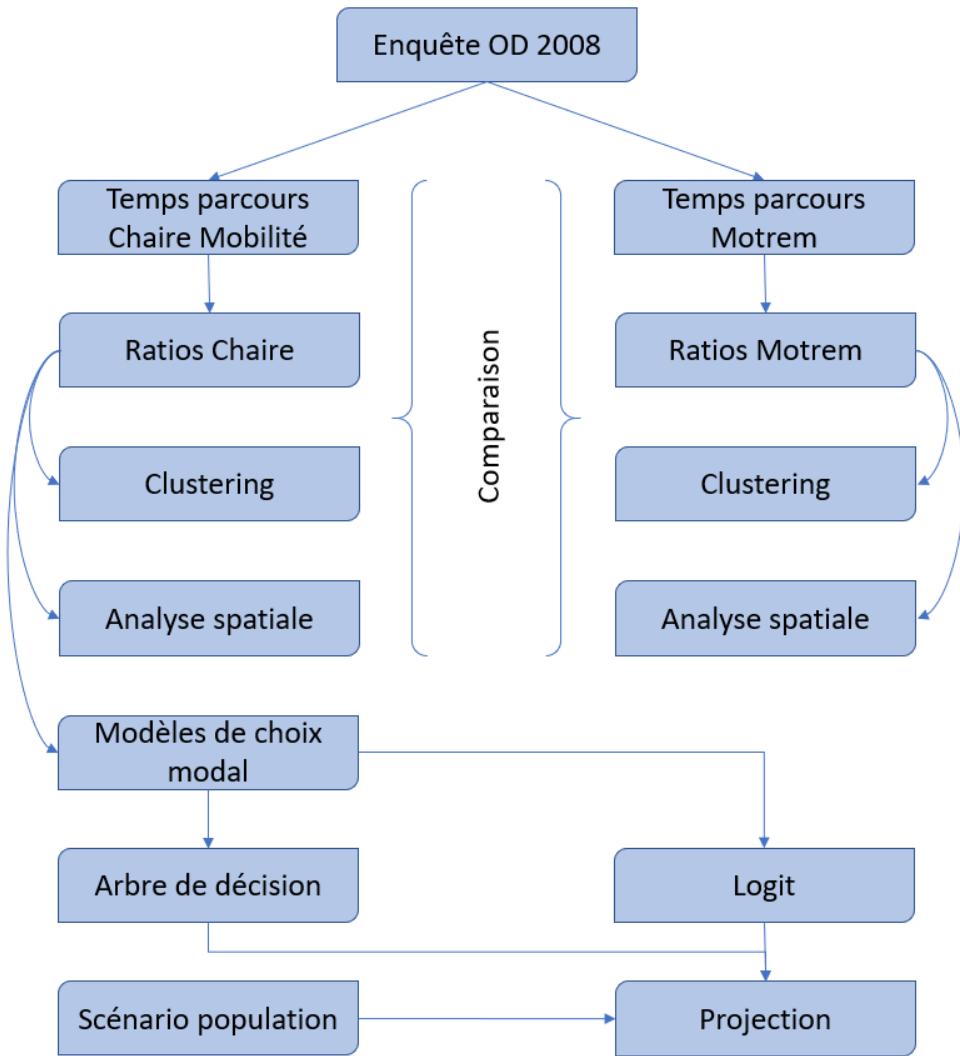


Figure 3-1 : Schéma méthodologique global

### 3.2 Définition/Justification des ratios de compétitivité

L'objectif de l'étude est de pouvoir comparer aisément la performance relative de chacun des modes par rapport à l'automobile. L'idée expérimentée dans ce mémoire est de créer un indicateur basé principalement sur les caractéristiques de l'offre de transport et sur celles intrinsèques aux déplacements.

Les données GTFS contiennent l'ensemble de la définition du réseau de TC et ses caractéristiques de service (les lignes desservies, leur fréquence de passage, etc.). Les caractéristiques du déplacement sont la paire OD, l'heure de départ et le motif du déplacement.

Il a été mentionné précédemment que les voyageurs se basent sur les coûts perçus plutôt que sur les coûts réels lorsqu'ils choisissent un mode de transport pour effectuer leur déplacement. En termes de coût monétaire, les détenteurs d'une voiture voient à coût nul son utilisation pour un déplacement en particulier, tout comme un titulaire d'abonnement mensuel au service de TC voit à coût nul l'utilisation du TC. Outre le coût monétaire, la durée du trajet intervient également au niveau du coût généralisé. Dans ce cas, avec un coût monétaire perçu comme étant nul, l'hypothèse suivante est posée : le coût généralisé des déplacements est uniquement dépendant du temps de parcours. Ces temps incluent le détail de l'ensemble des sections du parcours, débutant par le temps d'accès à l'origine, le temps d'attente initial, le temps passé en véhicule, le temps d'attente aux correspondances, si applicable, ainsi que le temps d'accès à la destination.

Considérant qu'un des objectifs de ce mémoire est de développer une méthode qui permet d'isoler les impacts de l'offre de transport, la méthode développée doit réduire l'impact des propriétés individuelles telles que la perception du temps ou les préférences modales. Bien que simplificatrice, cette méthode constitue la base d'un modèle d'analyse des scénarios projetés. Dans une éventuelle amélioration de celui-ci, différentes caractéristiques de la personne pourraient être ajoutées et des facteurs de pénibilité propres à différents segments de population pourraient aussi être définis, voir entre autres Nazem et al. (2011).

Se voulant un outil de comparaison modale, l'indicateur suggéré correspond, pour un déplacement défini par une paire OD et une heure de départ, au ratio entre le temps de parcours du mode choisi divisé par le temps de parcours du même déplacement fait en automobile sous congestion. Cet indicateur est nommé ratio de compétitivité. Un des intérêts principaux de ces ratios est de pouvoir mesurer rapidement l'impact d'un changement de temps de parcours hypothétique plutôt que de devoir définir les modifications de l'offre dans les logiciels de simulation et refaire des simulations.

Il sera possible de caractériser chacun des déplacements, aussi référencé sous le terme ipère qui est le numéro séquentiel unique associé à chacun des enregistrements, par ses trois ratios de compétitivité : le ratio TC, le ratio vélo et le ratio marche (les trois ayant comme dénominateur le temps de déplacement en automobile). L'équation 3.1 présente l'équation générale du ratio de compétitivité, ratio de temps de parcours du mode  $i$  par rapport à celui du mode  $j$  et le Tableau 3-1 illustre la forme vectorielle d'un ipère défini par ses ratios de compétitivité de chacun des modes par rapport à l'automobile. Ainsi, sous forme vectorielle, un trajet de 10 minutes en voiture, 15

minutes en TC, 20 minutes à vélo et de 45 minutes à pied sera défini par le vecteur de ratios suivant : 1,5; 2; 4,5.

$$(3.1) \quad Ratio_{i/j} = TempsParcours_i / TempsParcours_j$$

Tableau 3-1 : Structure du vecteur de ratios de compétitivité

Ipère	Ratio TC/Auto	Ratio Vélo/Auto	Ratio Marche/Auto
-------	---------------	-----------------	-------------------

En caractérisant chacun des déplacements par un vecteur de ratios caractérisant les alternatives modales, il pourrait être possible de trouver une corrélation entre le choix modal observé dans l'enquête OD et des familles de ratios identifiées par une méthode de classification. Ces familles pourraient aussi servir à identifier des zones de forte ou de faible compétitivité modale.

Dans le cadre de ce mémoire, chaque paire OD est simulée à l'heure de départ déclarée. Dans une optique de caractérisation de l'offre TC, il serait intéressant d'effectuer les simulations de chaque paire OD à chaque heure de départ afin de voir l'évolution temporelle de la qualité de l'offre TC par rapport à l'automobile. Il serait également possible de modifier la définition des vecteurs du Tableau 3-1 pour y inclure d'autres variables du déplacement, comme la distance de déplacement.

### 3.3 Sources et description des données

Deux sources de données sont à disposition pour développer cette méthode. Toutes deux simulées à partir des données issues de l'enquête OD 2008, la première provient de la Chaire Mobilité et la seconde du MTMDET. Cette section présente ces deux bases de données, leur source et les données qu'elles contiennent.

#### 3.3.1 Enquête OD

Un consortium d'autorités de transport de la région de Montréal assure la réalisation d'une enquête, tous les cinq ans, qui sonde environ 5% de la population du Grand-Montréal afin de dresser un portrait des habitudes de transport. L'enquête de 2008 utilisée pour ce projet contient 355 459 observations pondérées pour recréer les quelque 8 000 000 de déplacements quotidiens faits par les résidents. Notons toutefois que parmi les 355 459 observations, seules 319 437 représentent des

déplacements, les autres répondants ayant affirmé ne pas s'être déplacés. La base de données contenant 135 variables, contenant de l'information sur le logis, sur la personne et sur le déplacement lui-même notamment les modes empruntés; l'enquête est une riche source d'informations. La plus récente enquête OD a été effectuée en 2013. Toutefois, les prévisions basées 2013 n'étant pas encore disponibles celles de 2008 ont été utilisées, incluant un facteur de projection des déplacements à l'horizon 2031. Plus d'information au sujet de ces enquêtes peut être trouvé sur le site du RTM à l'adresse web suivante : <https://rtm.quebec/fr/a-propos/portrait-mobilite>.

### **3.3.1.1 Zone de couverture de l'enquête OD**

Afin d'encadrer les données pertinentes à ce travail, il est important de définir la zone d'intérêt. L'enquête OD s'intéresse à l'ensemble de la grande région de Montréal, soit l'île de Montréal, Laval, Longueuil ainsi que les couronnes nord et sud. Les ménages enquêtés sont représentés sous forme de carte de chaleur à la Figure 3-2. Par souci d'uniformité entre la taille des zones d'analyse, des partitions hexagonales d'un kilomètre carré ont été faites pour recouvrir l'ensemble de la région. Ce quadrillage servira à l'agrégation des données pour l'ensemble des analyses spatiales.

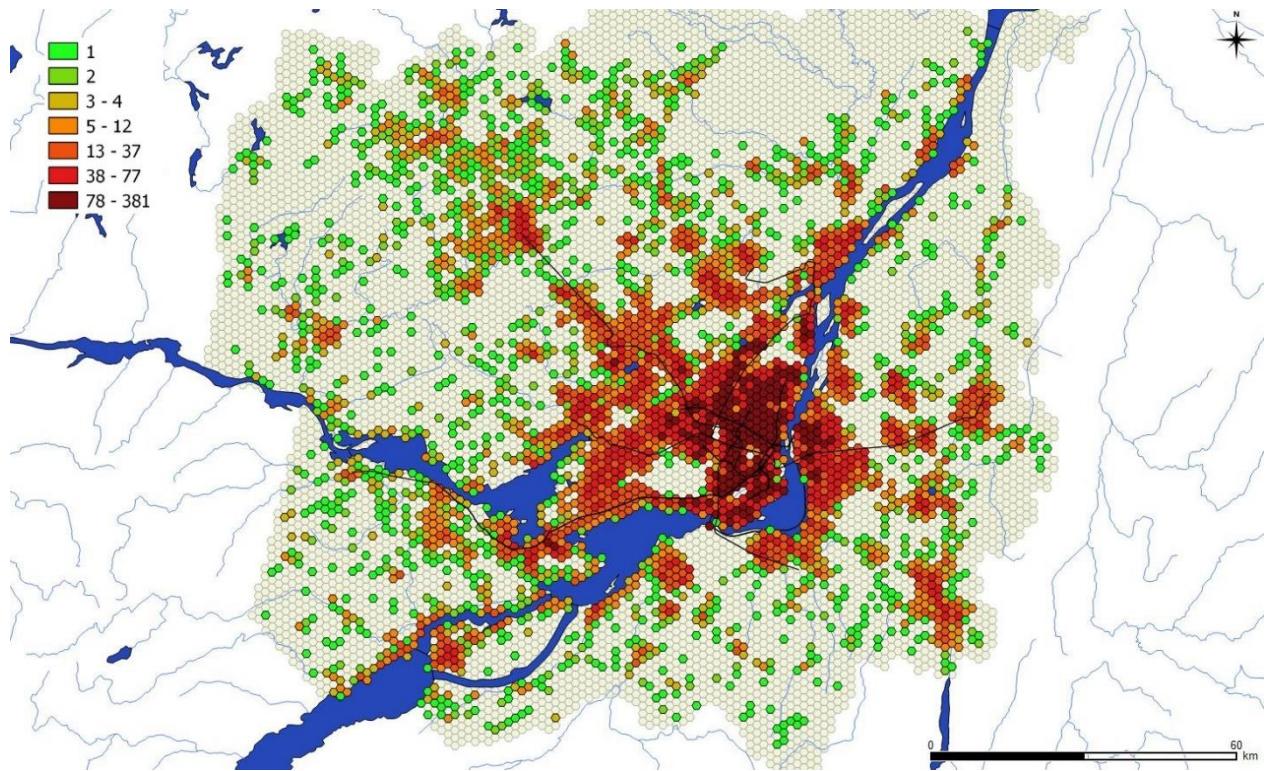


Figure 3-2 : Ménages enquêtés dans l'enquête OD 2008, par zone hexagonale de 1 km<sup>2</sup>

### 3.3.2 Simulations issues du modèle régional

Le MTMDET utilise différents outils pour simuler les déplacements sur les réseaux : MADITUC pour l'affectation des déplacements sur le réseau de transport en commun et EMME (MTMDET 2015) pour l'affectation des déplacements sur le réseau routier. L'ensemble des déplacements ayant été déclarés en mode motorisé dans l'enquête OD 2008 est simulé, soit un total de 262 653 déplacements. Le modèle prend en compte la congestion selon les trois étapes de la méthode suivante. Tout d'abord, les déplacements recueillis dans l'enquête OD forment une première catégorie de véhicule nommés véhicules particuliers. Ensuite, à partir de comptages, sont ajoutés trois autres classes de véhicules commerciaux : les légers commerciaux, les camions réguliers et les camions lourds. Finalement, ces déplacements sont distribués et affectés sur le réseau routier à l'aide du logiciel EMME qui effectue une affectation d'équilibre. Le niveau de congestion peut alors être évalué en termes de vitesse de déplacement.

Le logiciel MADITUC permet d'estimer le temps de déplacement sur le réseau TC. Il est à noter que, dans le cadre de ce projet, les déplacements bimodaux (impliquant un mode privé et un mode public, typiquement automobile et transport en commun) sont simulés tel que des déplacements unimodaux pour chacun des modes impliqués.

La Figure 3-3 présente les zones d'analyse transport (ZAT) utilisées dans ce dernier.

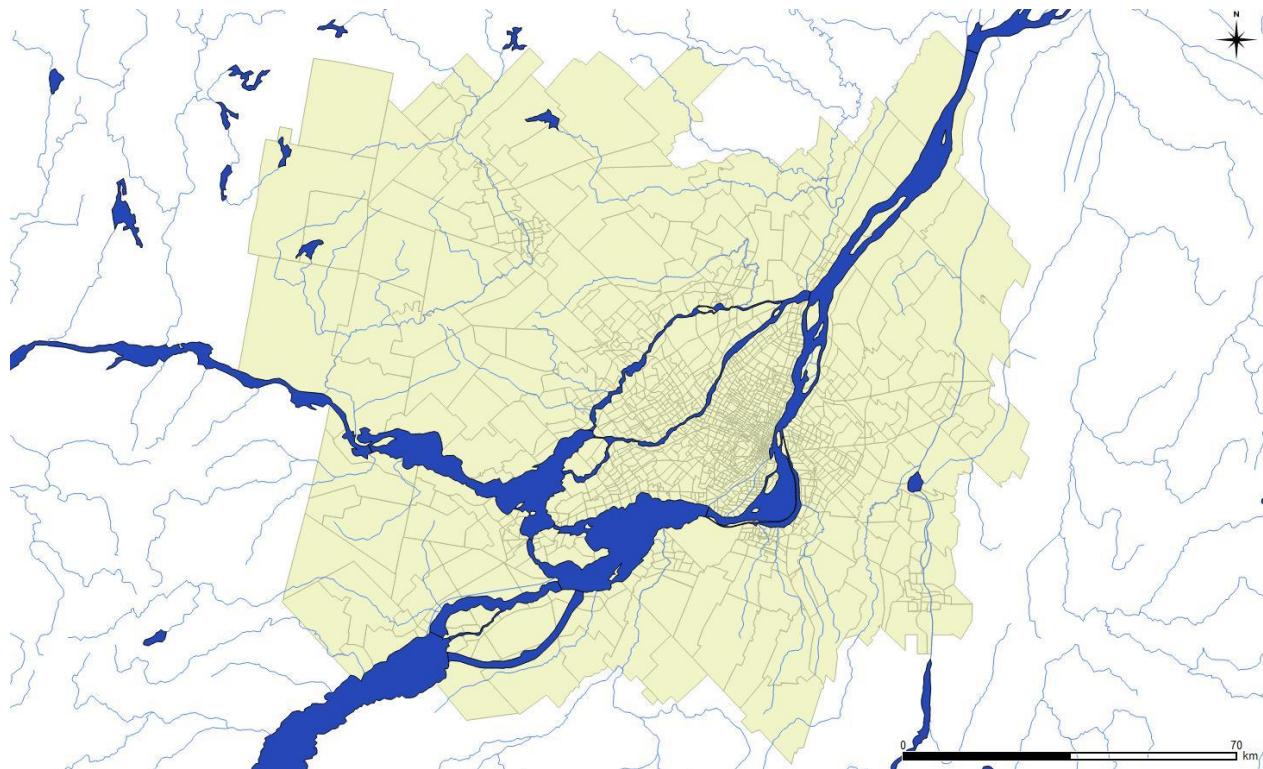


Figure 3-3 : Zones analyse transport (ZAT) utilisées dans le Motrem08

### 3.3.2.1 Simulations automobiles issues du MOTREM

Le modèle de transport de la région de Montréal de l'année 2008, nommé MOTREM08, fournit des débits et des vitesses moyennes par lien à « chaque heure d'une journée ouvrable typique de l'automne » (MTQ, 2014). Toutefois, ayant été calibrés sur la période de pointe AM, les résultats présentés pour les autres périodes horaires n'ont pas été validés.

Les informations routières proviennent des simulations effectuées avec le logiciel EMME en se basant sur la position des centroïdes de chacune des ZAT. La simulation calcule des temps de parcours moyens pour chaque paire de centroïdes zonaux pour chaque heure de la journée. Cette

méthode n'étant pas en mesure de simuler les déplacements internes, les temps de déplacements intrazonaux sont approximés selon la taille et la densité de population de la zone.

### 3.3.2.2 MOTREM TC

Les simulations de parcours en transport collectif ont été effectuées avec le système MADITUC du service de la modélisation des systèmes de transport (SMST), calibration AMT v3. Ce modèle de simulation se base sur des itinéraires simulés plutôt que déclarés : les résultats peuvent ne pas être significatifs, particulièrement pour les déplacements de courtes distances (MTQ, 2014). Aussi à noter, ce module de simulation segmente la journée en cinq périodes horaires (pointe am, hors-pointe jour, pointe pm, hors pointe soir et nuit) pour lesquelles des niveaux de service spécifiques sont codées. Les simulations ayant été calibrées sur la pointe am, la qualité des données d'autres périodes horaires n'a pas été validée et celles-ci ne sont donc pas utilisées.

Parmi les 262 653 déplacements contenus dans la base de données du MOTREM08, soit tous les déplacements ayant été effectués avec un mode motorisé, 241 147 ont trouvé un itinéraire en TC. Il y a donc 21 506 déplacements qui n'ont pas trouvé d'itinéraires en TC, soit 8,19% des déplacements.

### 3.3.3 Simulation Chaire Mobilité

En cohérence avec l'importance maintenant accordée au mode actif, il devient essentiel d'en tenir compte aussi dans les modèles de choix modaux. Pour ce faire, tous les déplacements sont simulés à l'heure déclarée pour chacun des modes d'intérêt, soit la voiture, le TC, le vélo et la marche. Cette section présente les diverses simulations effectuées à la Chaire Mobilité. Ces données étant la base de toutes les analyses et modèles, il est important de bien saisir leur nature.

#### 3.3.3.1 Temps auto

Les temps de parcours en voiture simulés sont basés sur les vitesses observées via une base de données GPS de flotte de taxis (voir Tessier 2014 et Morency et al. 2017 pour la présentation des données utilisées et des indicateurs qui peuvent être estimés avec celles-ci). Les simulations utilisent donc des vitesses observées sur les réseaux pour différentes heures de la journée, et ce afin d'assurer la meilleure représentativité possible des conditions routières typiques, c'est-à-dire sous

congestion. Lorsque les données sont manquantes, les véhicules simulés se déplacent à la vitesse affichée.

### 3.3.3.2 Temps TC

Les simulations de temps de parcours en TC effectuées à la Chaire Mobilité sont des simulations désagrégées qui simulent de manière indépendante le trajet le plus court d'une origine à une destination à une heure donnée, selon une offre de TC donné (définie à l'horaire). Les requêtes sont traitées sur la plateforme OSRM qui retourne des résultats dits « à l'horaire ».

Travaillant avec les données de l'enquête OD 2008, il aurait été intéressant d'avoir la définition de l'offre TC de 2008 mais, le standard GTFS étant encore à ses débuts à cette époque, aucune définition suffisamment complète de l'offre TC de 2008 n'était disponible. C'est donc le GTFS 2013 qui a été utilisé pour ces simulations. Ce choix n'a pas d'impact sur le développement méthodologique proposé dans ce mémoire mais peut affecter certains résultats.

Les temps de parcours issus de la simulation incluent le détail de chaque segment du voyage : le temps d'accès à l'origine, le temps d'attente initial, le temps passé en véhicule, le temps d'attente à la ou les correspondance(s), au besoin, et le temps d'accès à la destination.

Les paramètres utilisés pour les simulations fixent à un maximum de vingt minutes de marche les temps d'accès à l'origine, à la destination et de correspondance (transfert). Aussi, un minimum de cinq minutes d'attente a été fixé pour tout transfert. Selon ces paramètres, 274 982 temps de déplacements ont été calculés sur les 319 437 ipères soumis à la simulation. On retrouve donc 44 455 déplacements, soit 13,9% de l'ensemble des déplacements, n'ayant pas trouvé d'itinéraire TC correspondant aux critères. Si l'on compare avec MADITUC qui avait un taux d'impossibilité de simulation de 8,19%, l'écart peut s'expliquer le choix des paramètres de la fonction d'utilité, l'identification des nœuds d'accès, l'utilisation de fréquences plutôt que d'un horaire détaillé, etc. principalement par la différence dans la base de données initiale. Les informations nécessaires pour mieux évaluer les différences ne sont toutefois pas disponibles. En effet, le MOTREM ne simulant que les 262 653 déplacements ayant été faits en mode motorisé, une plus grande quantité de déplacements faits en TC est disponible avec les simulations de la Chaire Mobilité puisque les déplacements faits avec un mode actif sont aussi simulés. Toutefois, comme le démontre l'analyse de ces déplacements, pour la majeure partie des déplacements faits à pied, le TC n'est tout simplement pas une option viable.

Afin d'isoler les facteurs pouvant influencer sur les ratios de compétitivité, des temps de parcours minimisés ont été calculés. Ces temps représentent le temps de parcours en TC si le départ avait été parfaitement synchronisé avec l'offre TC, c'est-à-dire que le temps d'attente initial avait été nul. L'utilisation de ces données sera notée TC\* à des fins comparatives dans ce chapitre; elles seront utilisées par défaut dans le scénario de référence 2008. Cette hypothèse peut se justifier par le fait que la déclaration des heures de départ est souvent arrondie dans les enquêtes et par le fait que les déplacements auto font abstraction du temps dédié à la recherche de stationnement, à l'accès à l'origine et à la destination.

Les Figure 3-4, Figure 3-5 et Figure 3-6 comparent les données TC et TC\*, soit les temps simulés en transport collectif pour l'ensemble des déplacements de l'enquête OD 2008. La Figure 3-4 montre la distribution des temps d'attente initiaux en fonction de l'heure de départ. On observe que les déplacements effectués entre minuit et environ 5h am ont une grande composante de temps d'attente initial. Ce phénomène est le résultat de personnes se rendant à l'arrêt de TC au milieu de la nuit alors que le service n'est pas en opération avant le matin. En effet, tel que le montre la Figure 3-5, les temps de parcours moyen pour cette période sont principalement dictés par ces temps d'attente.

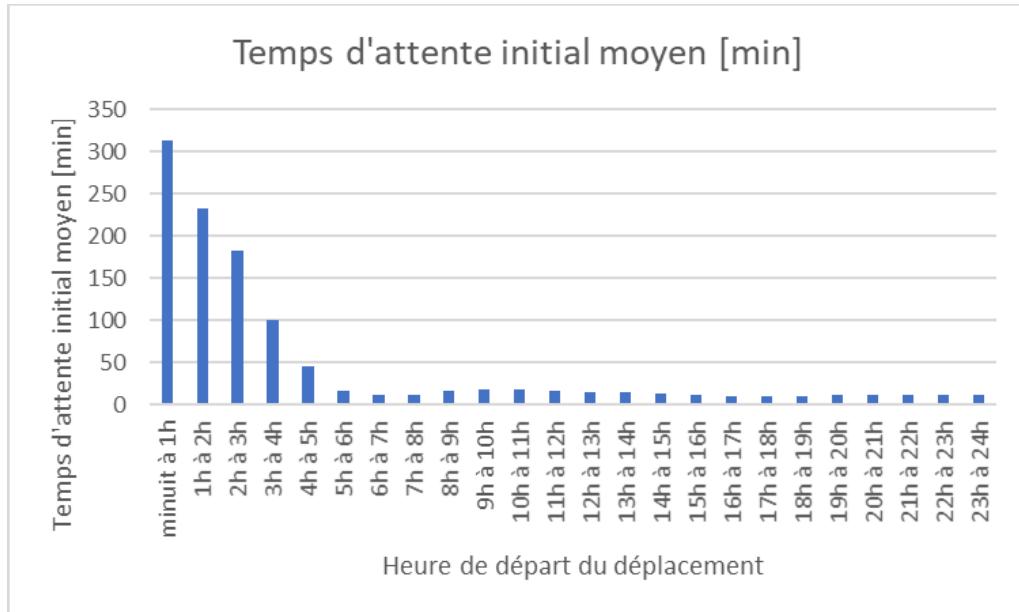


Figure 3-4 : Distribution des temps d'attente initiaux moyens TC

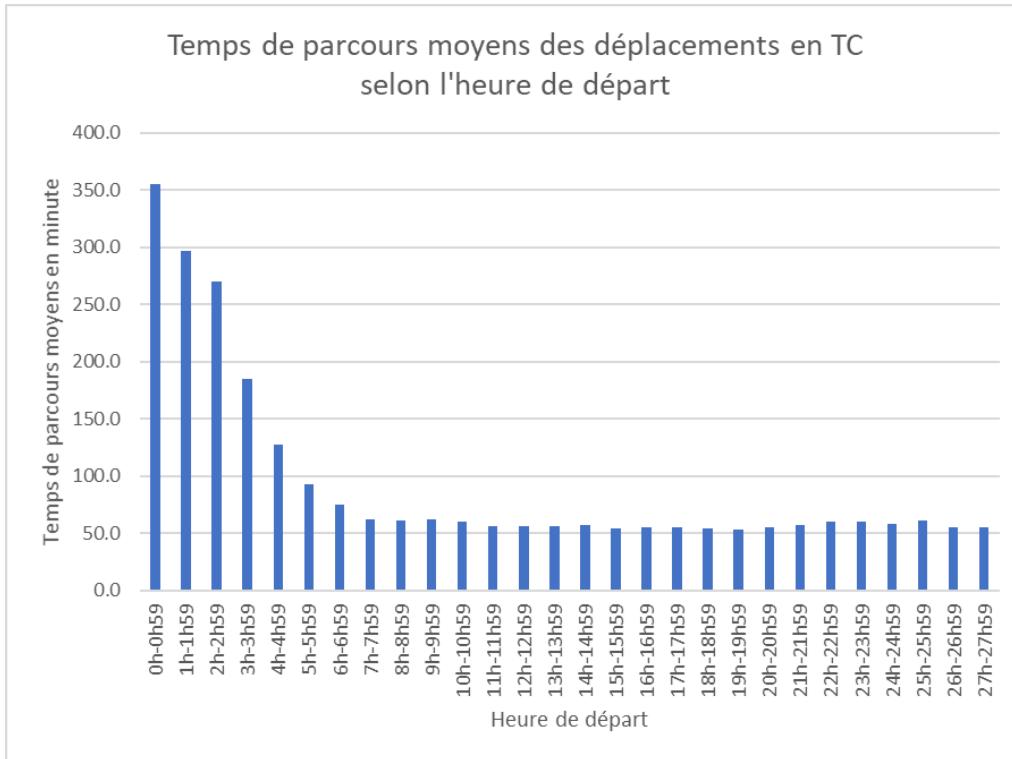


Figure 3-5 : Temps de parcours moyens des déplacements en TC selon l'heure de départ

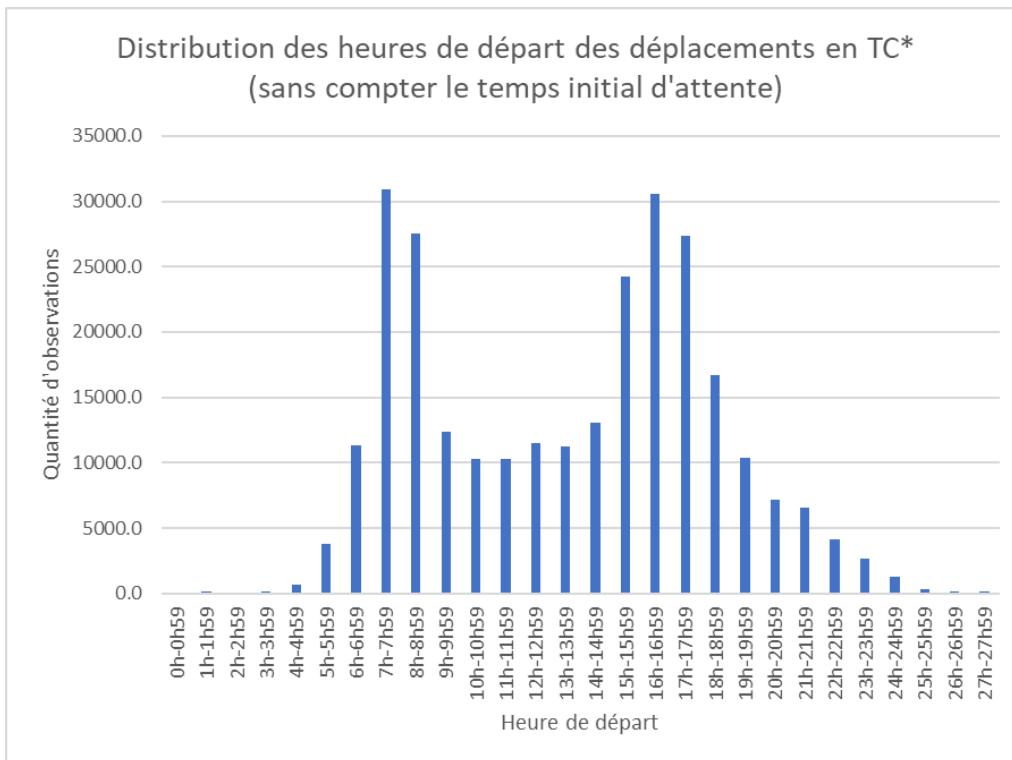


Figure 3-6 : Distribution des heures de départ des déplacements en TC\*

La Figure 3-6 illustre la distribution des départs dans le temps, sous l'hypothèse que ceux-ci sont synchronisés avec le service de TC. On y voit que les premiers départs débutent principalement vers 5h am, laissant la période de minuit à 5h dépourvue de service TC. Ce phénomène s'explique par la définition du service dans les GTFS qui définissent la journée de la 4<sup>e</sup> à la 28<sup>e</sup> heure. Cette période est plus représentative d'une journée de transport collectif. Toutefois, l'enquête OD contient des départs dès l'heure 0 et jusqu'à l'heure 28. Cette différence de définition de la journée est la cause de ces résultats.

De ces graphiques se dégagent deux conclusions. La première est que les temps TC pour la période minuit-5h am sont principalement dictés par le temps d'attente initial et ne sont donc pas représentatifs de la compétitivité modale puisqu'il n'y a pas de service TC.

La seconde est que l'utilisation des données TC\* semble plus représentative pour les fins de modélisation modale. Il est toutefois impossible d'utiliser les départs déclarés dans la période minuit - 5h puisque les ratios de compétitivité souffriraient d'une importante distorsion temporelle. En effet, un départ TC déclaré à 1h am comparera le temps de parcours auto à 1h am au temps TC effectué à 5h am. Ainsi, c'est un total de 1 518 déplacements, soit 0,6% des observations, qui seront exclus des analyses.

### 3.3.3.3 Temps vélo et marche

Les simulations des temps vélo et marche ont également été effectuées avec l'aide d'OSRM. Les paramètres de simulation ont été définis à la Chaire Mobilité tels que les cyclistes se déplacent à une vitesse moyenne de 15 km/h et subissent une pénalité de 20 secondes sur les virages à gauche alors que les piétons marchent à 5 km/h en toutes circonstances. La définition des dénivélés dans les modèles numériques n'étant pas assez précise, ce facteur n'est pas pris en considération.

## 3.4 Comparaison MOTREM vs Chaire Mobilité

Ayant des temps de parcours auto et TC provenant de deux outils différents, il est intéressant de les comparer, d'autant plus que les méthodes de simulation sont très différentes. Du côté du MOTREM08 (MO), les modèles sont calibrés sur une base régionale et par affectation dynamique jusqu'à équilibre, c'est-à-dire que les temps de déplacement auto sont estimés de centroïde à centroïde et que les déplacements internes sont estimés en fonction de la superficie de la zone et de sa densité de population. Les simulations de la Chaire Mobilité (CM) sont quant à elles basées

sur des itinéraires uniques pour chacun des déplacements dont les vitesses sont ajustées à l'aide de données historiques. En effet, les vitesses automobiles sous congestion sont basées sur une base de données GPS de déplacements taxis. Les temps de parcours des déplacements TC sont calculés à partir de fichiers GTFS, donc à l'horaire.

### 3.4.1 Comparaison des temps auto

La comparaison des données provenant du MO et de la CM est intéressante dans l'optique de l'évaluation des résultats obtenus des deux méthodes. Ainsi, une analyse comparative des temps par déplacement est proposée. Pour ce faire, deux facteurs comparatifs sont étudiés : les différences de temps de parcours (DT, équation 3.1) ainsi que le ratio de ceux-ci (RT, équation 3.2). Avec ces indicateurs, une représentation graphique des grands écarts pourra aider à identifier la source des différences.

$$(3.1) \quad DT = Temps_{Chaire} - Temps_{Motrem}$$

$$(3.2) \quad RT = \frac{Temps_{Chaire}}{Temps_{Motrem}}$$

Notons que les simulations du MO ont été calibrées pour la période de pointe AM (6h00 à 8h59). Les documents complémentaires (accompagnant les fichiers de données) mettent en garde contre une utilisation des données pour les périodes hors pointe AM. La comparaison des résultats des simulations sera donc limitée à la période de pointe AM qui contient 31 923 déplacements. De plus, les deux méthodes ont beaucoup de paramètres différents pouvant être à la source de divergences. Parmi les plus évidentes, le MO calcule les vitesses à partir de modèles en régime permanent alors que la CM s'appuie sur des vitesses historiques observées. Finalement, l'utilisation de l'offre TC 2013, à défaut d'avoir une définition complète de l'offre 2008, peut évidemment causer des différences entre les résultats provenant des deux modèles. Le Tableau 3-2 présente les principales caractéristiques statistiques de ces indicateurs, les temps de parcours étant présentés en minutes.

Tableau 3-2 : Statistiques descriptives de DT et RT, comparaison temps auto CM et MO

	Min	Max	Moyenne	Médiane	Écart type
DT [minute]	-83.2	21.1	-2.4	-0.9	6.2
RT	0.006	4.3	0.92	0.91	0.29

Les Figure 3-7 et Figure 3-8 illustrent que la majorité des déplacements sont plus courts dans les simulations de la CM et que plus le déplacement est long, plus la différence est marquée. Les données de la CM étant ajustées sur des données mesurées, il est probable que les vitesses simulées du MOTREM soient sous-estimées. Les Figure 3-9 et Figure 3-10 illustrent l'impact relatif que ces différences ont sur la durée des parcours. L'impact en termes de ratio est particulièrement marqué pour les déplacements de moins d'un kilomètre. Également, les déplacements de plus d'une dizaine de kilomètre sont de 15 à 20% plus rapides que ceux du MO.

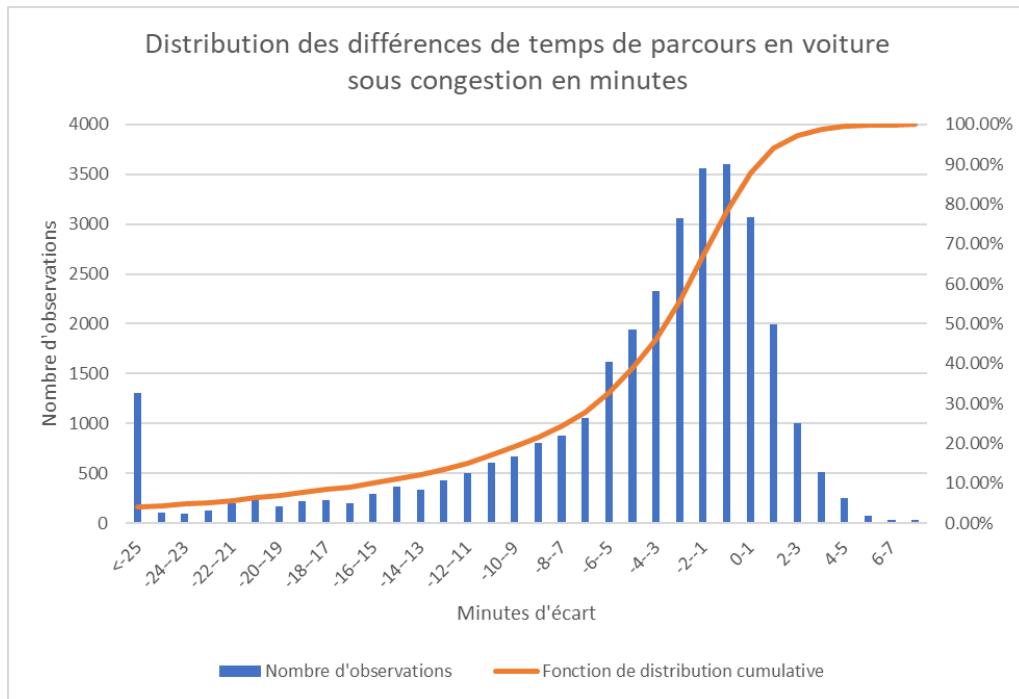


Figure 3-7 : Distribution des différences de temps de parcours

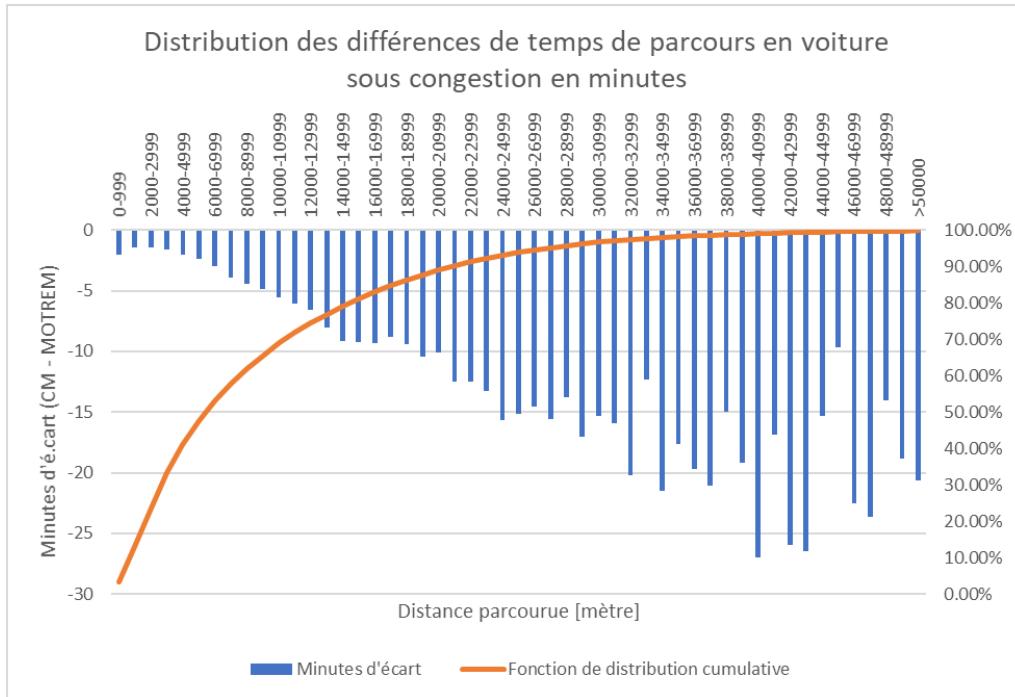


Figure 3-8 : Distribution des différences de temps de parcours Chaire-MOTREM (mins) en fonction de la distance parcourue

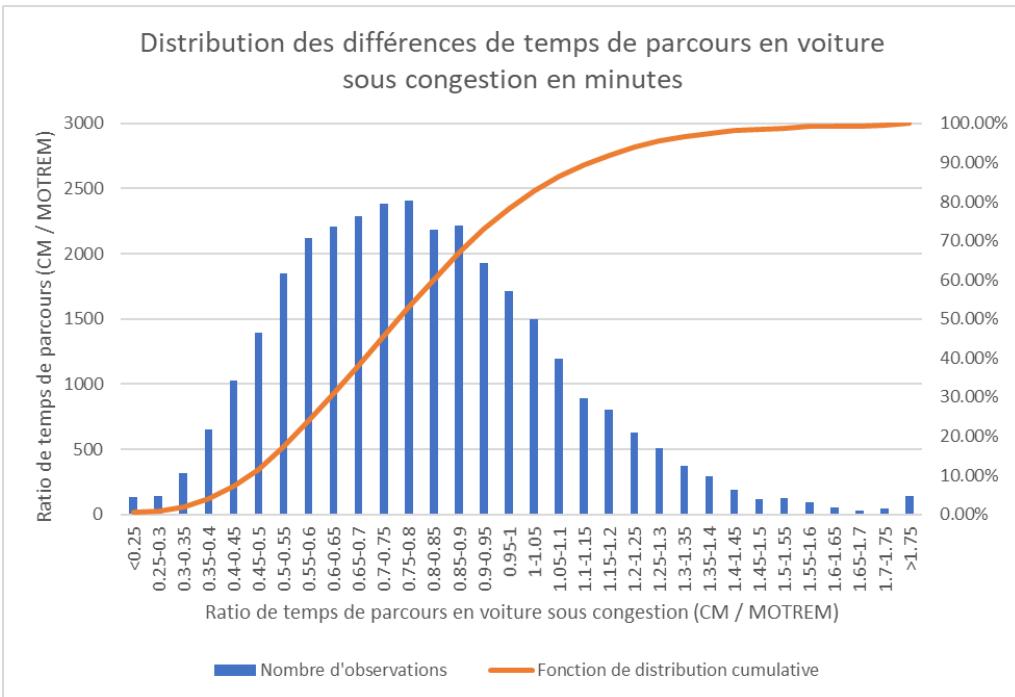


Figure 3-9 : Distribution des ratios de temps de parcours

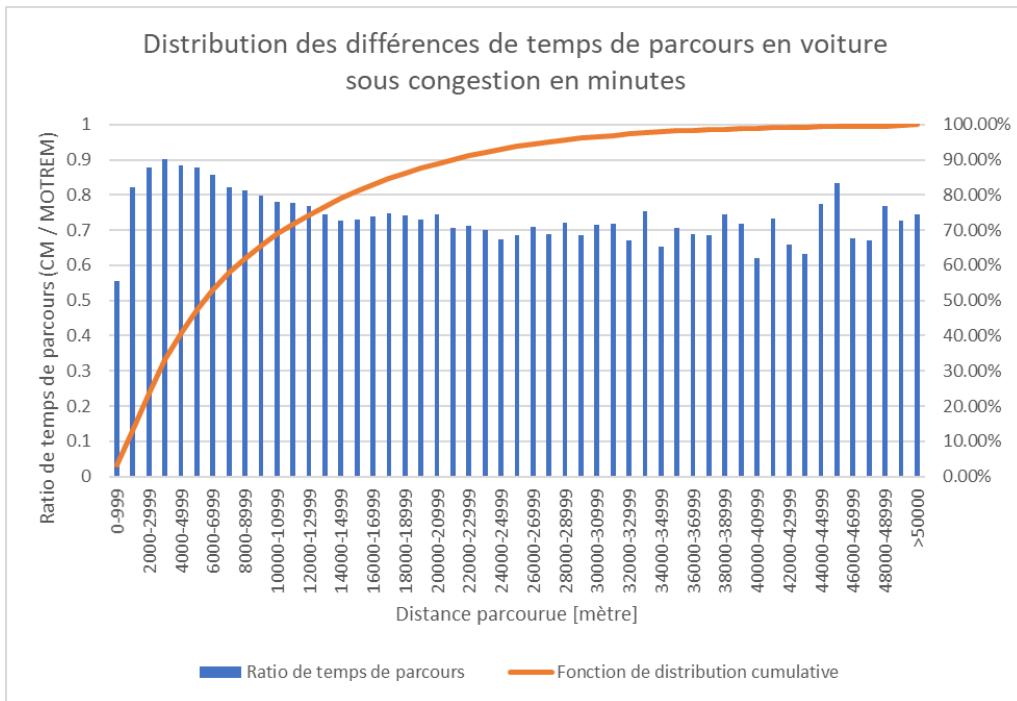


Figure 3-10 : Distribution des ratios de temps de parcours en voiture sous congestion fonction de la distance parcourue

### 3.4.2 Analyse spatiale des écarts dans les résultats

La répartition spatiale des écarts entre les résultats du MOTREM et ceux de la Chaire Mobilité peut nous informer sur la cause, la localisation et les regroupements de ceux-ci. Si on détecte la présence de tendances claires, ces dernières devront être prises en compte dans l’interprétation finale des données. Plus la valeur est élevée (rouge), plus les temps CM sont supérieurs à ceux du MO; inversement, plus la valeur est faible (négative – vert), plus les temps CM sont faibles par rapport aux temps MO.

Les Figure 3-11 et Figure 3-12 représentent la distribution des écarts bruts, en minutes, entre les simulations de la CM et celles du MO. Notons que dans ce mémoire, les données utilisées lors des analyses spatiales sont agrégées à leur point d’origine. La graduation des données correspond aux quintiles de la distribution des différences.

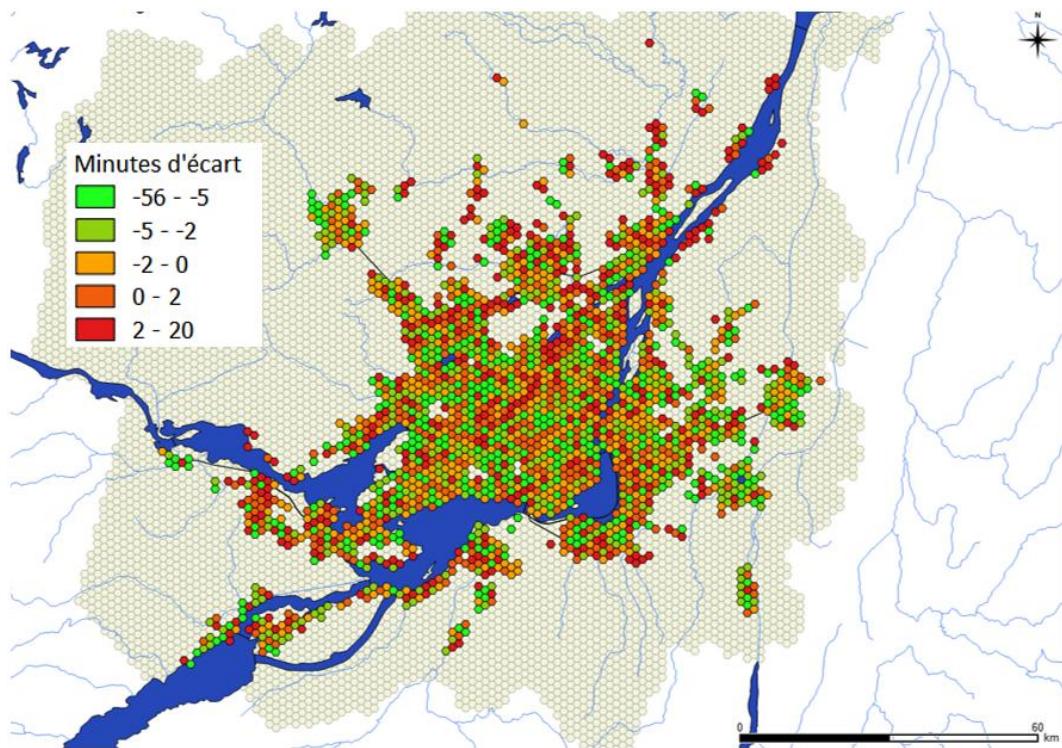


Figure 3-11 : Différences de temps de parcours moyens, auto sous congestion, en minutes

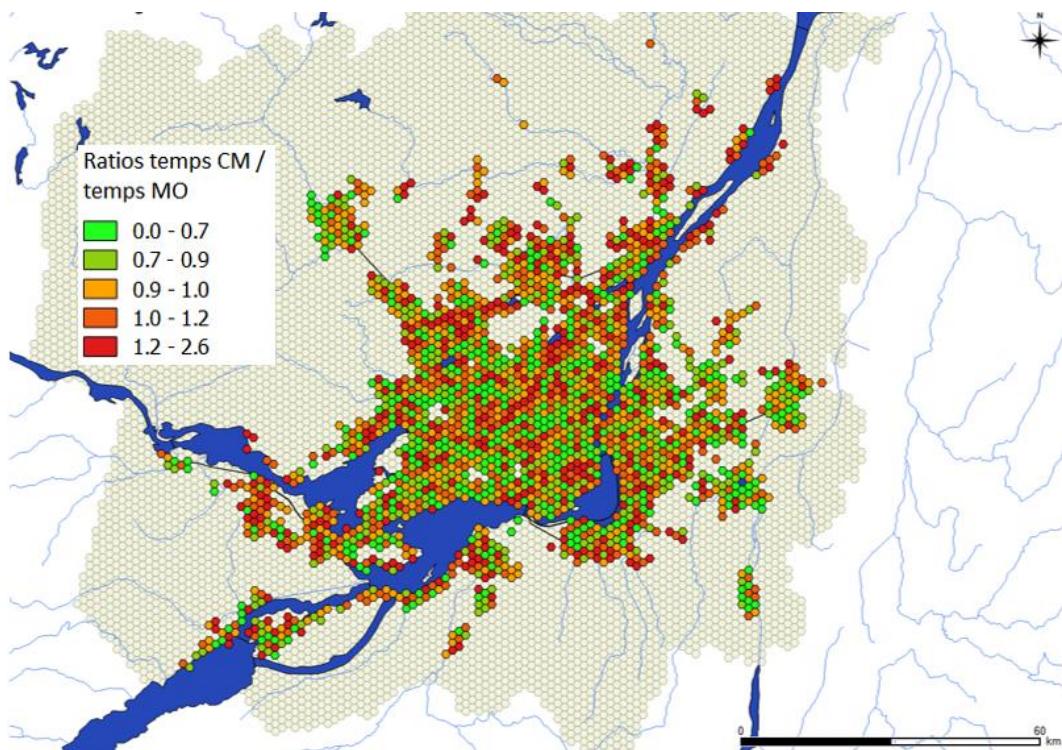


Figure 3-12 : Ratios moyens par zone, temps de parcours en auto sous congestion

Afin d'avoir une bonne compréhension des écarts entre les simulations des deux outils, analysons leur dispersion spatiale. Rappelons que l'indice global de Moran représente le niveau d'autocorrélation globale des résultats et peut prendre des valeurs allant de -1 à 1. Pour plus de détails à propos de cet indice, svp consulter les travaux de (Moran 1950). Le Tableau 3-3 présente les indicateurs de Moran. Les valeurs à peine plus grandes que zéro indiquent une auto corrélation spatiale quasi inexistante.

Tableau 3-3 : Indices de Moran des écarts de temps de parcours auto

	Indice global de Moran	z-value	Présence d'autocorrélation spatiale ?
Différence temps de parcours auto	0.014782	0.9665	non
Ratios temps de parcours auto	0.020609	1.3374	non

Des résultats du Tableau 3-3, les valeurs de l'indicateur de Moran sont très près de 0, ce qui indique que la distribution spatiale des données est majoritairement aléatoire. Les indices de significativité statistique z sont d'ailleurs en deçà du seuil de 1,96, ce qui confirme l'absence d'autocorrélation spatiale significative pour les deux indicateurs comparatifs.

De cette comparaison, il semble que les simulations auto effectués à la CM sous-estiment les temps de parcours, comparativement au MO, en moyenne de 8% et ce, de manière généralisée sur le territoire mais sans toutefois présenter de structure spatiale claire.

### 3.4.3 Comparaison des temps TC

En planifiant son départ en fonction de l'heure de desserte du TC, il est possible de sauver le temps d'attente initial. Un nouveau jeu de donnée est donc créé en posant comme nul le temps d'attente initial. L'utilisation de ce jeu de données (à temps d'attente initial nul) est identifiée par l'utilisation de l'étoile (\*).

Le Tableau 3-4 présente les statistiques descriptives des écarts calculés entre les temps de parcours provenant du MO et de la CM. La section TC représente les temps de parcours entier et la section TC\*, les temps de parcours avec l'attente initiale posée comme égale à zéro. On y voit qu'en moyenne, les temps de parcours TC calculés par la CM sont 13,27 minutes plus long, soit l'équivalent d'environ 36%. En isolant l'effet des temps d'attentes initiaux, cette différence moyenne est réduite à près de 7 minutes. Sachant que la définition du service TC utilisée par la CM

est « à l'horaire » et que le MO utilise des simulations en régime permanent, il est possible que les temps simulés par le MO soient sous-estimés.

Tableau 3-4 : Statistiques descriptives, comparaisons Chaire Mobilité et MOTREM

TC		TC*	
	Déférence [min]	Ratio	Déférence [min]
Moyenne	13.27	1.36	6.99
Médiane	7.54	1.21	4.21
Écart type	45.22	1.00	32.92
			0.64

Les distributions des différences de temps de parcours TC\* est présenté la Figure 3-13 et celle des ratios à la Figure 3-14. Ces figures illustrent que le temps de parcours de 70% des déplacements est plus long lorsque simulé par la CM. Les temps sont en moyenne 20% plus longs que leur contrepartie MO.

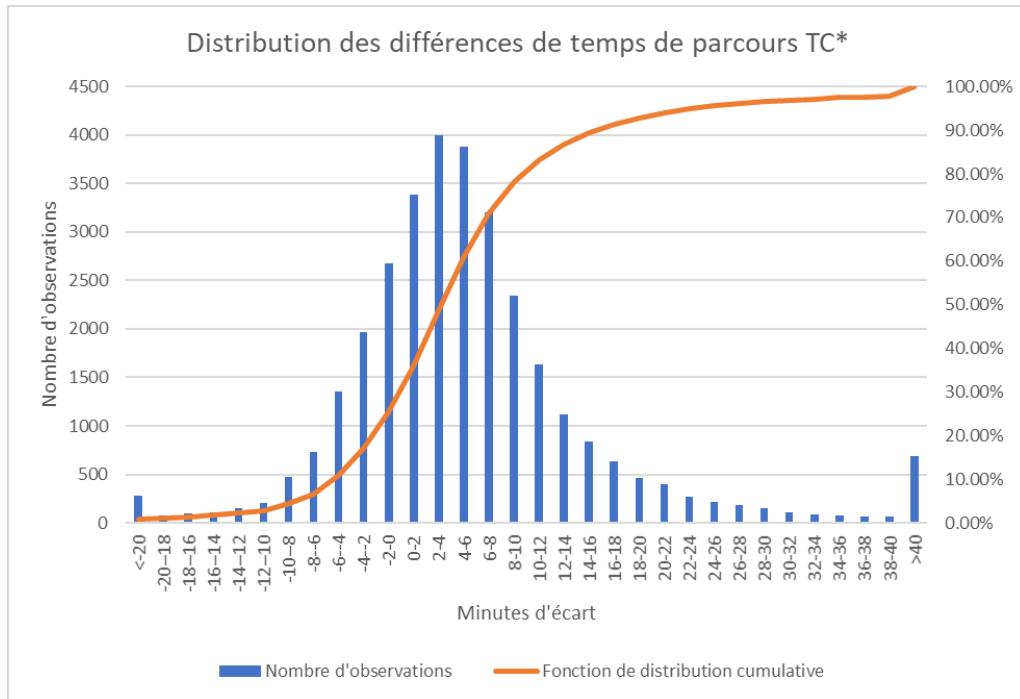


Figure 3-13 : Distribution des différences de temps de parcours TC\*-CM vs TC-MO

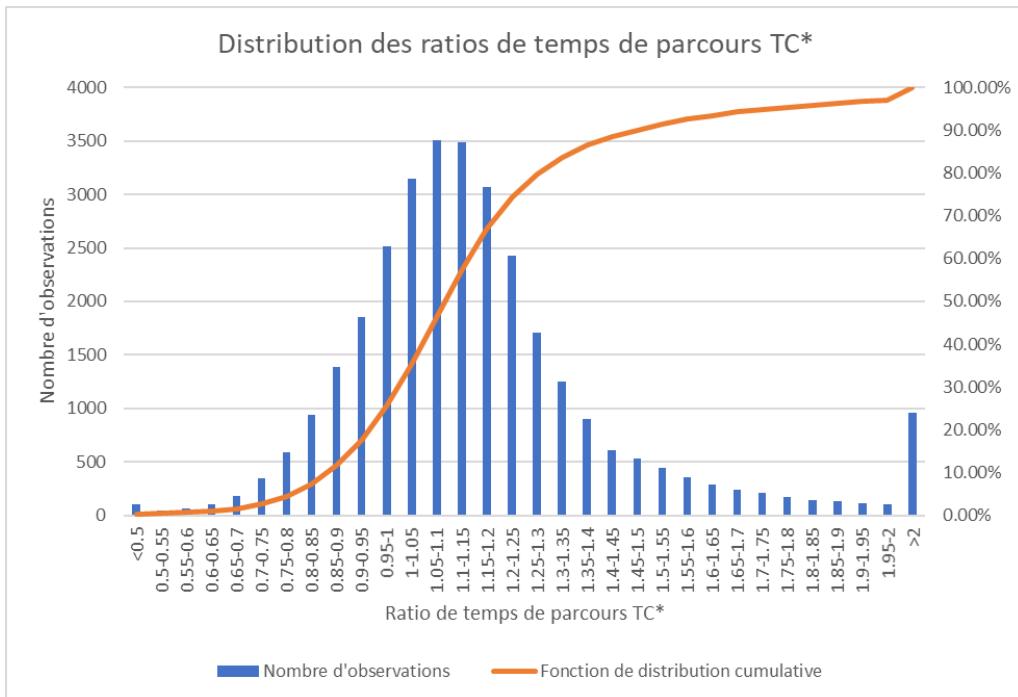


Figure 3-14 : Distribution des ratios de temps de parcours TC\*-CM vs TC-MO

### 3.4.4 Analyse spatiale des erreurs

Le Tableau 3-5 indique qu'en retirant les temps d'attentes initiaux, la valeur des indicateurs d'autocorrélation spatiale de Moran diminue grandement. Une telle différence entre le jeu de données TC et TC\* peut indiquer la présence d'autocorrélation spatiale dans la répartition des temps d'attentes initiaux. Voyons donc la répartition spatiale de ces temps d'attentes ainsi que les regroupements spatiaux associés à l'indice de Moran le plus élevé, soit les différences de temps TC\*.

Tableau 3-5 : Indicateurs de Moran

	Indicateur de Moran, TC	z-value	Indicateurs de Moran, TC*	z-value
Différence des temps	0,562839	300.06	0,215566	153.87
Ratios des temps	0,529382	56.0875	0,182224	-33.56

#### 3.4.4.1 Différences des temps de parcours TC\*

La Figure 3-15 présente les clusters de Moran associés aux différences de temps de parcours TC\*.

La légende identifie les couleurs associées aux différents clusters de Moran. Le #0 représente les

secteurs non significatifs, le #1 les secteurs où les DT sont élevées et qui sont entourés d'autres secteurs à valeur élevée de DT (haut-haut), le #2 les secteurs à valeur faible entourés d'autres secteurs à valeur faible (bas-bas), le #3 les secteurs à valeur faible entourés par des secteurs à valeur forte (bas-haut), le #4 les secteurs à valeur forte entourés de secteurs à valeur faible (haut-bas) et finalement, le #5 pour les secteurs sans voisin. Pour des fins de visualisation, les cours d'eau ont été retirés pour les graphiques de clusters de Moran.

Les différences de temps de parcours ne sont pas regroupées sur l'île de Montréal, mais plutôt en périphérie. On y distingue les valeurs élevées de DT en rouge, c'est-à-dire les endroits où les temps de la CM sont significativement inférieurs à ceux du MO. Ces regroupements de valeurs DT élevées sont situés dans les municipalités de St-Jean-sur-Richelieu, Châteauguay tout en rouge, c'est-à-dire que la valeur DT est élevée. En bleu sont identifiées les régions où les temps de la CM étaient significativement inférieurs à ceux du MO. Ces endroits se retrouvent à proximité des lignes de trains identifiées en noir sur la Figure 3-15. En plus de St-Jérôme, St-Lazare, Vaudreuil, Candiac et Mont-Saint-Hilaire, L'Épiphanie semble se démarquer en termes de performance dans les temps de parcours TC.

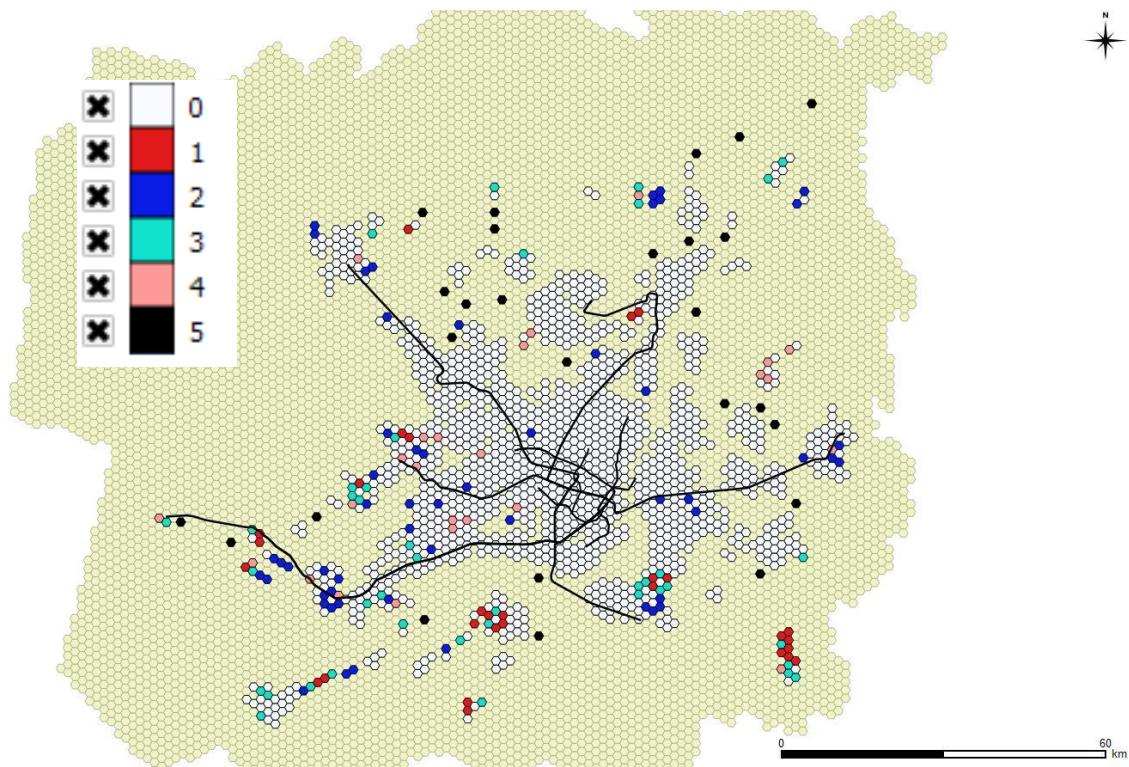


Figure 3-15 : Clusters de Moran pour différences de temps de parcours TC\*

#### 3.4.4.2 Temps d'attentes initiaux

Il est commun dans la littérature d'assumer un temps d'attente initiale de la moitié de la valeur de l'intervalle pour les lignes où l'intervalle est de moins de 10 à 15 minutes. Au-delà de ces valeurs d'intervalles, les gens ne se rendent plus à l'arrêt de manière aléatoire et il n'y a pas de consensus, à l'heure actuelle, sur la meilleure hypothèse de temps d'attente pour ces déplacements.

La différence d'indice de Moran du Tableau 3-5 suggère que l'autocorrélation remarquée dans le jeu de données TC vient du temps d'attente initial moyen inclus dans ces temps de parcours.

La Figure 3-16 représente les temps d'attente initiaux moyens par secteur. La desserte sur l'île est bien plus fréquente que dans les couronnes. En effet, l'indice de Moran de la variable temps d'attente moyen indique une forte autocorrélation d'une valeur de 0,387. La Figure 3-17 identifie spatialement les clusters de Moran avec la même définition de cluster que précédemment.

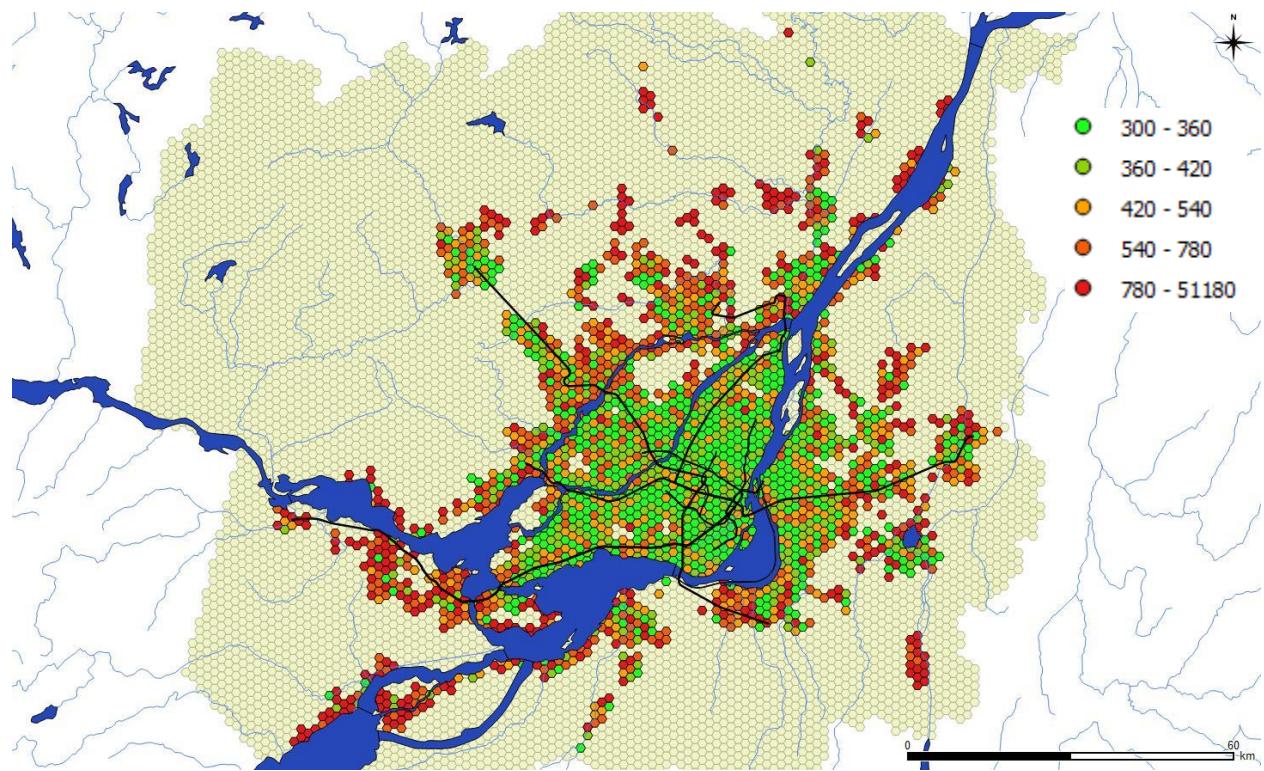


Figure 3-16 : Répartition spatiale des temps d'attentes initiaux en TC, légende en secondes

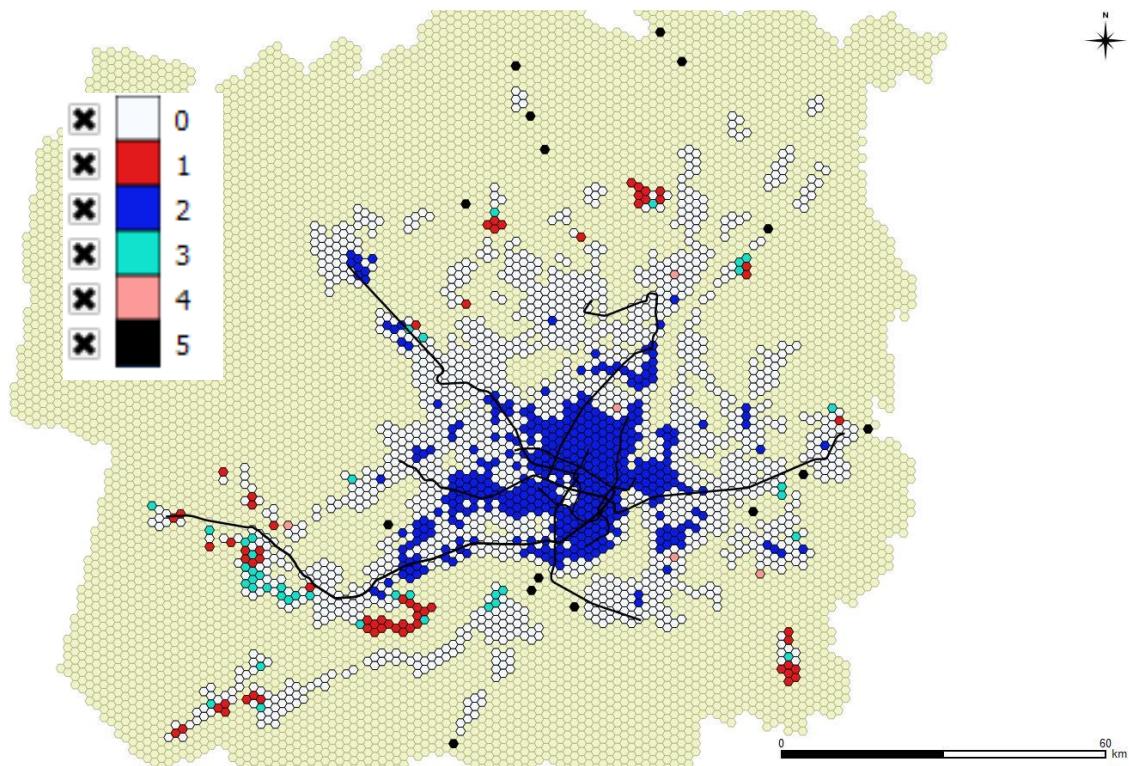


Figure 3-17 : Cluster de Moran des temps d'attentes initiaux

## 3.5 Préparation des données

Les données provenant du MOTREM étant calibrées uniquement en pointe am, elles ont servi à des fins comparatives mais ne seront pas utilisées dans les modèles présentés dans ce mémoire. Ce sont plutôt les temps de parcours issus des simulations faites à la Chaire Mobilité qui seront utilisés, notamment puisqu'ils peuvent être simulés pour toute heure de départ.

La toute première étape de la méthode d'analyse est la création de la base de données qui servira à l'étude. Les données de l'enquête OD et les temps simulés pour chaque mode sont mis en commun, puis filtrés en plusieurs étapes, tel que présenté dans le Tableau 3-6. L'objectif est d'écartier les données non pertinentes, aberrantes ou pouvant nuire à l'élaboration des modèles. Chaque étape de filtration est expliquée dans les paragraphes suivants. Suite à ces étapes, la base de données sera prête à utilisation pour la modélisation.

Tableau 3-6 : Étapes de filtration des données

	Étapes de filtration				
	Description	Observations filtrées	% retirées	Observations restantes	% du total
0	Enquête OD 2008	N/A	N/A	355 459	100%
1	Observations des non-mobiles	35 871	10.1%	319 588	90%
2	Départ de nuit (avant 5h am ou après 24h)	1 991	0.6%	317 597	89%
3	Observations des personnes de 4 ans ou moins	2 811	0.9%	314 786	89%
4	Déplacements hors zone	55 742	17.7%	259 044	73%
5	Vecteurs partiels	3 075	1.2%	255 969	72%

### 3.5.1 Filtre #1 – Observations des non-mobiles

Cherchant à dresser le portrait de la mobilité des montréalais, l'enquête OD reflète également qu'une partie de la population a déclaré ne pas s'être déplacée. Seules les observations correspondant à un déplacement sont gardées. Cette première étape de filtration retire 35 871 observations, soit 10,1% des enregistrements disponibles dans le fichier de l'enquête OD 2008.

### 3.5.2 Filtre #2 – Départs avant 5h et après 24h

Les départs simulés en transport en commun avant 5h génèrent des résultats inadaptés pour les objectifs de ce travail. Le problème vient probablement d'une disparité dans la concordance des données utilisées lors des simulations. En effet, les départs de l'enquête OD 2008 sont définis de 0h à 28h alors que l'horaire de service défini dans le GTFS est défini de 5h à 28h. Le résultat de cette situation est telle qu'illustrée en section 3.3.3.2. Ces déplacements auraient pu être juxtaposés avec ceux de 24 h à 28 h, mais puisque cela n'a pas été le cas initialement, les 1518 départs dans cette période horaire ne peuvent être utilisés. Dans la période entre 24 h et 28 h, on retrouve 473 observations. Bien que potentiellement utilisables, ces données ont été retirés puisque les autres départs de nuit l'ont également été. Cette exclusion n'affecte pas la démonstration méthodologique présentée dans ce mémoire et ces déplacements pourront être réintégrés dans des démarches ultérieures, en conjonction avec les autres déplacements nocturnes retirés dans cette section. Ensemble, les deux plages horaires nocturnes contiennent 1991 déplacements, soit 0,6% de l'enquête OD.

### 3.5.3 Filtre #3 – Personnes de 4 ans et moins

L'enquête OD contient l'information sur les déplacements effectués, mais aussi de l'information sur les ménages et la composition de ceux-ci. Plusieurs enfants de quatre ans ou moins sont ainsi recensés dans les ménages bien que leurs déplacements ne soient pas systématiquement documentés. Dans l'enquête OD 2008, 2841 observations proviennent de tels enfants, soit 0,9% du total, et sont retirés.

### 3.5.4 Filtre #4 – Déplacements hors zone

Lors des entrevues téléphoniques, tous les déplacements effectués lors de la journée enquêtée sont recueillis et il arrive que certains aient une extrémité de déplacement hors du territoire d'enquête. Ainsi, tel qu'illustrées sur la Figure 3-18, les coordonnées associées aux extrémités de déplacements hors du territoire d'enquête sont réparties de manière circulaire autour de la région et ne sont pas représentatives des extrémités réelles du déplacement. Puisque les déplacements ayant l'une de ses extrémités hors-territoire ne peuvent pas être simulés en totalité avec les réseaux disponibles, seuls les déplacements dont les origines et les destinations sont à l'intérieur de la zone

sont analysés. Ainsi, 55 742 observations sont retirées, soit 17,7% des enregistrements du fichier de l'enquête OD.

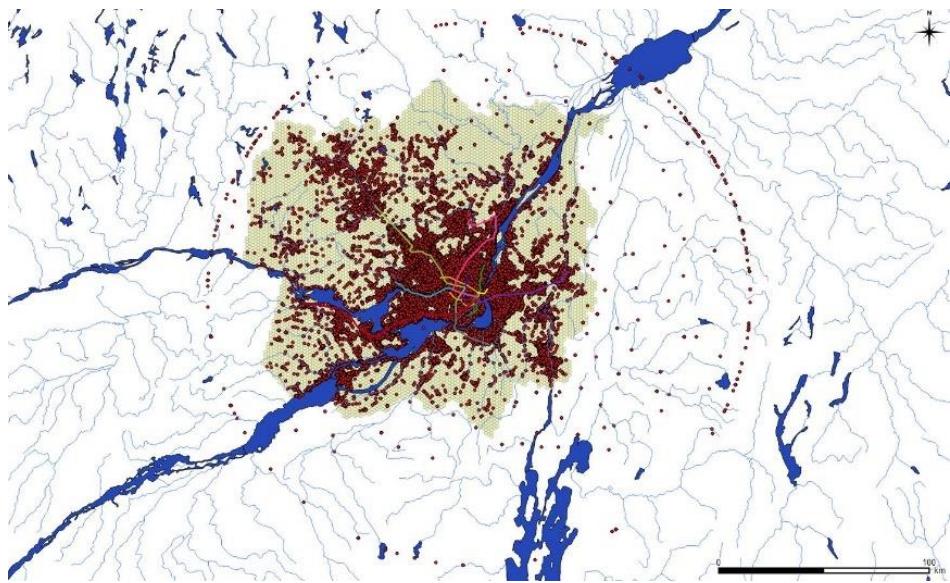


Figure 3-18 : Déplacements avant filtre #4

### 3.5.5 Filtre #5 – Impossibilité de simuler les déplacements avec certains modes

La base de données finale utilisée pour l'analyse est formée de données provenant de plusieurs simulations. Toutefois, tel que présenté en section 3.3, chacune des simulations est parvenue à calculer des quantités différentes de temps de parcours. Le Tableau 3-7 présente la quantité de données issue de chaque simulation.

Tableau 3-7 : Données disponibles par mode

Mode	Nombre d'enregistrements
Temps auto en congestion	311 128
Temps TC	274 982
Temps vélo	320 254
Temps marche	320 320

La simulation TC est la plus restreignante, limitant la base de données à un maximum de 274 982 enregistrements. Un temps de parcours manquant empêche le calcul d'un ratio et crée un vecteur incomplet. La raison principale pour le nombre inférieur d'enregistrements en TC est la disponibilité du service. Lors des simulations, un seuil de vingt minutes à pied a été fixé comme limite maximale, tant pour l'accès initial au service que pour l'accès à la destination, que pour

l'accès au transfert. Ainsi, 13,9% des enregistrements n'ont pas pu trouver de parcours TC. Cette étape retire 3075 observations qui avaient une définition partielle de leur vecteur de compétitivité. Suite à ce 4<sup>ème</sup> filtre, la base de données finale est composée de 255 969 observations, soit 72% de l'enquête OD.

Bien que cette méthode propose un filtre efficace aux fins d'analyse et de modélisation, certaines paires OD pourraient trouver un itinéraire suite à des modifications du réseau TC. Avec la méthode actuelle, l'évolution de ces déplacements ne pourrait être capturée puisqu'ils ont été retirés de la base de données. Cette limitation pourrait être abordée dans la poursuite des travaux de recherche portant sur cette méthode.

## CHAPITRE 4 SITUATION DE RÉFÉRENCE 2008

Ce chapitre présente la situation de référence qui sera la base de toutes les analyses et modélisations. Puisque les données de base utilisées proviennent de l'enquête OD 2008, la demande de transport de 2008 servira de référence. Toutefois, la définition du réseau de transport disponible en 2008 est incomplète. Tel que mentionné précédemment, pour remédier à cette situation, le choix a été fait d'utiliser le réseau de transport offert aux usagers en 2013. Ce choix n'est pas idéal car les réseaux ont pu évoluer, alors que les choix comportementaux observés l'ont été à la lumière de l'offre de transport disponible en 2008. Il faudra analyser les résultats étonnantes plus spécifiquement en tentant de voir si les évolutions 2008-2013 des réseaux de transport pourraient contribuer à les comprendre.

### 4.1 Statistiques descriptives

Le scénario de référence 2008 est basé sur l'enquête OD 2008. Le Tableau 4-1 présente en parallèle les statistiques de l'enquête OD 2008 en entier et les données conservées aux fins de modélisation. On y voit les impacts des étapes de filtration des données.

Tableau 4-1 : Statistiques descriptives du scénario de référence

		Statistiques descriptives, scénario référence 2008			
		Enquête OD		Données modélisation	
		Observé	Pondéré	Observé	Pondéré
Déplacements	Déplacements	355 459	8 303 685	255 969	6 662 389
	Distance moyenne	8.2	8.2	6.7	6.8
Population	Taille	156 968	4 068 708	96 728	2 513 136
	% Femmes	51.61%	50.70%	51.86%	50.68%
Part modale	Auto cond	53.28%	53.17%	50.47%	50.23%
	Auto pass	13.60%	13.17%	13.12%	12.69%
	TC	15.39%	16.00%	18.48%	19.20%
	Vélo	1.37%	1.41%	1.41%	1.46%
	Marche	10.92%	10.96%	11.88%	11.91%
	Autre	5.45%	5.28%	4.65%	4.51%
Personne-km	Auto cond	1 707 321	44 729 614	1 046 375	27 371 662
	Auto pass	349 984	8 780 533	209 569	5 280 224
	TC	396 206	10 711 153	365 933	9 877 504
	Vélo	14 606	394 444	12 176	328 998
	Marche	27 723	728 961	23 588	617 664
	Autre	122 532	3 094 003	63 596	1 578 511

Le retrait des déplacements ayant une extrémité hors du territoire d'enquête est probablement la cause majeure de la réduction de la distance moyenne des déplacements et de la légère augmentation de la part modale du TC au dépend de l'automobile conducteur et des autres modes. La catégorie autre mode inclut l'autobus scolaire, un autre autobus, le transport adapté, un avion, un train ou un autobus interurbain.

## 4.2 Compétitivité et heure de départ

Plusieurs facteurs ont un impact sur la compétitivité modale. Étant donné la définition des ratios de compétitivité comme étant un ratio de temps de parcours, les facteurs influençant la compétitivité modale affectent les temps de parcours, que ce soit celui des automobiles ou ceux du mode à l'étude. Voyons les facteurs qui varient au cours de la journée et qui peuvent avoir un impact sur le temps de parcours. Du côté de l'automobile, la congestion routière, l'itinéraire et la distance à parcourir sont les facteurs principaux. Du côté du TC, nommons la présence de lignes express, la présence et les périodes de validité des voies réservées, la fréquence de service ainsi que la congestion routière étant donné que la majorité des lignes opèrent en circulation mixte.

Les temps de parcours des modes actifs sont, quant à eux, indépendants de la congestion automobile ou d'une offre extérieure. Ces déplacements sont toutefois plus sujets à être affectés par les aléas du climat. Leurs ratios de compétitivité varieront donc principalement selon la performance de l'automobile, tel que l'illustre la Figure 4-1. On y voit les ratios moyens de chacun des modes en fonction de l'heure à laquelle les départs ont été effectués.

Les ratios TC sont illustrés en bleu et les ratios TC\* en orange. La différence entre ces deux courbes est uniquement due à l'attente initiale, tel que discuté précédemment. Très marquée, cette différence s'atténue lors des périodes de pointe, périodes à laquelle l'offre est plus importante pour accommoder la demande. De manière contraire, cette différence s'amplifie en période de faible demande, que ce soit en milieu de journée ou durant la nuit lorsque le système de métro est fermé. La variation des ratios TC et TC\* au cours de la journée s'explique par un ajustement de l'offre à la demande et la présence ou non de congestion. Les ratios plus compétitifs des heures de pointe sont le reflet de cet ajustement de l'offre.

Les ratios des modes actifs, vélo en gris et marche en jaune, illustrent indirectement la performance de l'automobile. Tel que mentionné plus haut, leur vitesse étant indépendante de la congestion automobile, seuls les temps de parcours automobile et la distance influencent leur variation. Finalement, on remarque que les déplacements effectués en début de journée sont en moyenne beaucoup plus longs.

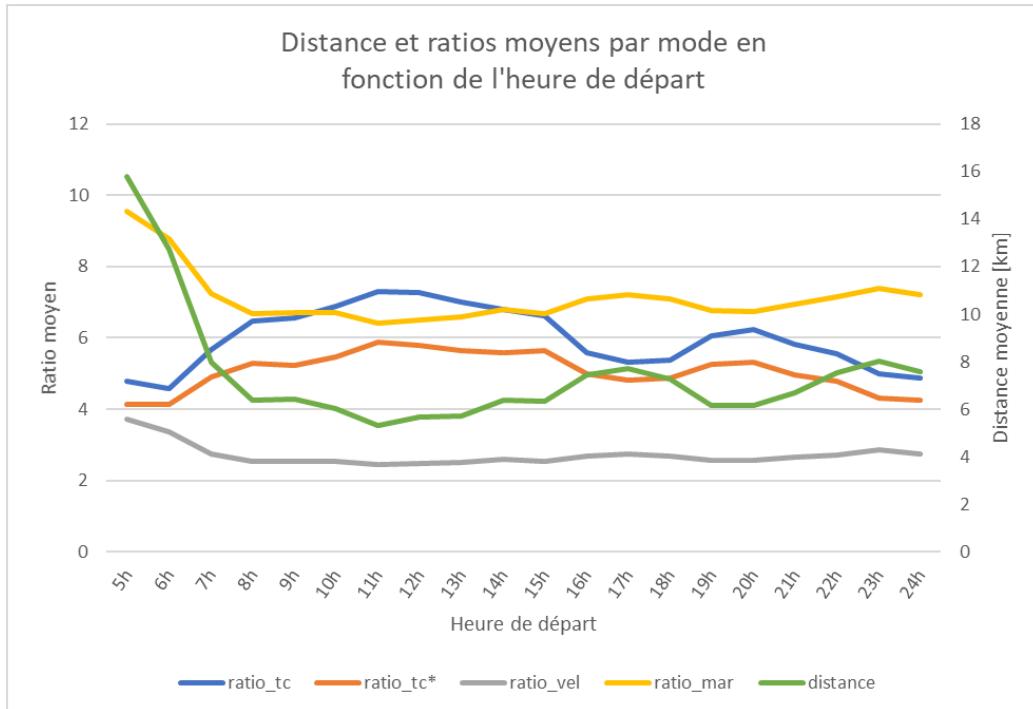


Figure 4-1 : Ratios moyens par période horaire

### 4.3 Segmentation horaire des déplacements

La segmentation de la journée en termes de transport se définit souvent en cinq sections : la nuit/matin, la pointe am, le milieu de la journée, la pointe pm et le soir/nuit. Ces segmentations « classiques » sont généralement bien adaptées pour représenter les différents niveaux de demande de transport. Les pointes représentent les moments de la journée où la demande est maximale, les moments où les travailleurs se dirigent à leur lieu de travail et en reviennent. Toutefois, la segmentation horaire relève souvent d'un choix arbitraire. En utilisant le test de Kolmogorov-Smirnov (KS) pour regrouper les paires d'heures aux fonctions de distribution cumulative (FDC) similaires, il est possible de créer une segmentation horaire basée sur une comparaison statistique.

La méthode proposée est inspirée du travail de Tessier (2016) qui avait utilisé cette méthode à des fins similaires mais avec un autre ensemble de données.

Le test KS consiste à comparer deux fonctions de distribution cumulative tirées de deux échantillons afin de déterminer si elles sont issues d'une même loi. Dans le cadre de ce mémoire, les FDC comparées sont celles des déplacements qui ont été classés en fonction de leur ratio TC/Auto, et ce, pour des périodes d'une heure, représentant chacune des heures de la journée. Ainsi, le test KS présenté cherche à identifier les paires d'heures de départ ayant un profil de performance TC similaires et dissimilaires. Le résultat du test s'exprime en pourcentage et représente la plus grande différence entre les FDC comparées. La Figure 4-2 illustre ce test KS pour deux FDC. Cette distance est accompagnée d'une value p de significativité statistique. Défini à 5%, ce seuil est le critère de rejet de l'hypothèse nulle qui suggère que les distributions analysées sont tirées de la même distribution statistique.

En effectuant le test pour chacune des paires d'heures, on obtient une matrice 24x24 des distances maximales entre chaque paire de FDC. La Figure 4-3 présente un extrait de la matrice des distances pour les heures allant de 5h à 17h59. Les cases rouges témoignent d'une grande similarité de FDC alors que plus les cases sont vertes, plus elles sont dissimilaires. Les encadrés noirs représentent les groupements finaux. La matrice complète du test KS est disponible en annexe A.

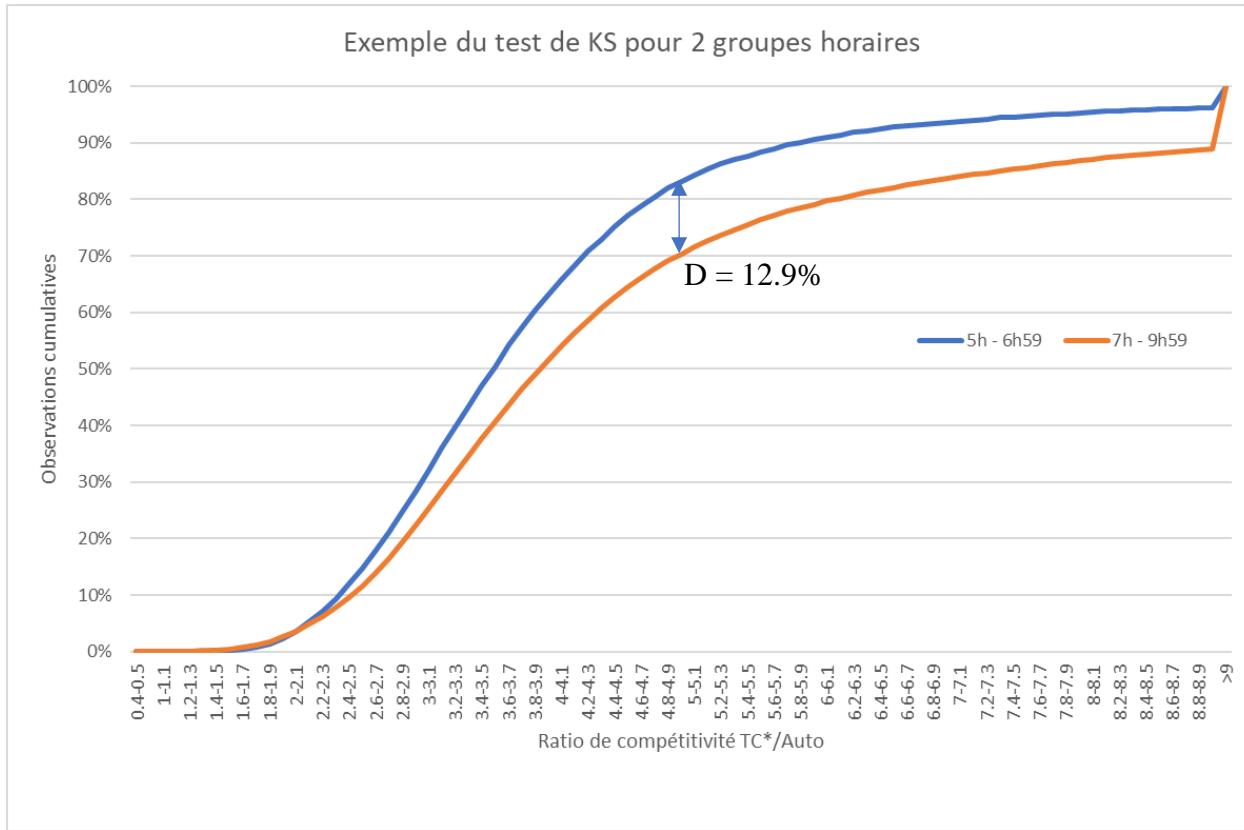


Figure 4-2 : Illustration de la statistique D du test KS comparant les distributions de 5h-6h59 et 7h-9h59

	5	6	7	8	9	10	11	12	13	14	15	16	17
5		2.1%	11.4%	17.0%	16.4%	20.5%	26.3%	24.3%	23.9%	22.7%	22.3%	13.8%	11.4%
6	2.1%		11.8%	17.5%	17.1%	21.0%	27.1%	25.0%	24.6%	23.6%	23.3%	14.8%	12.0%
7	11.4%	11.8%		5.9%	6.1%	10.3%	15.9%	13.7%	13.3%	13.2%	12.5%	5.0%	1.1%
8	17.0%	17.5%	5.9%		1.9%	4.9%	10.1%	7.9%	7.5%	7.8%	7.4%	4.7%	5.8%
9	16.4%	17.1%	6.1%	1.9%		4.4%	10.2%	8.1%	7.7%	7.2%	6.9%	3.4%	5.4%
10	20.5%	21.0%	10.3%	4.9%	4.4%		6.4%	4.7%	3.9%	3.4%	2.8%	6.9%	9.5%
11	26.3%	27.1%	15.9%	10.1%	10.2%	6.4%		2.4%	3.0%	4.7%	4.4%	12.8%	15.2%
12	24.3%	25.0%	13.7%	7.9%	8.1%	4.7%	2.4%		1.9%	3.5%	2.5%	10.8%	13.1%
13	23.9%	24.6%	13.3%	7.5%	7.7%	3.9%	3.0%	1.9%		2.0%	2.0%	10.3%	12.9%
14	22.7%	23.6%	13.2%	7.8%	7.2%	3.4%	4.7%	3.5%	2.0%		1.9%	9.0%	12.4%
15	22.3%	23.3%	12.5%	7.4%	6.9%	2.8%	4.4%	2.5%	2.0%	1.9%		8.8%	11.7%
16	13.8%	14.8%	5.0%	4.7%	3.4%	6.9%	12.8%	10.8%	10.3%	9.0%	8.8%		4.3%
17	11.4%	12.0%	1.1%	5.8%	5.4%	9.5%	15.2%	13.1%	12.9%	12.4%	11.7%	4.3%	

Figure 4-3 : Extrait de la matrice des statistiques D issues du test KS pour chaque paire de FDC

Le critère utilisé pour regrouper deux heures consécutives est défini tel que la distance D maximale est inférieure à 7%. Ce critère, bien que subjectif, est inspiré des travaux de Tessier (2015) et semble approprié pour la segmentation désirée des données. En procédant de la sorte, on obtient la matrice illustrée à la Figure 4-4. Identifiées en gris sont les paires d'heures répondant au critère. On peut également y remarquer des similitudes entre des paires d'heures non consécutives telles que la pointe du matin et la pointe du soir ainsi que les heures du milieu de la journée et celles de soirée.

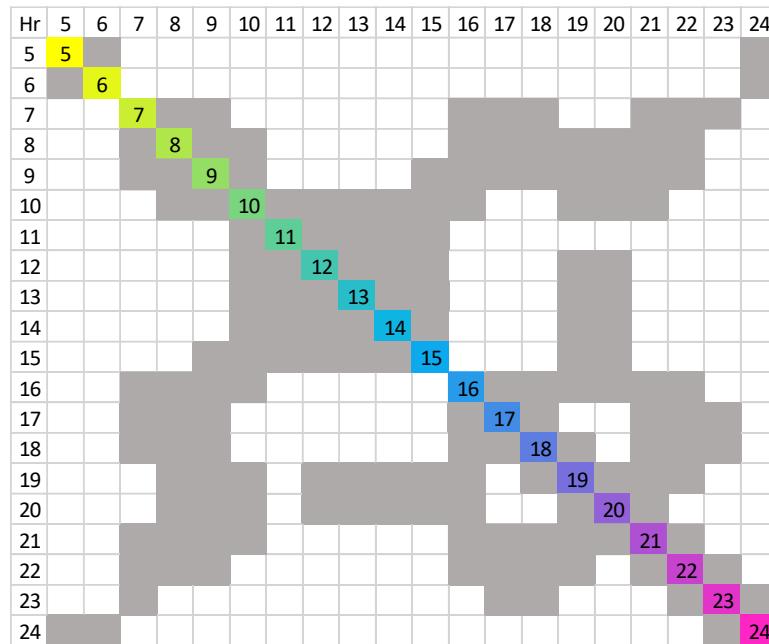


Figure 4-4 : Paires d'heures correspondant au critère  $D < 7\%$

Suite au test KS, six périodes horaires ont été définies et sont présentées au Tableau 4-2. La Figure 4-5 présente la répartition des déplacements durant la journée.

Tableau 4-2 : Groupes horaires issus du test KS

Groupe 1	5h - 6h59 Pré-pointe AM
Groupe 2	7h - 9h59 Pointe AM
Groupe 3	10h - 15h59 Jour
Groupe 4	16h - 18h59 Pointe PM
Groupe 5	19h - 20h59 Post-pointe PM
Groupe 6	21h - 24h59 Soir

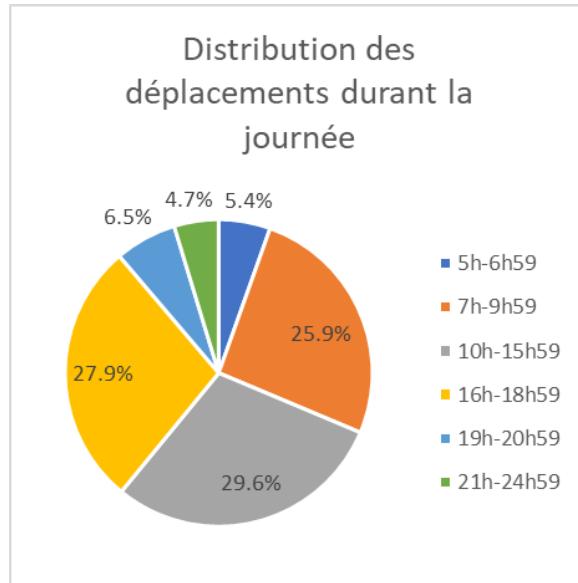


Figure 4-5 : Distribution des déplacements durant la journée

La Figure 4-6 illustre les distributions des ratios TC\* pour chacun des groupes horaires issu du test KS. Elle nous permet de comparer l'évolution de la performance du TC tout au long de la journée. Plus les courbes sont élevées dans la partie gauche du graphique, plus le TC est compétitif.

La période 1, soit la pré-pointe am, se différencie des autres distributions. Elle est particulièrement performante, se démarquant par une courbe beaucoup plus concentrée dans les basses valeurs de ratios.

Outre cette période, la performance du TC est à son maximum durant les pointes am en orange et pm en jaune. On remarque ensuite les queues de distributions illustrant les déplacements où le TC n'est pas une option compétitive.

À la lumière de ce graphique, on confirme que le service TC est plus compétitif en heures de pointe et c'est durant la journée, soit de 10h à 16h où le service TC est le moins performant. Il est toutefois difficile d'expliquer l'ampleur de la démarcation de la pré-pointe. Une piste d'explication serait l'effet combiné du moins grands nombre de déplacements et la présence de lignes express pour les travailleurs.

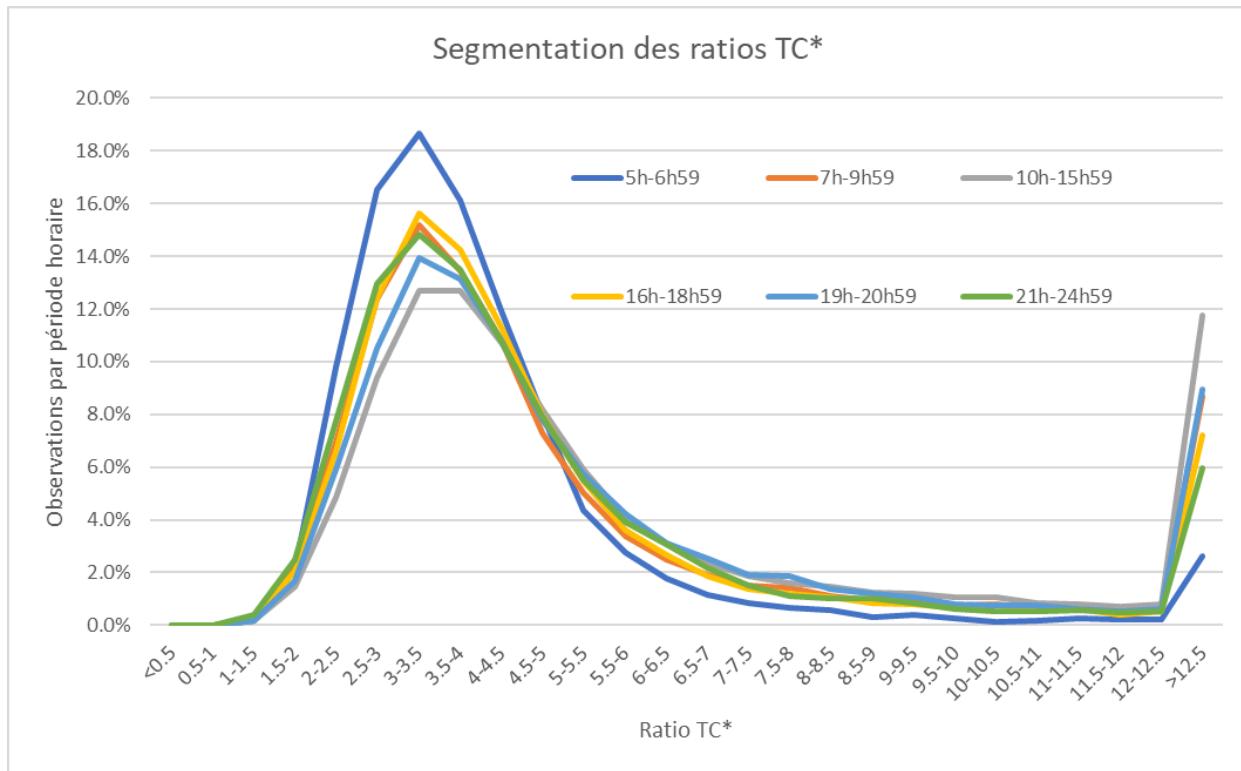


Figure 4-6 : Distribution des Ratios TC\* par groupe horaire

#### 4.4 Compétitivité et distance

L’attractivité modale peut dépendre de la distance du trajet à parcourir. Comme le montre la Figure 4-7, les ratios TC sont extrêmes pour les courts déplacements et ils se stabilisent rapidement, alors que les ratios moyens des modes actifs se dégradent avec la distance. On peut facilement imaginer que les modes actifs sont moins attractifs, voire non-plausibles, pour les longs déplacements. Les ratios les concernant sont tout de même estimés car ils permettent d’assurer l’applicabilité de la méthode sur la base de tous les ratios, et ce, pour tous les déplacements, nonobstant leurs propriétés. De manière similaire, les temps d'accès et d'attente inhérents au TC sont dissuasifs pour les courts déplacements. Basées sur les travaux de distance seuil de Demers et Morency (2009), les distances seuils (D80) sont présentées et un concept de distance seuil minimal pour le TC est proposé.

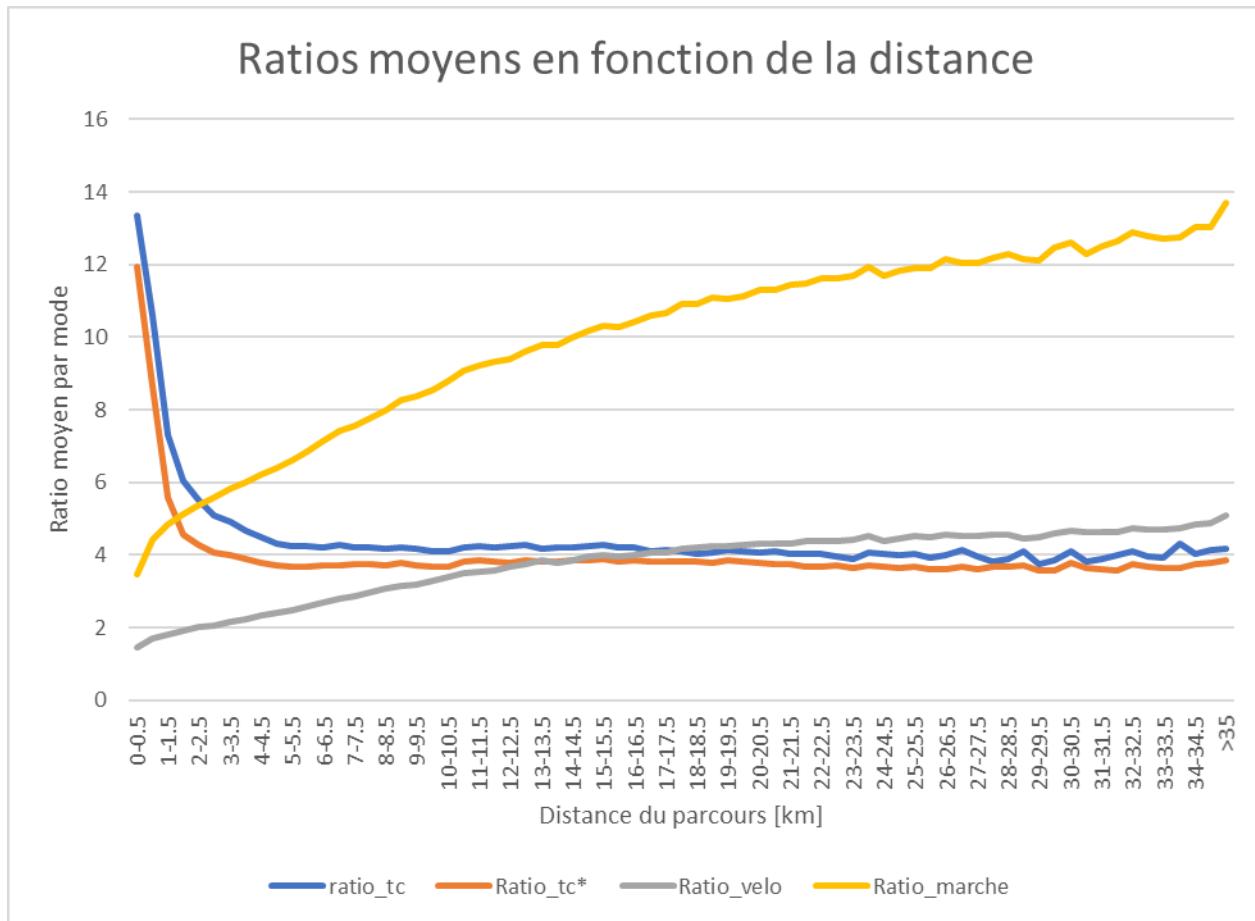


Figure 4-7 : Ratios moyens en fonction de la distance

#### 4.4.1 Distances seuils

Tel que mentionné ci-dessus, la distance à parcourir peut influencer le choix modal. Le transport en commun ne semble pas être une option pour les courts déplacements, mais qu'en est-il des autres modes? La Figure 4-8 ci-dessous montre les FDC des distances des déplacements selon le mode, alors que la Figure 4-9 montre la répartition modale des déplacements en fonction de la distance.

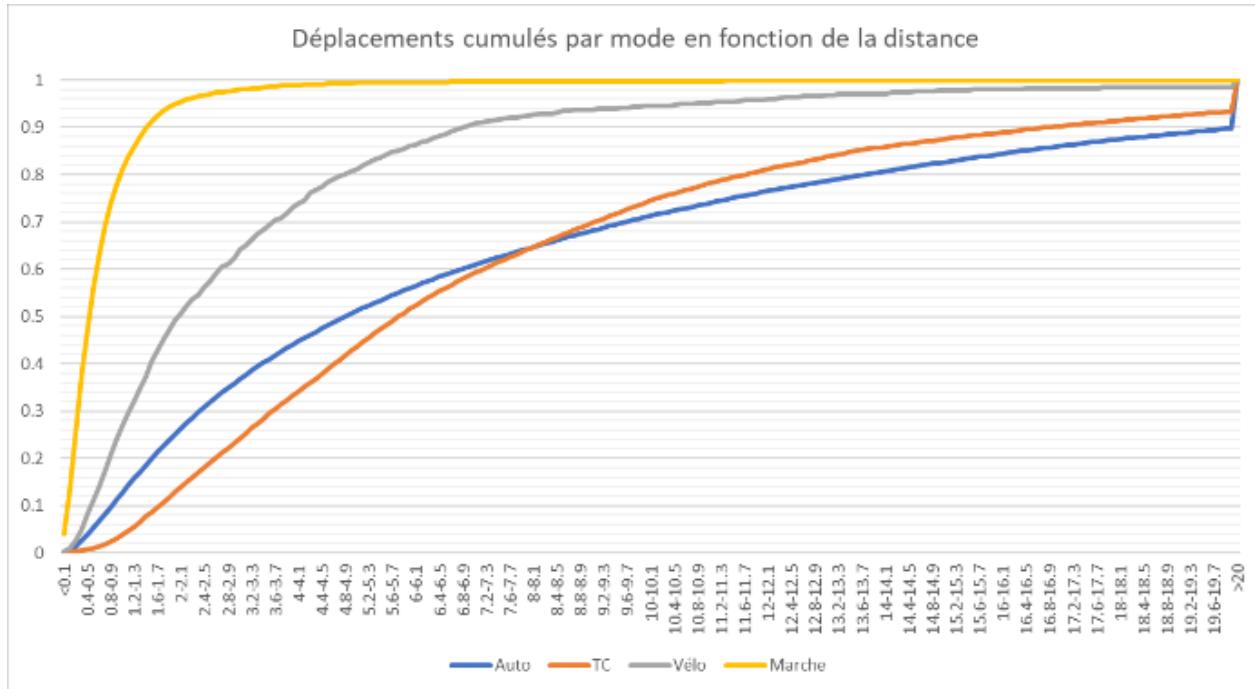


Figure 4-8 : Déplacements cumulés par mode en fonction de la distance

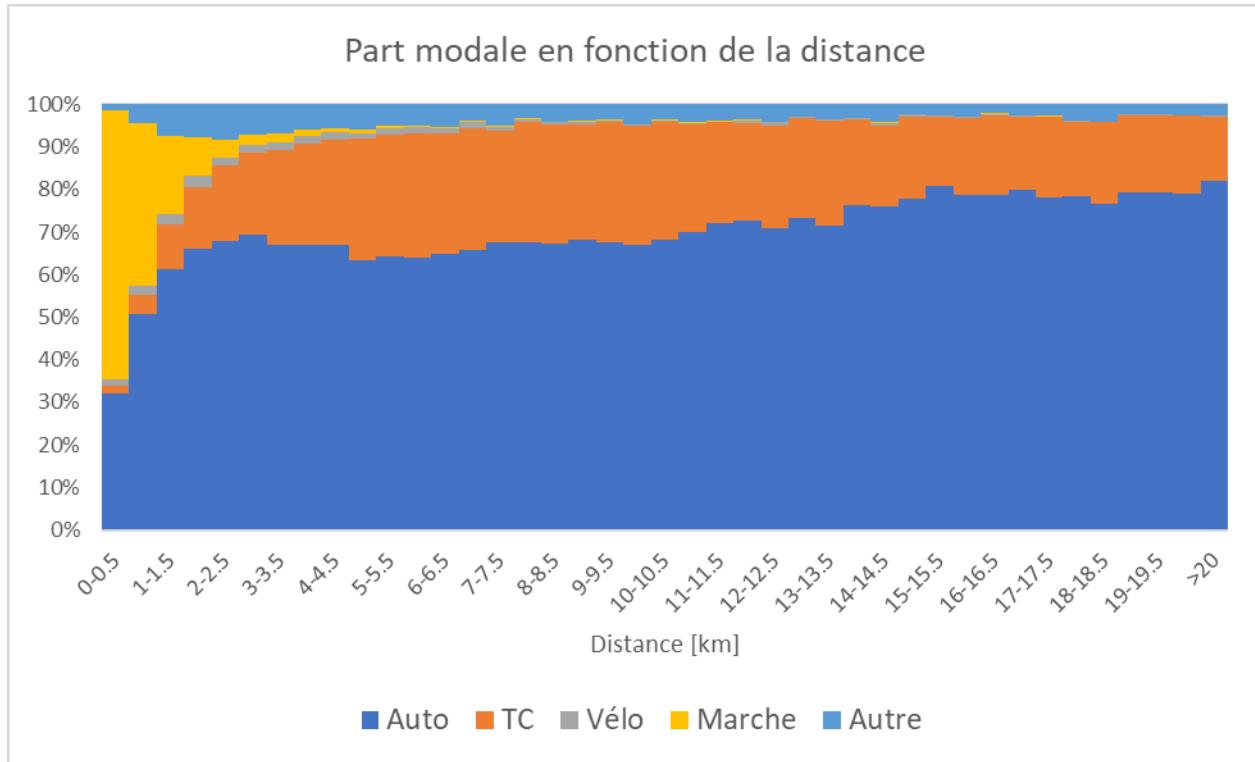


Figure 4-9 : Part modale en fonction de la distance

Le concept de distance seuil, tel que proposé par Demers et Morency (2009), est défini comme la distance « qui cumule 80 % des déplacements faits à pied par un segment démographique particulier ». Des analyses par groupe démographique ont aussi été reprises et présentées par Martel-Poliquin (2012). La figure 4-10 présente les distances seuil pour la marche par groupe démographique. Les distances seuil pour les adultes (15 ans et plus) sont de 1,15 km pour les femmes et de 1,25 km pour les hommes.

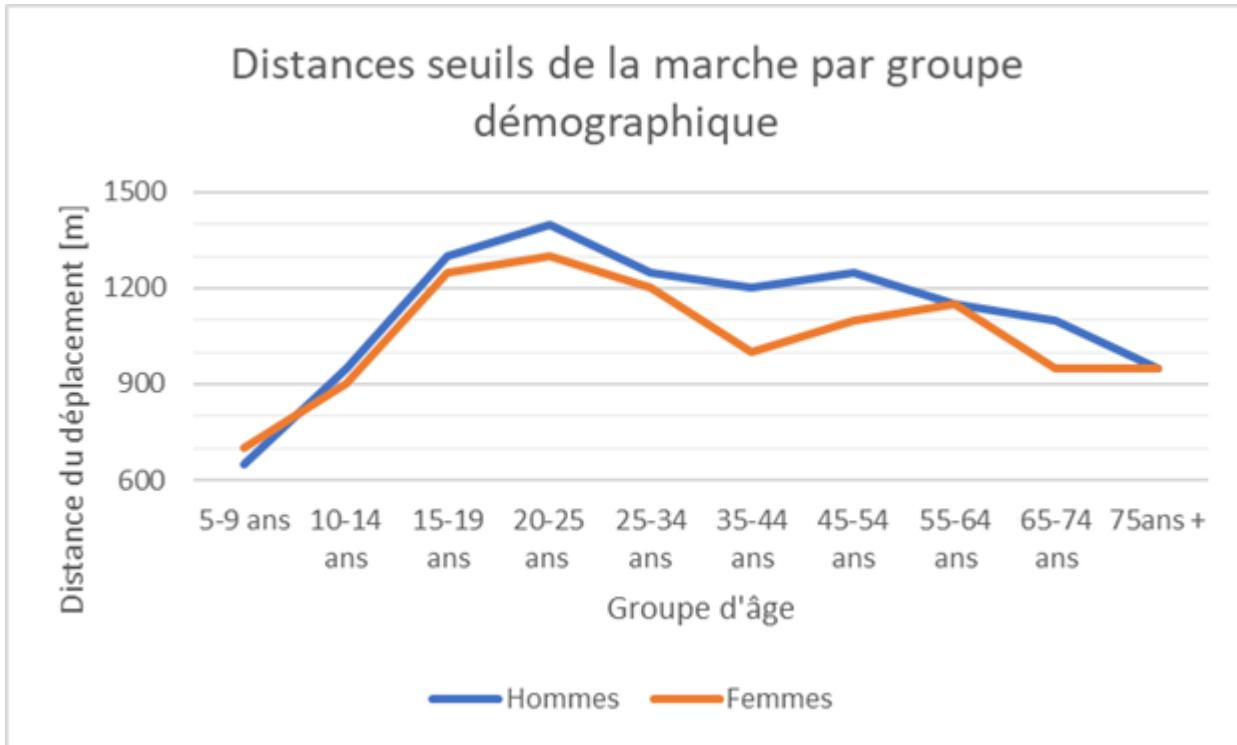


Figure 4-10 : Distances seuil pour la marche par groupe démographique

De manière similaire, la Figure 4-11 présente les distances seuils du vélo. Elles sont en moyenne, pour les adultes, de 4 km pour les femmes et de 5 km pour les hommes.

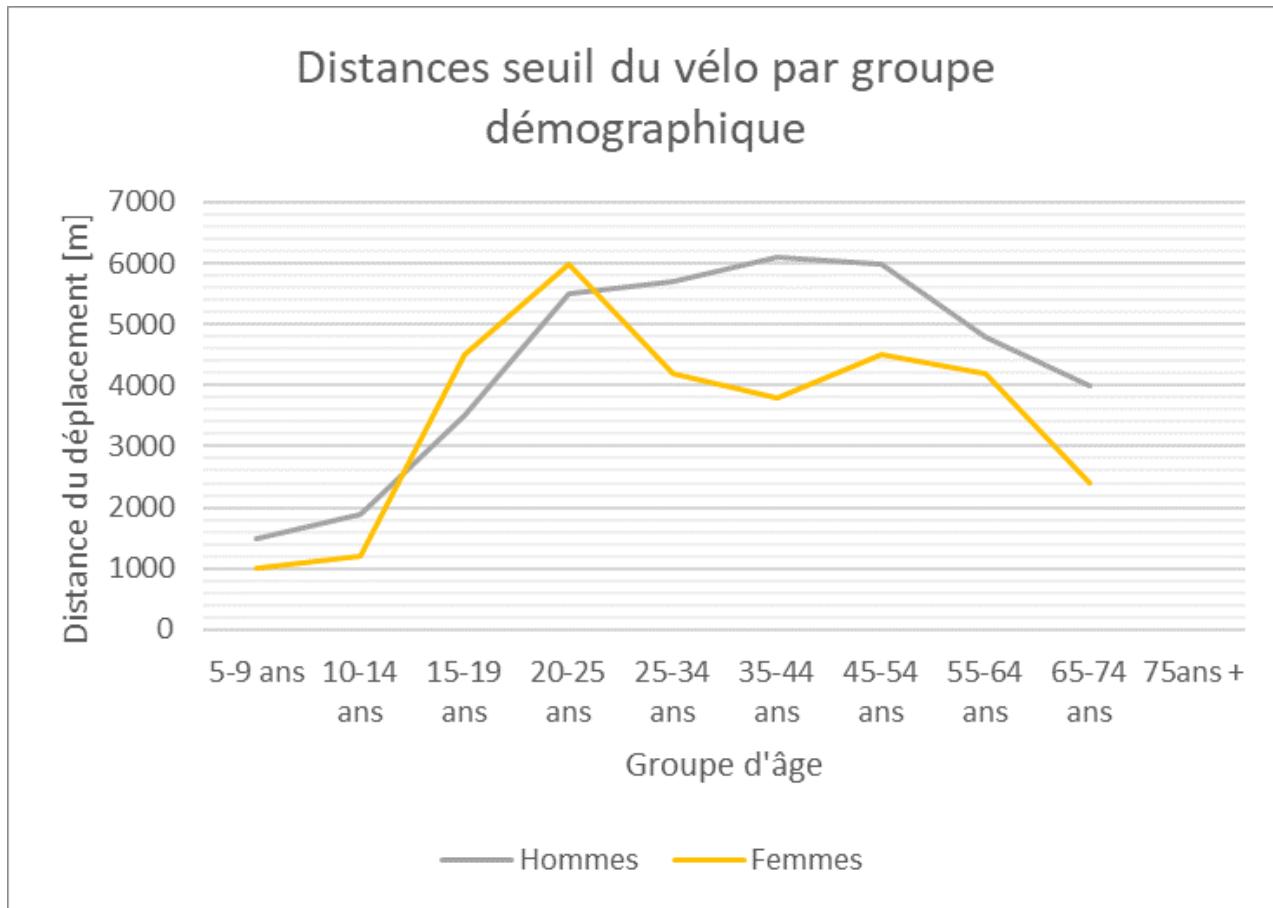


Figure 4-11: Distances seuil pour le vélo par groupe démographique

La Figure 4-7 met en évidence les ratios très élevés du TC pour les petites distances. Ces ratios s'expliquent par la proportion élevée du temps d'accès à l'origine, à la destination, ainsi que du temps d'attente par rapport au temps du déplacement lui-même. Pour ces distances, souvent à l'intérieur de la distance seuil de la marche, très peu de gens choisissent le TC pour faire leur déplacement tel qu'illustré sur la figure 4-8.

Serait-il approprié alors de définir une « distance seuil minimale » pour le TC? Ce seuil correspondrait à la distance minimale à parcourir pour que le TC devienne compétitif comparativement à la marche.

Pour identifier ce nouveau seuil, la définition suivante est suggérée : « La distance à partir de laquelle le ratio moyen marche/TC\* devient plus grand que 1. »

L'identification de ce seuil passe par la création d'un nouveau ratio de compétitivité, le ratio marche/TC\*. L'évolution de ce dernier en fonction de la distance est illustrée en Figure 4-12. On

y voit que la distance à laquelle, en moyenne, le temps de parcours à la marche équivaut à celui du transport en commun est entre 1200 et 1300 m.

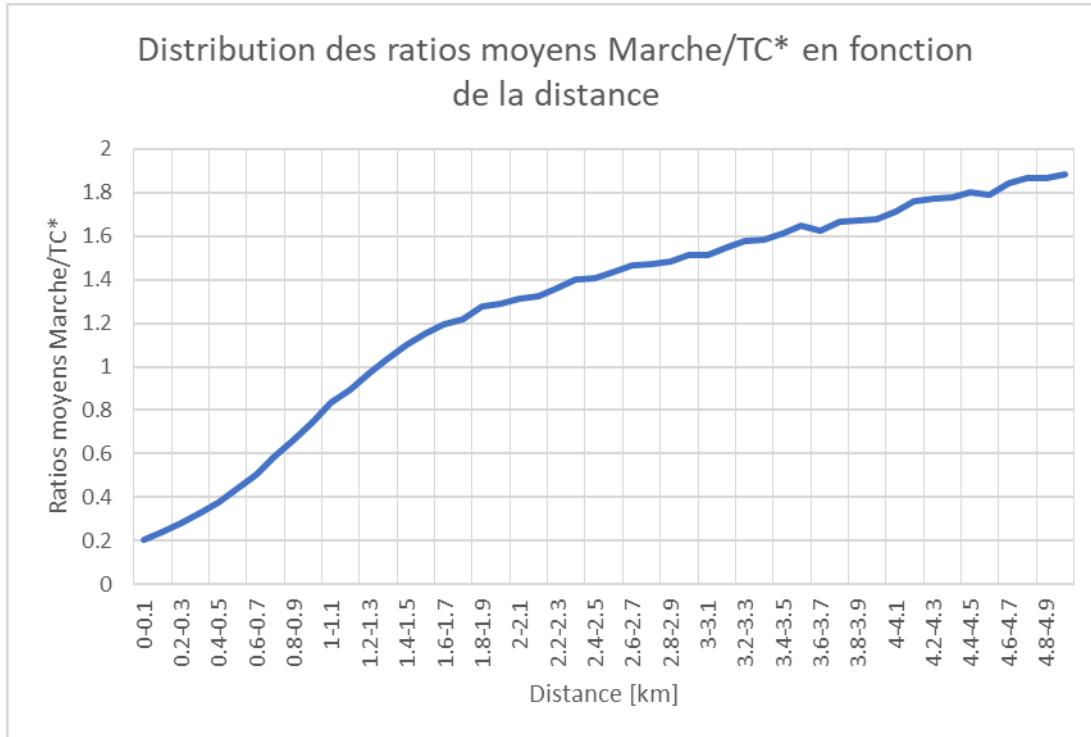


Figure 4-12 : Ratios moyens marche/TC\* en fonction de la distance réseau

La distance seuil minimum pour le TC serait donc de 1300 m. La Figure 4-13 illustre la distribution des déplacements faits en TC par rapport au ratio marche/TC\*, où près de 6% des déplacements fait en TC avaient un ratio marche/TC\* inférieur à un, ratio correspondant en moyenne à une distance inférieure au seuil de 1,3 km.

Une distance seuil de 1,3 km peut toutefois sembler élevée en milieu urbain. Plusieurs facteurs ont été négligés et pourraient venir affecter cette distance seuil. Tout d'abord, la pénibilité associée à la marche aurait comme effet de réduire la distance seuil. Aussi, contrairement à la marche, le temps de parcours TC est dépendant d'une offre extérieure et il existe un grand éventail de densité d'offre dans la région. Il serait donc plus pertinent de calculer une distance seuil locale pour le TC qui serait lié au niveau d'offre observé.

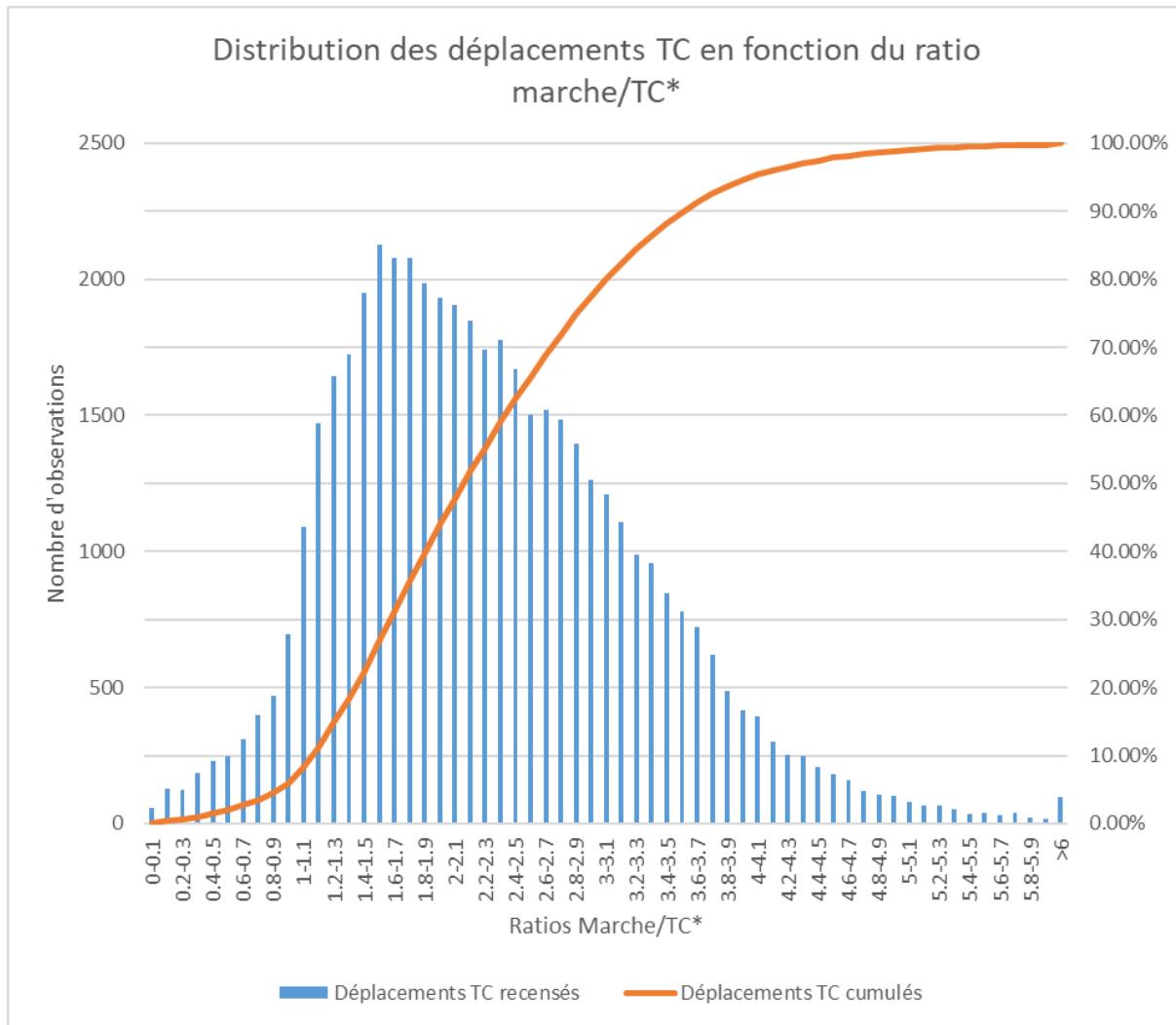


Figure 4-13 : Distribution des déplacements TC en fonction du ratio marche/TC\*

## 4.5 Regroupement des vecteurs de compétitivité

Autant l'analyse des regroupements des vecteurs de compétitivité sur une base horaire est intéressante pour étudier l'évolution temporelle de l'attractivité du système de TC, autant l'analyse par clustering est intéressante pour identifier des familles de déplacements. Ces familles pourront servir dans la définition d'un modèle de choix modal. La méthode utilisée est la méthode de clustering à deux étapes, c'est-à-dire que dans un premier temps un clustering en k-moyennes regroupe la base de données en 50 groupes, et ce sont les centroïdes de chacun de ces groupes qui seront utilisés lors du clustering hiérarchique.

### 4.5.1 Définitions des vecteurs

Pour cette analyse, des vecteurs à trois dimensions sont utilisés. Chaque valeur est un ratio de temps de parcours (TP) tel que montré par le Tableau 4-3.

Tableau 4-3 : Définition des vecteurs pour clustering

Ipère	Ratio $TC^*/auto$	Ratio vélo/auto	Ratio marche/auto
1	3.478	3.263	9.227
2	3.594	5.063	14.028
...	...	...	...
n	$TP_{TC^*}/TP_{auto}$	$TP_{vel}/TP_{auto}$	$TP_{mar}/TP_{auto}$

### 4.5.2 Résultat du regroupement

Dans un premier temps, l'ensemble de la base de données a été regroupé en 50 groupes avec la méthode de regroupement des k-moyennes. Ensuite, la seconde étape de regroupement s'est faite par regroupement hiérarchique. La Figure 4-14 représente la somme de la différence des carrés dans chaque groupe pour différentes quantités de groupes finaux. Ce graphique aide à définir combien de groupes il est pertinent d'analyser. Pour une analyse pertinente, l'intérêt est au point d'infexion où l'impact de l'ajout d'un groupe devient moins important. Ainsi, la Figure 4-14 montre la somme quadratique des erreurs dans chaque groupe, c'est-à-dire le carré de la distance entre chaque observation et la moyenne du groupe. Une première infexion aux environs de 5 groupes est intéressante mais à l'aide du dendrogramme de la Figure 4-15 et des liens de regroupement hiérarchique représentés, un regroupement en 9 groupes est choisi.

### Somme des carrés résiduels en fonction du nombre de groupes

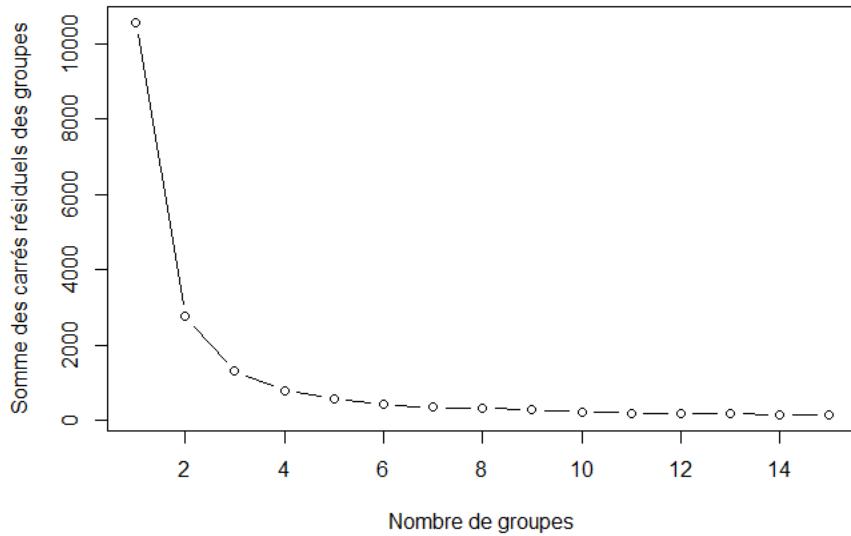


Figure 4-14 : Sommes des carrés résiduels par groupe

### Dendrogramme complet, division en 9 groupes

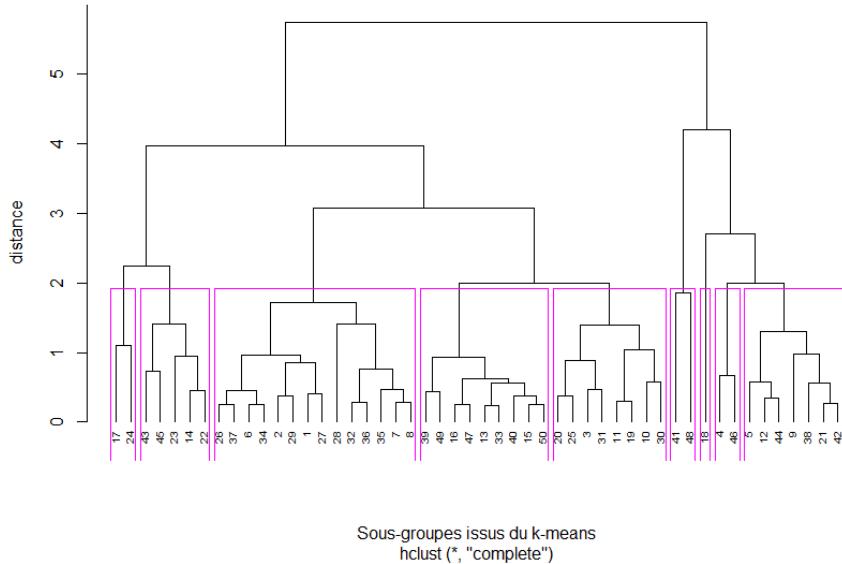


Figure 4-15 : Dendrogramme complet avec 9 clusters

Le choix d'une division en 9 groupes se justifie par le désir d'avoir un niveau de résolution assez fin pour isoler les petits clusters afin de pouvoir les analyser de manière indépendante. Le Tableau 4-4 présente quelques caractéristiques de clusters obtenus. Aux fins d'identification dans la Figure 4-15, les groupes sont numérotés en ordre, de la gauche vers la droite.

Tableau 4-4 : Caractéristiques des 9 clusters

Groupes	Observations		Vecteurs de Ratios			Distance [km]		Part modale				
	Nombre	Pourcentage	TC*	Vélo	Marche	Moyenne	Écart-type	Auto	TC	Vélo	Marche	Autre
1	1532	0.6%	16.66	2.86	7.39	2.47	4.41	65.4%	2.1%	2.0%	26.9%	3.6%
2	11550	4.4%	17.00	1.65	4.17	0.47	0.52	37.2%	1.6%	1.8%	57.3%	2.1%
3	99123	38.2%	4.02	2.91	7.77	8.24	5.51	70.5%	22.2%	0.9%	1.5%	4.9%
4	81518	31.4%	3.66	1.85	4.95	3.02	1.99	59.7%	23.0%	2.3%	8.3%	6.7%
5	27125	10.4%	9.12	1.67	4.29	0.95	1.02	52.5%	3.4%	2.2%	36.5%	5.4%
6	495	0.2%	15.42	5.29	14.06	15.31	12.53	74.9%	5.3%	0.4%	10.7%	8.7%
7	1147	0.4%	3.76	8.21	10.88	14.99	4.88	81.2%	16.4%	0.0%	0.0%	2.4%
8	2469	1.0%	5.46	5.80	15.43	26.69	13.31	92.1%	4.9%	0.4%	0.6%	0.0%
9	34818	13.4%	4.23	4.54	11.91	19.63	9.37	81.5%	15.5%	0.2%	0.1%	0.7%

On remarque que dans les deux premiers groupes qui regroupent 5% des données, les ratios TC sont très grands. Avec des ratios moyens de l'ordre de 17, le TC n'est tout simplement pas une option pour ces déplacements. Toutefois, pour ces deux groupes, les distances moyennes sont courtes, ce qui rend les modes actifs attrayants.

Le troisième groupe est celui qui contient le plus d'observation avec 38,2% des déplacements. Il contient les déplacements de moyennes-longues distances où l'offre TC est compétitive. Les distances étant hors de portées de la marche, la répartition modale est principalement divisée entre l'auto et le TC.

Le quatrième groupe, également important avec 31,4% des déplacements, contient les déplacements de moyennes distances où l'offre TC est compétitive. Comparé au groupe précédent, la part modale du TC est toujours très forte à 23%, la marche bénéficiant d'un transfert modal depuis l'automobile de près de 7 points de pourcentage. La distance moyenne de 3,02 km demeure toutefois un obstacle à la pratique de la marche.

Le cinquième groupe contient les déplacements de courtes distances où l'offre de TC est présente. En concordance avec les tendances identifiées précédemment, le TC n'est pas attrayant pour les déplacements d'une distance de moins d'un kilomètre. En effet, comparé au groupe précédent, il y a un transfert modal majeur du TC vers la marche. Plusieurs automobilistes délaissent également leur voiture et choisissent de marcher.

Les derniers groupes contiennent les déplacements de longues distances. Pour ces déplacements, sans surprise, l'utilisation de l'automobile est prédominante. Le groupe 6 semble bien marginal avec 0,2% des déplacements, une part modale de 10,7% pour la marche est bien surprenante pour des déplacements de longueur moyenne de 15 km. Toutefois, avec un écart-type de 12,53 kilomètre, les distances parcourues au cours des déplacements présents dans ce groupes sont très variables. Une analyse plus approfondie des déplacements effectués à pied montre effectivement que ces déplacements ont une distance moyenne de 0,53 km. Les groupes 7 et 9 contiennent les déplacements où le TC est compétitif et compte pour environ 16% des déplacements. Finalement, le groupe 8 contient les plus longs déplacements qui sont réalisés quasi-exclusivement en voiture.

L'analyse de corrélation entre les parts modales observées, les ratios et distances moyennes par groupe du Tableau 4-4 est présentée dans le Tableau 4-5. Malgré l'influence prononcée de la distance sur l'utilisation de l'automobile et des modes actifs, l'attractivité du TC n'est pas corrélée avec ce facteur. L'augmentation du ratio TC bénéficie principalement à la marche qui est une option pour les courtes distances. Le coefficient de corrélation de -0.797 confirme que plus le ratio TC\* est élevé, moins la part modale du TC est élevée. Les ratios vélo et marche sont principalement liés à la distance à parcourir, ainsi plus ils sont grands, plus la voiture est attractive, le TC restant peu affecté. Finalement, l'utilisation des autres modes ne semble que peu corrélée aux divers ratios ou à la distance à parcourir.

Tableau 4-5 : Coefficients de corrélation linéaire

	Ratio TC*/auto	Ratio vélo/auto	Ratio marche/auto	Distance
Auto	-0.516	0.797	0.893	0.905
TC	-0.797	0.115	-0.040	0.093
Vélo	0.358	-0.882	-0.867	-0.877
Marche	0.739	-0.644	-0.677	-0.723
Autre	0.146	-0.230	-0.060	-0.241

## 4.6 Portrait global des ratios de compétitivité

Une des caractéristiques intéressantes des ratios de compétitivité est le dénominateur commun utilisé, ce qui permet une comparaison aisée entre les modes. Ainsi, la distribution de ces ratios par mode de transport illustrée à la Figure 4-16 donne un bon aperçu de la performance relative de chacun, pour les déplacements observés dans l'enquête (après filtrage).

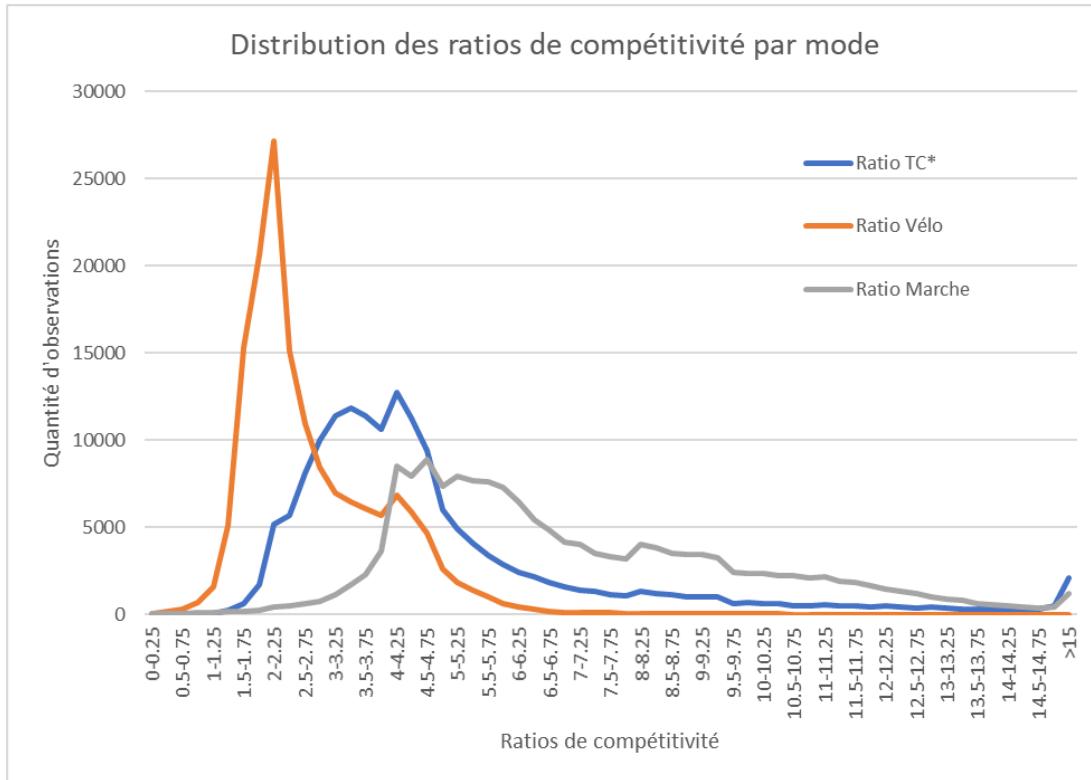


Figure 4-16 : Distribution des ratios de compétitivité par mode

Tableau 4-6 : Valeur seuils des quintiles de ratios de compétitivité

Quintile	TC*	Vélo	Marche
1	0-3	0-1.75	0-4.65
2	3-3.65	1.75-2.1	4.65-5.65
3	3.65-4.45	2.1-2.6	5.65-7
4	4.45-6.2	2.6-3.6	7-9.5
5	6.2+	3.6+	9.5+

À la lumière de la Figure 4-16 et du Tableau 4-6 présentant les seuils des quintiles des ratios pour chacun des modes, il est possible d'affirmer que le vélo est particulièrement performant. Rappelons que l'hypothèse posée pour le calcul du temps de parcours en vélo néglige les temps d'accès au vélo, la congestion routière, les conditions météorologiques, le dénivelé présent lors du parcours, les caractéristiques de la personne et suppose un accès aisément à ce mode en tout temps. Avec un tel niveau de performance, il peut être surprenant de ne pas avoir une plus grande part modale pour ce mode. D'autres facteurs tel que l'effort physique nécessaire, des enjeux de sécurité ou hygiéniques pourraient probablement expliquer un tel phénomène.

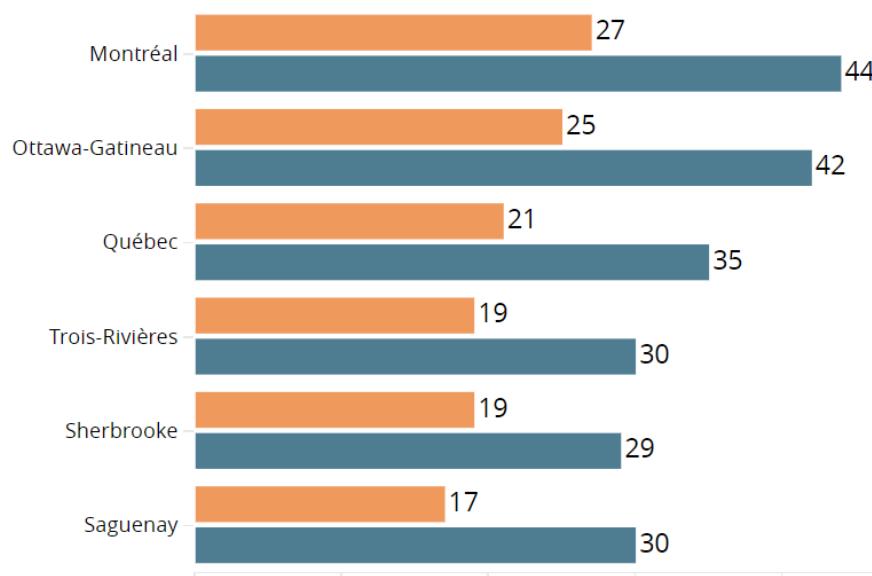
Avec des valeurs médianes se situant entre 3,5 et 4,5, la performance du transport en commun semble grandement sous-évaluée. Étonnamment, seulement 20% des déplacements TC ont un ratio inférieur à 3, ce qui semble bien peu considérant la densité d'offre de service de métro de Montréal qui est particulièrement efficace en période de pointe. Cette sous-estimation des ratios est probablement le résultat combiné d'une sous-estimation des temps de parcours auto en congestion et d'une surestimation des temps de parcours TC. Mentionnons que le temps d'accès à l'automobile et de recherche d'un espace de stationnement n'ont pas été considérés.

Dans un article paru le 4 décembre 2017 dans *Le Devoir*, on y cite des données de Statistique Canada 2016 dans lesquelles sont comparés les temps de parcours moyens pour se rendre au travail en voiture et en TC pour différentes villes du Québec. La Figure 4-17 illustre que les temps de parcours domicile-travail pour la ville de Montréal sont de 27 minutes en voiture et 44 minutes en TC, ce qui donne un ratio de 1,63, loin du ratio médian de 4 observé plus tôt. Cet article, basé sur des données de Statistique Canada, laisse supposer que les ratios issus des temps simulés ne sont pas représentatifs de la compétitivité réelle du TC.

### Temps de trajet domicile-travail

En minutes

Voiture Transport en commun



Source: Statistique Canada



Figure 4-17 : Temps de trajet domicile-trajet (*Le Devoir*, 4 décembre 2017)

Finalement, la marche ne semble pas du tout attrayante comparée à l'automobile, à s'en fier à la Figure 4-16. Toutefois, ce modèle de simulation néglige les temps d'accès à l'automobile et le temps requis pour stationner la voiture rendu à destination. Pour les courts déplacements, ce temps pourrait être non-négligeable et pourrait influencer grandement la mesure de la performance de la marche. Il semble aussi y avoir une sous-estimation systématique du temps de déplacement automobile sous congestion.

## 4.7 Analyse d'autocorrélation spatiale par période horaire

La demande de transport variant au cours de la journée, les sociétés de transport modulent le niveau de service afin d'optimiser leurs coûts d'opération. Il est intuitif d'augmenter le niveau de service et la capacité des bus en heures de pointe, périodes où la majorité des travailleurs se déplacent. L'afflux d'automobiles dans ces périodes et la présence de voies réservées aux autobus pourraient affecter les ratios de manière plus significative à un moment de la journée qu'à un autre. Ainsi, afin d'identifier la présence et la localisation d'un tel phénomène, les indices de Moran ont été calculés pour chacun des modes pour chacune des périodes horaires.

La Figure 4-18 montre que la période #2, soit de 7h à 9h59, est celle qui montre les plus grands signes d'autocorrélation spatiale pour l'ensemble des modes. Pour cette analyse, les variables d'intérêt étaient les différents ratios de compétitivité selon la position de l'origine du déplacement. Cette section présentera l'analyse spatiale des ratios de cette période horaire, les autres présentant peu d'intérêt.

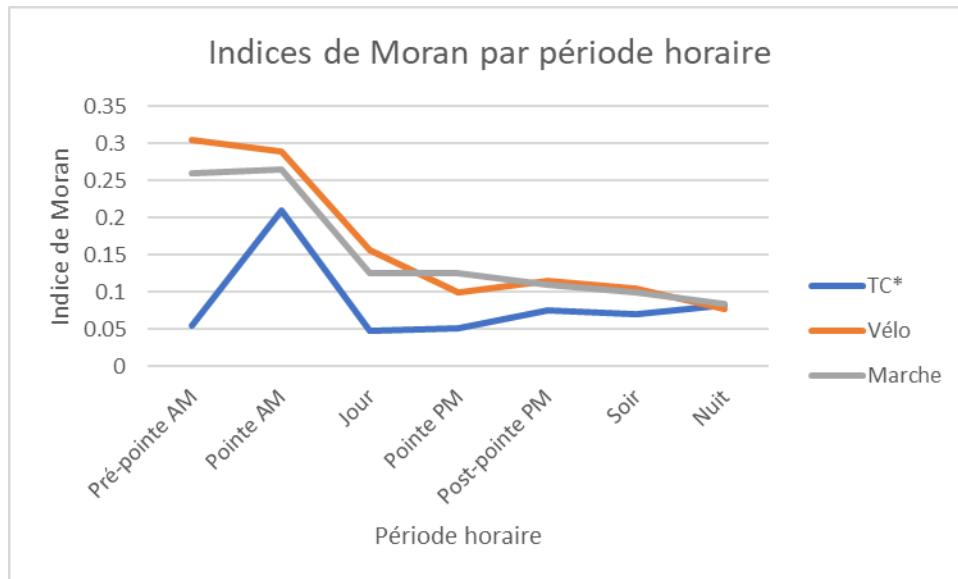


Figure 4-18 : Indices de Moran par période horaire

La Figure 4-19 illustre la performance du TC avec une prédominance de secteur bleus, c'est-à-dire avec des regroupements de secteurs avec des ratios TC\* significativement inférieurs à leurs voisins. La majorité des municipalités des couronnes, dont Sainte-Anne-des-Plaines, Saint-Lin, l'Épiphanie, Lavaltrie, Otterburn Park, Carignan et Beauharnois semblent bien performer, tandis qu'on observe l'inverse à cette période de la journée dans les municipalités de Les Coteaux, Les Cèdres, St-Jérôme et St-Jean-sur-Richelieu. Quelques secteurs bleus longeant les lignes de métro peuvent également être observés. Afin de mieux voir cette région, une carte des ratios TC moyens aux abords des lignes de TC est disponible en annexe B.

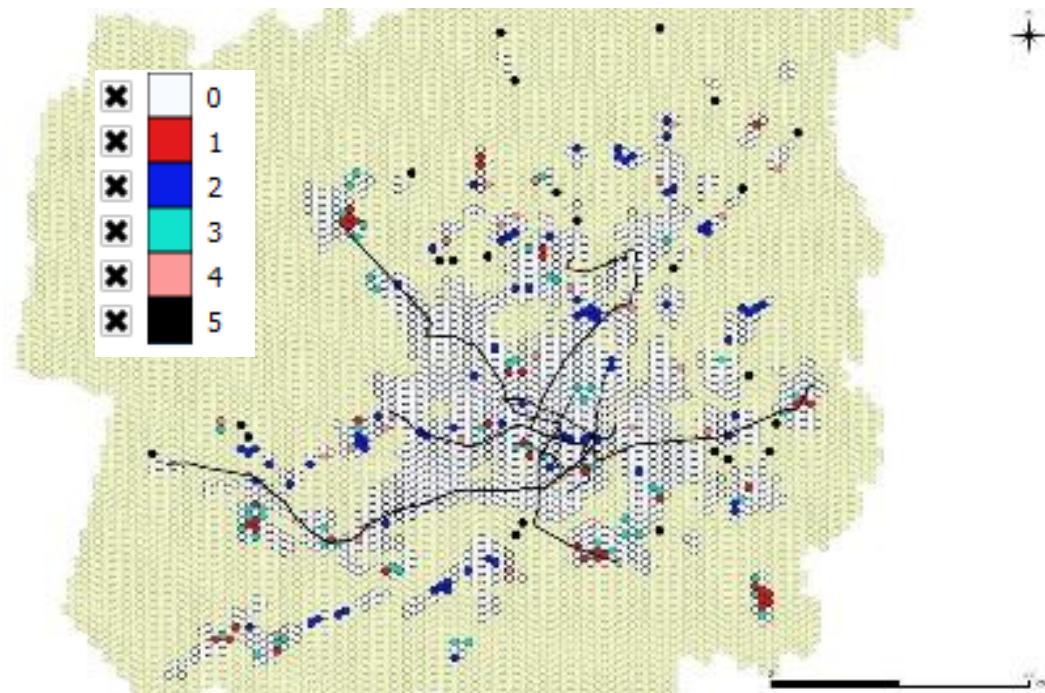


Figure 4-19 : Clusters de Moran TC\*, 7h am à 9h59

Les Figure 4-20 et Figure 4-21 montrent que c'est l'île de Montréal qui est la plus favorable aux modes actifs. Les seules régions de l'île qui ne sont pas identifiées en bleu sont l'ouest de l'île, Pointe-aux-Trembles et la région de l'aéroport. Les régions les moins favorables aux cyclistes sont situées principalement sur la rive sud, en bordure de l'île.

Rappelons que les zones blanches (0) sont les zones non significatives, les zones rouges (1) sont les valeurs élevées adjacentes à d'autres valeurs élevées, les zones bleues (2) sont les valeurs basses adjacentes à d'autres valeurs basses, les zones cyans (3) sont les valeurs basses adjacentes à des valeurs élevées, les zones roses (4) sont les valeurs élevées adjacentes à des valeurs basses et finalement, les zones noires (5) sont les zones isolées.

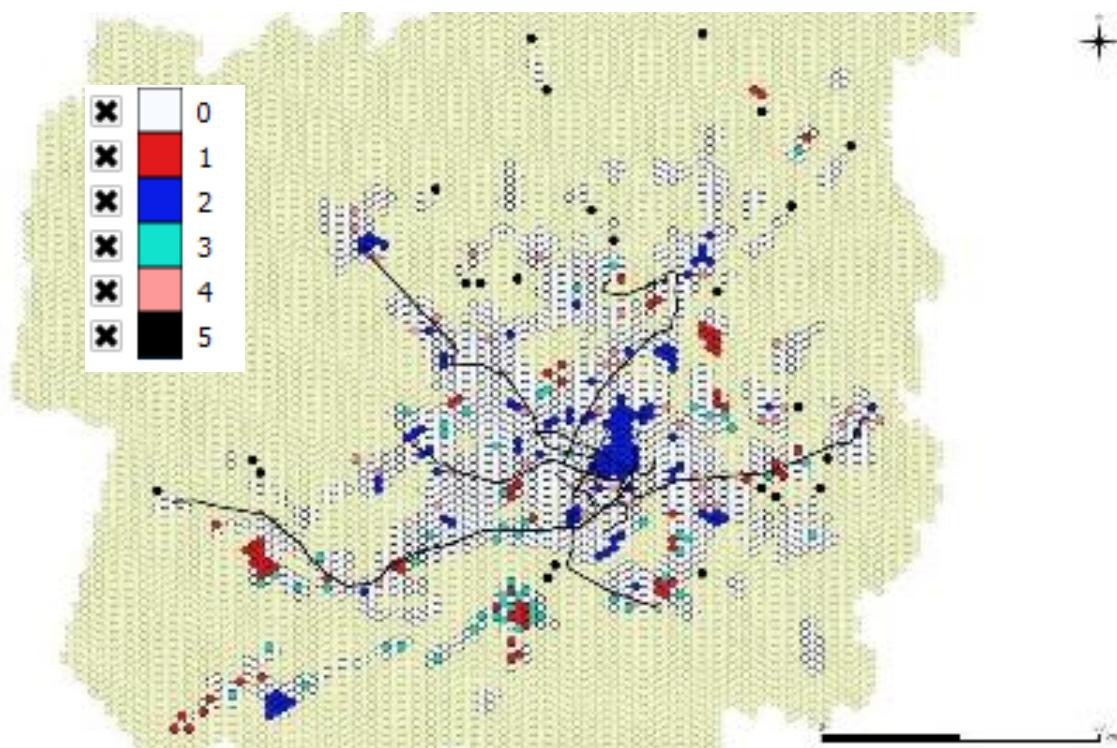


Figure 4-20 : Clusters de Moran Vélo, 7h am à 9h59

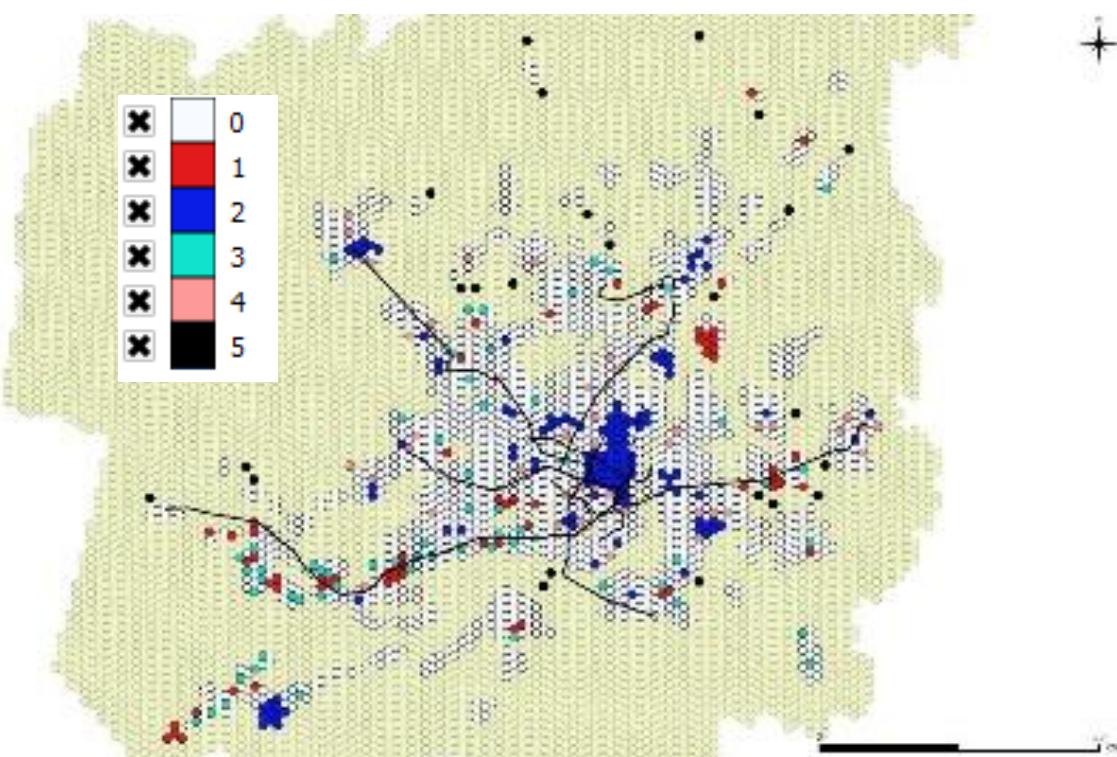


Figure 4-21 : Clusters de Moran Marche, 7h am à 9h59

## CHAPITRE 5 MODÈLES DE CHOIX MODAL

La modélisation modale sert à prédire le choix modal qu'une personne fera pour effectuer son déplacement. Deux méthodes de modélisation du choix modal sont proposées dans ce mémoire afin de prévoir les habitudes modales sur une population projetée. La première est un arbre de décision alors que la seconde est un modèle classique de type logit multinomial.

### 5.1 Arbre de décision

Un premier arbre de décision avec l'ensemble des données a été créé, soit les variables suivantes : le genre, le groupe d'âge, le niveau de revenu, le motif du déplacement, la possession d'un permis de conduire, la possession d'un titre mensuel de transport, le niveau de motorisation du ménage, l'heure de départ, la distance et les différents ratios de compétitivité. La variable à prédire est le choix modal tiré de l'ensemble des alternatives suivant : auto conducteur, auto passager, TC, vélo, marche et autres modes. L'arbre est composé de cellules, chacune contenant l'information reliée aux déplacements inclus dans le groupe. Dans chacune de celles-ci, l'information est présentée telle qu'illustrée en Figure 5-1. La première ligne identifie le mode principal (ou dominant) qui est aussi identifié par la couleur de la feuille, suivi des indices de mixité. Ces indices présentent la répartition des modes déclarés associés aux déplacements regroupés dans la feuille.

De gauche à droite sont identifiées les parts modales de l'auto conducteur, de l'auto passager, des autres modes, de la marche, du TC et du vélo. Finalement, le pourcentage indiqué représente la part des déplacements qui sont regroupés dans chacune des feuilles.

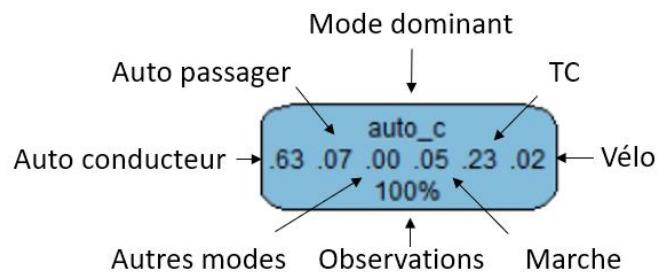


Figure 5-1 : Contenu des feuilles des arbres de décision

Avec l'aide de la bibliothèque « rpart » du logiciel RStudio, intitulée « Recursive Partitioning and Regression Trees », un premier arbre a été calculé et est présenté en Figure 5-2. Cet arbre, constitué de six tests de variables et de sept feuilles arrive à classer 255 969 déplacements avec un taux

d'erreur de 49,53%. Ce taux reflète la proportion des déplacements dont le mode choisi n'est pas le mode majoritaire de la feuille.

Les principales variables qui sont mises en évidence par les tests sont la possession du permis de conduire (percond), la possession d'un titre mensuel de TC (passeTC). Ensuite, pour les non-détenteurs de permis de conduire ni de titre mensuel, la distance, le niveau de motorisation et le motif deviennent les facteurs de division.

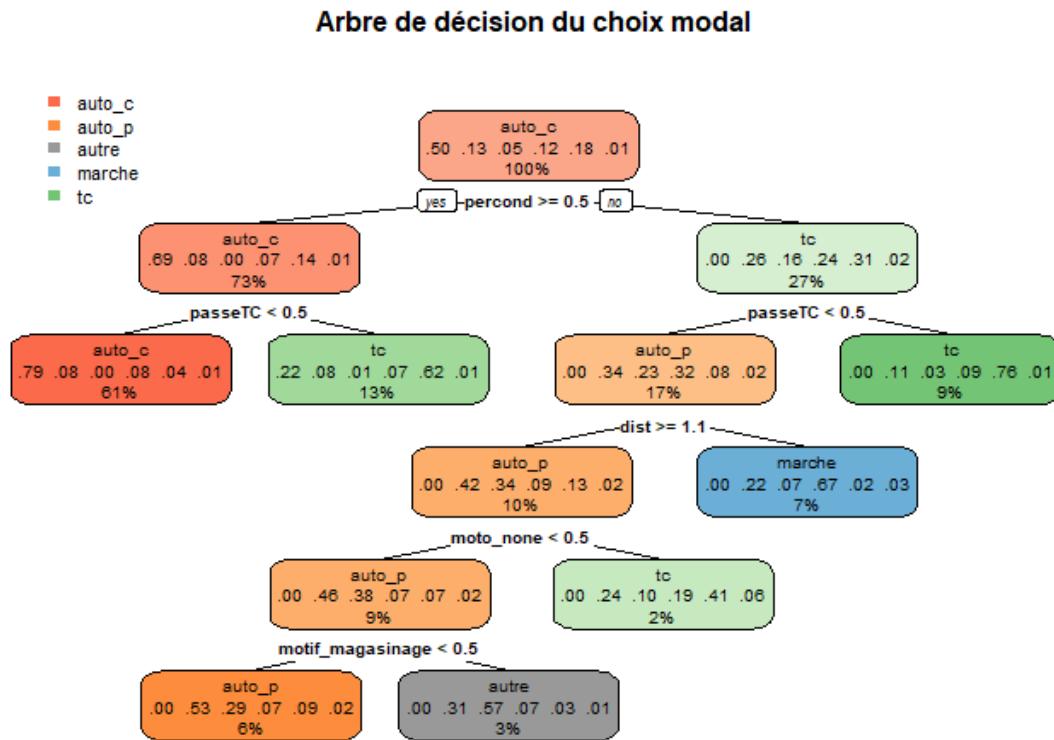


Figure 5-2 : Arbre de décision, tous les déplacements

L'état de possession du permis de conduire ou du titre mensuel TC, ainsi que les niveaux de motorisation des ménages peuvent être la conséquence de l'offre de transport existante. Suivant une évolution de l'offre TC, certains ménages pourraient choisir de délaisser leur(s) voiture(s) et passer au TC, ou l'inverse. Ainsi, il peut être difficile d'identifier si ces variables sont la cause ou la conséquence du choix modal.

Afin d'avoir un modèle excluant toutes les variables dépendantes de l'état de possession d'un objet et ainsi éviter ce problème potentiel, un autre modèle sera réalisé. Ce modèle exclura les variables

suivantes : la possession du permis de conduire, du titre mensuel de TC, ainsi que le niveau de motorisation du ménage. Il deviendra ainsi plus facile à utiliser dans une démarche prédictive.

### 5.1.1 Non-détenteurs de permis de conduire

Pour ce modèle, seuls les non-détenteurs de permis de conduire seront utilisés. Cet arbre arrive à classer les 67 866 déplacements avec un taux d'erreur de 68,72%. Avec un taux d'erreur frôlant les 70%, les variables explicatives utilisées ne semblent pas en mesure d'expliquer le choix modal de l'ensemble de la population.

Il est intéressant de remarquer, à la Figure 5-3, la distance seuil utilisée par le modèle à savoir 1,1 km. Que les gens aient ou non de titre mensuel de TC, en deçà du seuil fixé à 1,1 km, la majorité des gens choisissent de marcher. Cette distance est très similaire aux distances seuil de 1,15 et 1,25 km mentionnées plus haut. Pour les plus longs déplacements, ceux qui ont accès à la voiture emprunteront majoritairement ce mode. Afin d'aider à la lecture de cette figure, une version plus grande est jointe en annexe C.

### Arbre de décision du choix modal pour les non-détenteurs de permis de conduire

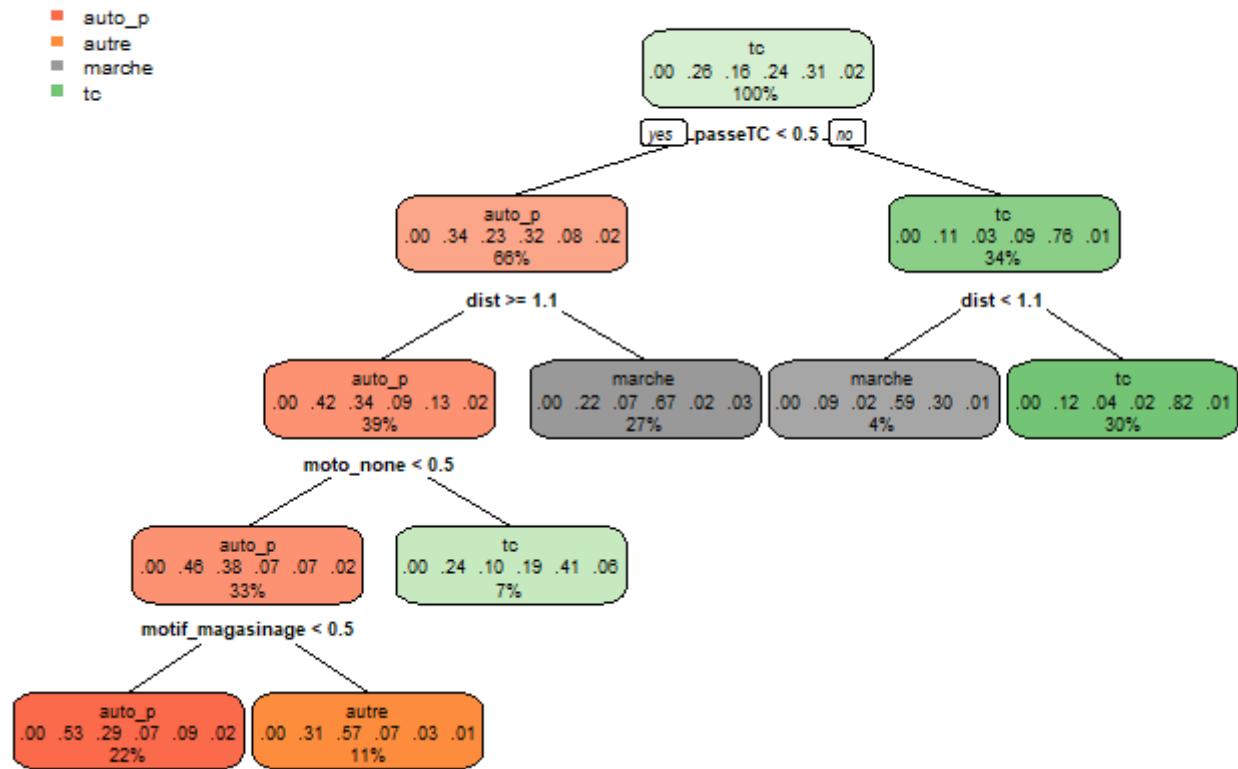


Figure 5-3 : Arbre de décision, non détenteurs de permis de conduire

### 5.1.2 Détenteurs de permis de conduire

De manière similaire, un arbre a été créé pour les déplacements des détenteurs de permis de conduire. Cet arbre comportant 7 feuilles et 6 tests de variables classe 188 103 déplacements avec un taux d'erreur de 31,38%.

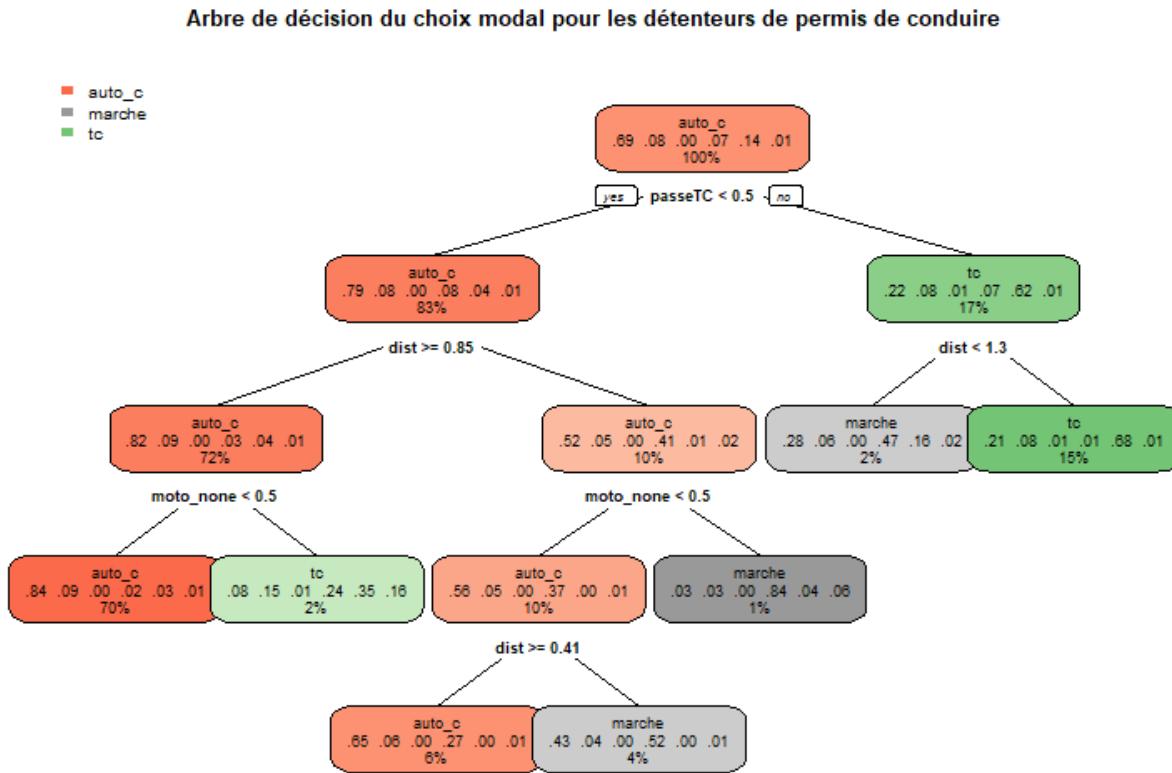


Figure 5-4 : Arbre de décision, détenteurs de permis de conduire

Le premier facteur utilisé dans l'arbre de la Figure 5-4 est également la possession du titre mensuel de TC. Pour ceux qui n'en n'ont pas, le second facteur est une distance seuil établie à 0,85 km. Au-delà de cette distance, s'il y a présence de véhicule au domicile, le déplacement a de fortes chances de se faire en automobile comme passager. Sinon, le TC devient le mode de rechange. Pour les distances en deçà de 0,85 km, les personnes vivant dans un ménage dépourvu de voiture marcheront tandis que celles pouvant se faire conduire seront majoritaires à le faire jusqu'à une distance d'environ 400 m, distances-en deçà de laquelle elles marcheront. Toute comme l'arbre précédent, une version plus grande est disponible en annexe D.

L'arbre est bien plus simple du côté des détenteurs d'un titre mensuel de TC. De ce côté, la marche est le mode de prédilection jusqu'à une distance de 1,3 km. Il est intéressant de noter la différence de distance seuil de la marche entre les détenteurs de titre mensuel TC et ceux qui n'en n'ont pas.

Il existe toutefois une limitation à l'utilisation de la possession d'un titre mensuel comme variable explicative puisque suite à une modification dans l'offre, quelqu'un pourrait être tenté de s'en

procurer un. Pour remédier à cette situation, un dernier arbre a été calculé sans l'utilisation des variables « possession d'un permis de conduire » et « possession d'un titre mensuel ». Puisque le niveau de motorisation des ménages est également un facteur de possession matérielle, cette variable est également retirée.

### Arbre de décision du choix modal excluant le niveau de motorisation, la possession du permis de conduire et du titre mensuel de TC

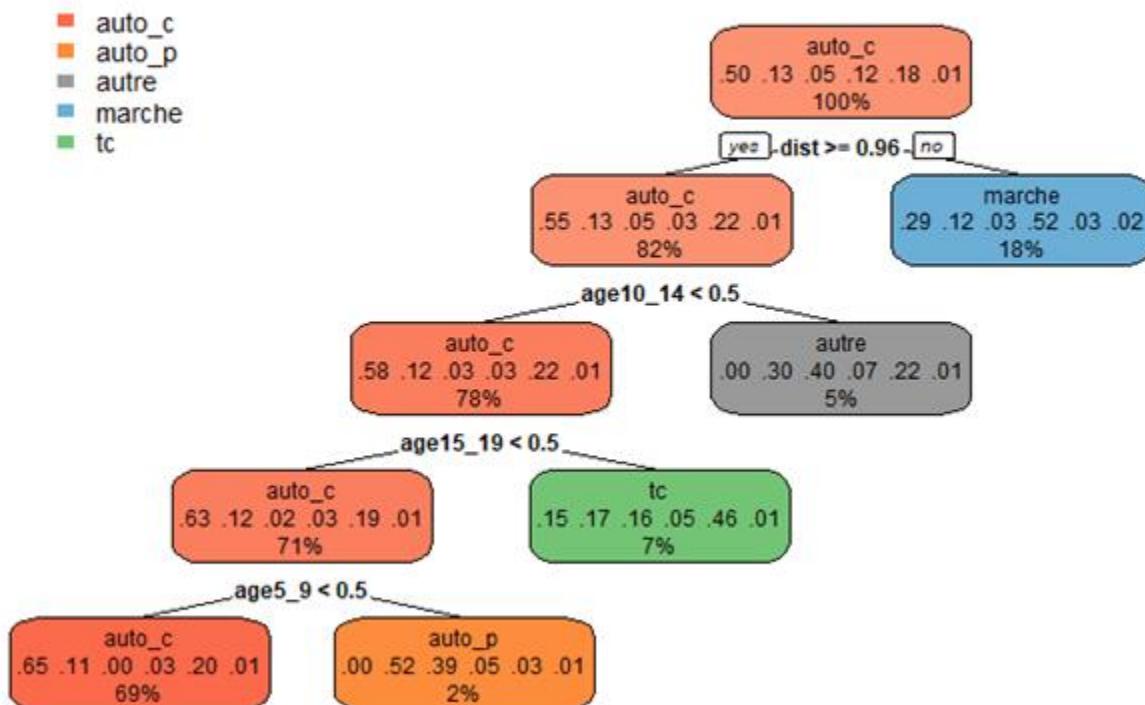


Figure 5-5 : Arbre de décision excluant les facteurs de possession du permis de conduire, de titre mensuel TC et de motorisation du ménage

Avec une erreur relative de 49,53%, le modèle de la Figure 5-5 arrive à trier adéquatement 129 184 des 255 969 déplacements. Pour cet arbre, les variables les plus importantes sont la distance et l'âge. En effet, le premier critère de division est la distance définie à 0,96 km. En deçà de cette distance, les piétons sont majoritaires. Toutefois, pour les distances plus élevées, l'âge semble le critère qui dicte le mode choisi. Les plus jeunes enfants (5 à 9 ans) sont plus propices à être passagers dans une voiture alors que les pré-adolescents (10 à 14 ans) sont majoritaires à utiliser un autre mode, catégorie dont font partie les autobus scolaires. Les jeunes adultes (15 à 19 ans)

sont quant-à-eux majoritaires à choisir le TC. Passé cet âge, environ les deux tiers des déplacements sont effectués en tant qu'automobilistes conducteurs.

### 5.1.3 Travailleurs adultes en heure de pointe am

Jusqu'ici, les modèles ont été calés sur l'ensemble ou de larges sous-ensembles des données. L'hétérogénéité des habitudes de transport nuit à l'élaboration de modèles plus précis. La modélisation de groupes plus spécifiques d'échantillons aurait le potentiel d'être plus précise pour la classe étudiée. Le premier groupe ainsi étudié est les travailleurs adultes en heure de pointe am. Ce groupe comprend 28 863 observations; l'arbre résultat a un taux d'erreur de 36,819%.

La Figure 5-6 présente les seuils obtenus en modélisant le choix modal selon le sexe, le groupe d'âge (à partir de 20-24 ans), la distance et les ratios de compétitivité. En premier lieu, la figure indique que les gens se rendent principalement au travail soit en conduisant leur véhicule (63%) ou en transport collectif (23%). Toutefois, lorsque le trajet est 3,1 fois plus long en TC, soit 66% des déplacements à motif travail, la majorité des gens choisissent leur voiture. Pour ces déplacements, la répartition modale est la suivante : 71% d'auto conducteur, 6% d'auto passager, 13% de TC et 9% de modes actifs.

Pour les déplacements en deçà de ce seuil de compétitivité, le TC est majoritaire pour les déplacements inférieurs à 14 km. Tout en étant minoritaire dans les déplacements de plus de 14 km, 25% de ces déplacements ont été effectués en TC.

Finalement, pour les ratios TC très compétitifs, soit inférieurs à 2,4, la majorité des déplacements (59%) ont lieu en TC. Il est intéressant de remarquer que, contrairement à la majorité des travailleurs qui privilégiennent leur voiture personnelle, les 25-34 ans sont majoritaires à utiliser le transport collectif pour se rendre au travail.

### Arbre de décision du choix modal des travailleurs adultes en heure de pointe am

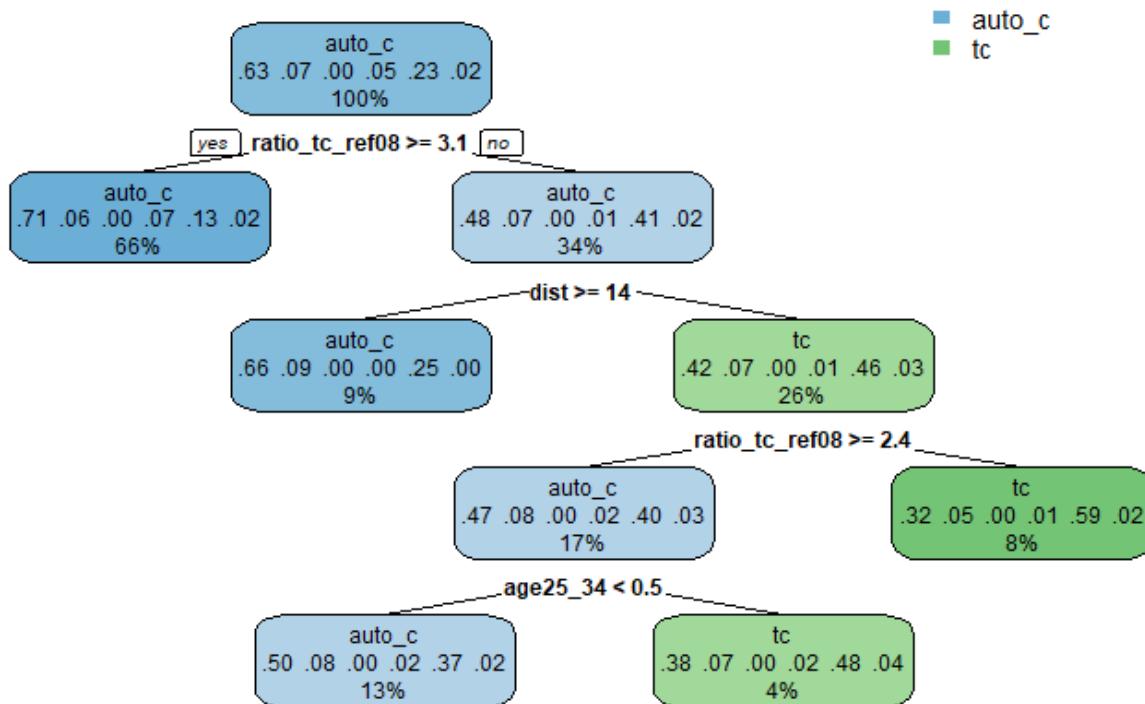


Figure 5-6 : Arbre des travailleurs en heure de pointe am

#### 5.1.4 Déplacements travail

En élargissant les données à l'ensemble des 50 991 déplacements à motif travail, l'arbre évolue tel qu'ilustré à la Figure 5-7. Similaire à celui de la section précédente et légèrement plus précis avec un taux d'erreur de 35,549%, celui-ci met en évidence l'attrait de la marche pour les courts déplacements. En effet, la part modale de la marche devient équivalente à l'automobile pour les déplacements de 1,2 km et moins, tout en atteignant 54% contre 36% pour l'automobile conducteur pour ceux de 730 m et moins. Pour les plus grands déplacements, l'arbre est essentiellement identique au précédent, ayant une feuille en moins. Cette similarité était à prévoir étant donnée la réutilisation de certaines données.

## Arbre de décision du choix modal des travailleurs adultes

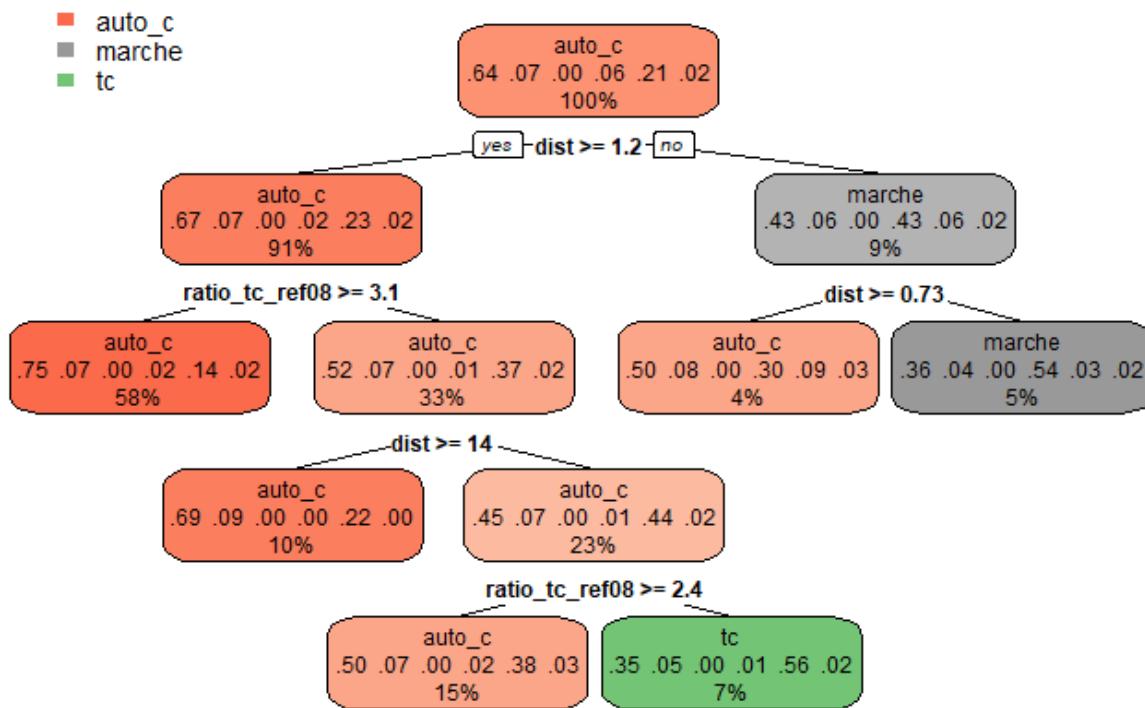


Figure 5-7 : Arbre de décision, adultes motif travail

### 5.1.5 Déplacements école

Les déplacements des écoliers sont une source importante de déplacements et les arbres présentés dans cette section pourraient permettre de mieux comprendre leurs habitudes de transport. Classant 25 422 observations avec un taux d'erreur de 71,155%, les modèles présentés dans cette section présentent une grande hétérogénéité dans les groupes. Une des causes de ce niveau de performance est la grande variété de modes utilisés par les écoliers. La feuille au sommet indique une répartition à 8% automobile conducteur, 20% automobile passager, 22% autre mode, tel que l'autobus scolaire, 19% marche, 29% TC et 2% vélo. Avec une telle variété, les modèles mettent en lumière les caractéristiques générationnelles.

### Arbre de décision, déplacements école tous les âges

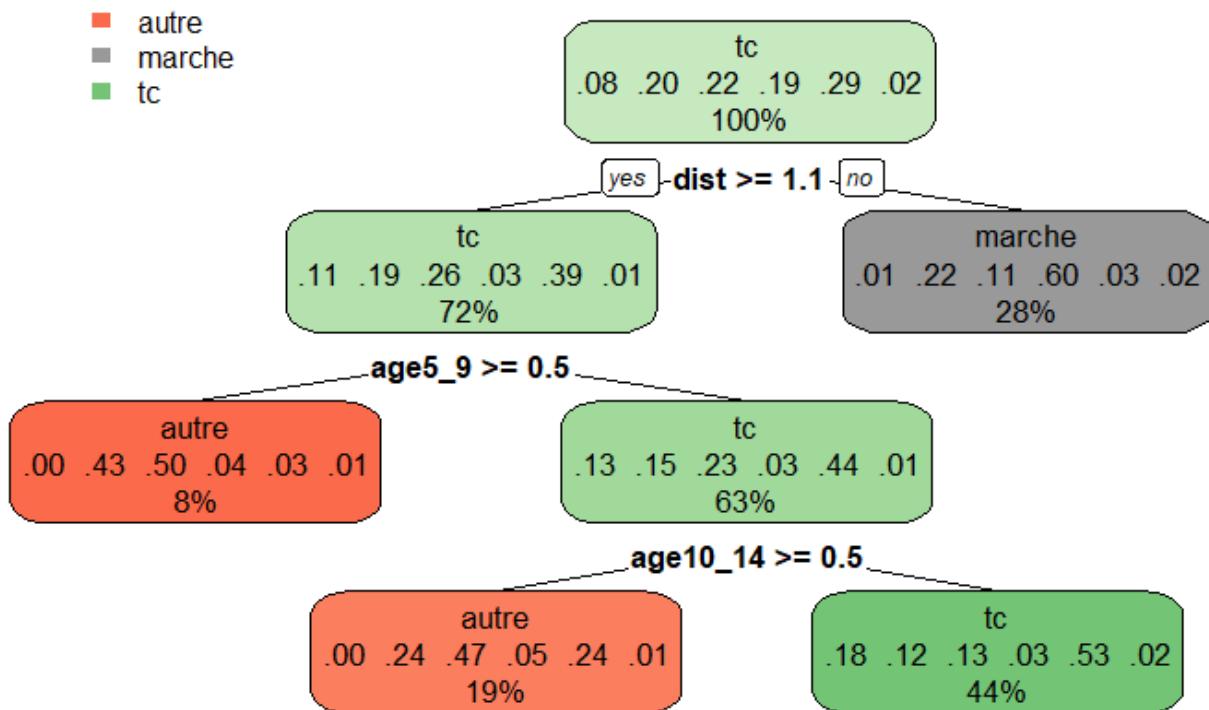


Figure 5-8 : Arbre de décision, déplacements école, tous les âges

Dans le premier arbre de la Figure 5-8, les facteurs explicatifs incluent le genre, tous les groupes d'âge, la distance et les ratios de chacun de modes. Les facteurs déterminants du choix modal sont l'âge et la distance. L'utilisation de la marche est principalement limitée par la distance. En effet, pour les déplacements en deçà de 1,1 km, la majorité (60%) des écoliers se rendent à l'école à pied alors que pour les déplacements plus longs, la marche obtient une part modale d'à peine 3%. Pour les déplacements plus longs, le facteur principal est l'âge.

En effet, les enfants de 5 à 9 ans sont majoritaires à prendre le bus scolaire (50%) et à se faire conduire à l'école (43%). Rendus entre 10 à 14 ans, les écoliers sont toujours majoritaires à prendre le bus scolaire (47%) alors que l'on observe un transfert de l'auto passager vers le TC, passant respectivement de 43% à 24% et de 3% à 24%. Chez les plus vieux, le TC domine avec une part modale de 53% alors que les autres modes tel le bus scolaire ne sont presque plus utilisés avec une part modale de 13%.

L'arbre de la Figure 5-9 tente d'expliquer le comportement des écoliers mais en excluant les groupes d'âge 5-9 ans et 10-14 ans des facteurs explicatifs. La distance seuil de 1,1 km pour les piétons demeure alors que le reste de l'arbre met en lumière les caractéristiques générationnelles de l'ensemble des écoliers. Chez les étudiants de 20 à 24 ans, le TC domine largement avec une part modale à 59% alors que 27% se tournent vers leur automobile et 9% s'y font conduire. Rendu de 25 à 34 ans, les comportements de mobilité ne semblent pas changer étant très similaires à ceux des 20-24 ans. Finalement, on retrouve les étudiants n'étant pas inclus dans les 20-34 ans, soit 57% des étudiants. Ceux-ci se déplacent principalement en TC ou en autre mode, tel le bus scolaire. On s'aperçoit que le TC est plus attractif sous la barre des ratios à 5,1 atteignant une part modale de 38% contre 30% pour le bus scolaire. Pour les déplacements ou le TC n'est pas attractif, près de la moitié des écoliers optent pour un autre mode, 22% se font conduire et 15% utilisent tout de même le TC.

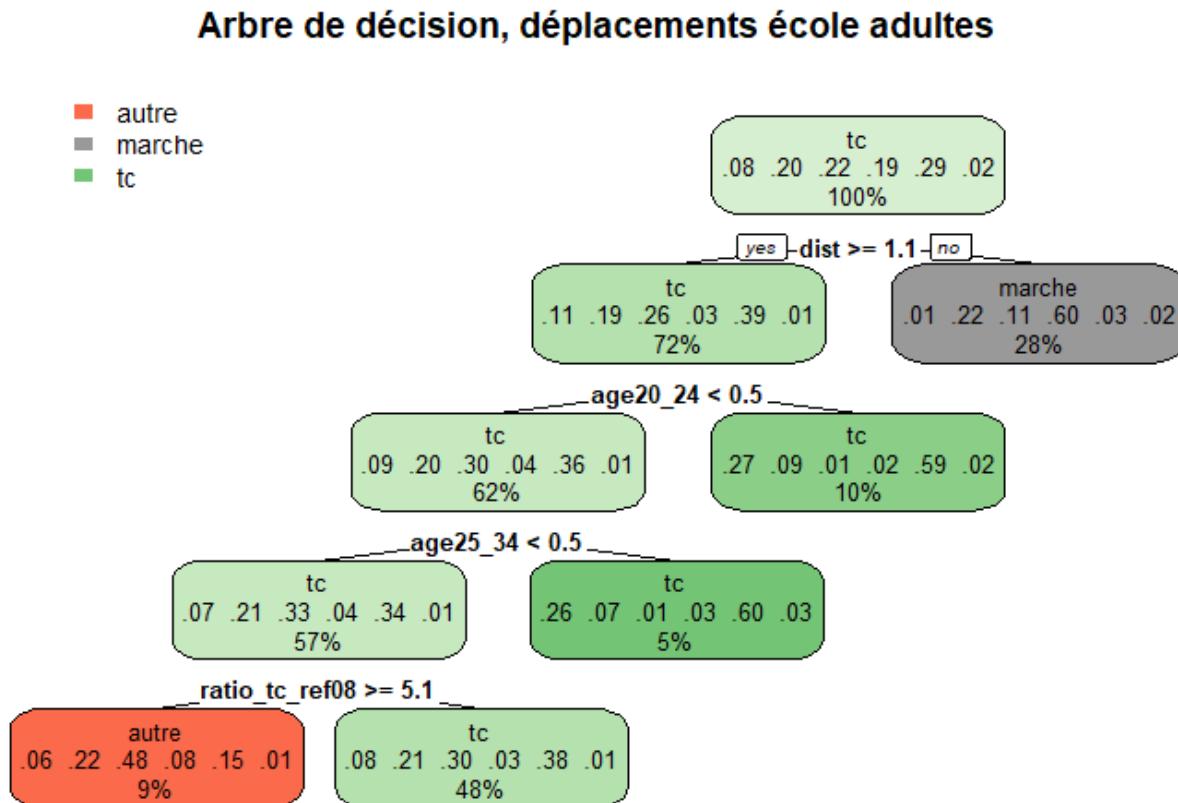


Figure 5-9 : Arbre de décision, déplacements école adultes

### 5.1.6 Déplacements magasinage

Contrairement au modèle scolaire où le TC et l'autobus scolaire étaient dominants, la Figure 5-10 illustre clairement que le TC n'est pas une option pour aller magasiner. Parmi les 21913 déplacements à motif magasinage, uniquement 9% des déplacements ont été effectués en TC. La plupart des déplacements se font en voiture (57% conducteur, 16% passager) ou à pied (16%). Pour ce motif de déplacement, la distance est le premier facteur qui influence le choix modal. Pour les déplacements de plus de 850 m, l'automobile conducteur prime à 64% de part modale alors que pour les déplacements inférieurs à 850 m, ce mode n'est emprunté que dans 34% des déplacements, laissant la majorité à la marche et ses 53% de part modale. La marche s'impose d'autant plus dans les déplacements de moins de 460 m ou pour ceux qui ont un ratio de marche inférieurs à 3,9. Il y a toutefois une grande hétérogénéité dans les déplacements ayant une distance entre 460 m et 850 m où il y a 44% d'automobilistes conducteurs et 40% de piétons.

**Arbre de décision du choix modal, déplacements magasinage**

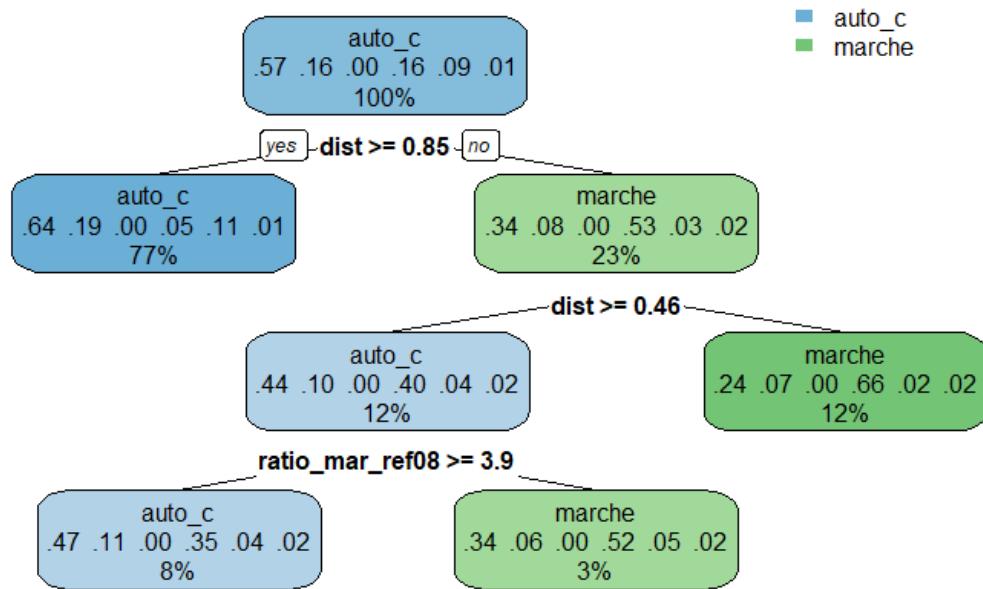


Figure 5-10 : Arbre de décision, déplacements motif magasinage

À la lumière de l'ensemble de ces modèles, il est intéressant de comparer les différents ratios TC retrouvés dans les modèles. Chez les travailleurs, on retrouve un seuil à 3,1, au-delà duquel la majorité des déplacements se font en voiture. Bien qu'à partir de ce seuil on retrouve pratiquement autant d'utilisateurs de TC que de conducteurs, c'est sous le seuil de 2,1 que les utilisateurs du TC sont nettement majoritaires avec une part modale de 59%. Chez les étudiants, le seuil est bien plus élevé, se situant à 5,1. Bien que les étudiants utilisent un choix de mode très varié, on pourrait en conclure que les travailleurs sont plus exigeants envers le TC que les étudiants, ce qui est corroboré par la valeur du temps plus importante des travailleurs.

## 5.2 Logit multinomial

Le second type de modèle est celui qui sera utilisé pour l'analyse des scénarios. Il s'agit du modèle logit multinomial. L'intérêt de ce modèle est la prise en compte d'une vaste étendue de variables explicatives, dans la mesure où elles ne sont pas corrélées entre elles et de calculer un facteur de pondération pour chacune dans l'équation probabiliste du choix modal. Le Tableau 5-1 présente les différentes variables ayant été utilisées dans le modèle logit. Pour chacune des paires de variables qui étaient trop fortement corrélées, une variable a été choisie et l'autre a été retirée des modèles. Ces variables sont présentées en rouge dans le Tableau 5-1.

Tableau 5-1 : Variables utilisées (noir) et retirées (rouge) dans le modèle logit

### 5.2.1 Indépendance des variables explicatives

L'indépendance des variables est vitale au bon fonctionnement des modèles prédictifs de type logit. Tel que mentionné en section 2.7, onze variables ont été sélectionnées parmi lesquelles 7 sont de type catégorielle et 4 sont de type continu. La présentation détaillée de ces variables est exposée au Tableau 5-2.

Tableau 5-2 : Présentation détaillée des variables explicatives

Variables de la personne	Type
Genre	Catégorielle à 2 classes
Âge	Catégorielle à 10 classes
Possession du permis de conduire	Catégorielle à 2 classes
Possession d'un titre mensuel TC	Catégorielle à 2 classes
<b>Variables du ménage</b>	
Revenu	Catégorielle à 7 classes
Motorisation	Catégorielle à 3 classes
<b>Variables du déplacement</b>	
Motif	Catégorielle à 6 classes
Heure de départ	Catégorielle à 6 classes
Distance	Continu
Ratio_TC	Continu
Ratio_Vélo	Continu
Ratio_Marche	Continu

Contrairement aux arbres de décisions, les modèles logit nécessitent que l'ensemble des variables explicatives soient indépendantes. Afin d'assurer ce niveau d'indépendance des variables, la matrice de corrélation linéaire (coefficient de Pearson) entre variables a été calculée. Toute paire de variables présentant une valeur absolue du coefficient de corrélation supérieure à 0,4 impliquera que ces deux variables ne peuvent être simultanément intégrées dans le modèle. La

Figure 5-11 présente la matrice de corrélation complète sur laquelle ces paires sont identifiées en bleu. Pour chacune de ces paires, un choix devra être fait afin de ne conserver qu'une variable par paire. Les variables suivantes sont donc susceptibles d'être retirées des modèles : âge 10-14, motif retour, motif magasinage, motorisation partielle, motorisation complète, pointe matin, midi, pointe soir, ratio vélo, ratio marche.

	homme	age5_9	age10_14	age15_19	age20_24	age25_34	age35_44	age45_54	age55_64	age65_74	age75plus	age5_74	age75plus	rev20kmoins	rev20k_39k	rev40k_59k	rev60k_89k	rev80k_100k	rev100kplus	rev autre	permisoui	motif_magasinage	motif_ecole	motif_loisir	motif_retour	motif_autre	moto_full	moto_part	moto_none	grphre1	grphre2	grphre3	grphre4	grphre5	grphre6	dist	ratio_tc_ref08	ratio_vel_ref08	ratio_mar_ref08
homme																																							
age5_9	2%																																						
age10_14	2% -6%																																						
age15_19	2% -6% -8%																																						
age20_24	0% -5% -6% -7%																																						
age25_34	-2% -9% -11% -11% -9%																																						
age35_44	-2% -10% -13% -14% -11% -18%																																						
age45_54	0% -11% -13% -14% -12% -19% -23%																																						
age55_64	0% -9% -11% -12% -9% -15% -18% -19%																																						
age65_74	0% -6% -8% 8% -7% -11% -13% -14% -11%																																						
age75plus	0% -4% -5% -6% -5% -8% -9% -10% -8% -6%																																						
rev20kmoins	-3% 0% -1% -2% 3% 1% -3% -6% 0% 7% 7% 7%																																						
rev20k_39k	-1% -1% -1% -3% 0% 3% -2% -5% 1% 7% 5% -11%																																						
rev40k_59k	1% 0% -1% -1% 0% 3% -1% 0% 1% -1% -2% -12% -18%																																						
rev60k_89k	1% 1% 0% 0% -1% 3% 2% 2% -1% -5% -5% -10% -16% -17%																																						
rev80k_100k	1% 2% 1% 0% -1% 2% -2% 6% 5% -3% -8% -7% -12% -18% -18% -16% -14%																																						
rev100kplus	2% 2% 3% 2% -1% -2% 6% 5% -3% -8% -7% -12% -18% -18% -16% -14%																																						
rev autre	-1% -2% -1% -1% -3% -6% -3% 2% 4% 6% 6% -12% -19% -19% -17% -15% -19%																																						
permisoui	7% -36% -45% -23% 1% 12% 19% 20% 14% 5% -4% -13% -5% 2% 4% 4% 5% 3%																																						
passcoui	-4% -11% -1% 22% 16% 6% -4% -5% -8% -7% -5% 7% 3% -1% -2% -3% -5% -2% -18%																																						
motif_travail	4% -11% -13% -10% 0% 8% 10% 14% 2% -10% -9% -6% -3% 1% 2% 2% 3% 0% 18% 2%																																						
motif_magasinage	2% 26% 31% 27% 8% -8% -14% -15% -13% -9% -7% 1% -1% -1% 0% 1% 1% -2% -37% 10% -17%																																						
motif_ecole	-4% -6% -7% -7% -4% -3% -3% -1% 8% 14% 11% 5% 4% 0% -2% -3% -5% 2% 5% -6% -15% -10%																																						
motif_loisir	-1% -3% -3% 0% 1% -1% -4% -3% 3% 7% 5% 2% 1% 0% -1% -1% -1% 0% 2% -2% -14% -9% -8%																																						
motif_retour	0% 1% 2% 1% 0% -1% -2% 0% 0% 0% 0% 0% 0% 0% 0% 0% 0% 0% 0% -44% -30% -27% -24%																																						
motif_autre	-3% -6% -8% -8% -5% 3% 10% 1% 0% 3% 3% 1% -2% -5% -10% -7% 0% 4% 10% 2% 17% -26% 4% -3% -1% 1% -2% 2%																																						
moto_full	-1% 7% 5% -13% -8% 1% 9% -1% 2% -2% -5% -10% -7% 0% 4% 5% 10% 2% 17% -26% 4% -3% -1% 1% -2% 2%																																						
moto_part	5% -5% -3% 15% 6% -5% -6% 3% -2% -1% -2% -9% -2% 3% 1% 1% -1% 0% 1% 9% -2% 4% -2% -3% 1% 0% -79%																																						
moto_none	-5% -3% -4% -3% 3% 6% -4% -4% 0% 4% 10% 29% 13% -4% -8% -9% -14% -3% -28% 26% -3% -2% 5% 3% 1% -3% -30% -34%																																						
grphre1	6% -4% -4% -2% 0% 1% 3% 7% 1% -5% -4% -3% -1% 1% 2% 2% 0% 0% 6% 2% 35% 1% -7% -5% -21% -3% 2% 0% -3%																																						
grphre2	-1% 8% 9% 4% -1% 1% 2% -6% -6% -7% -6% -3% -2% -1% 0% 2% 4% 0% -8% 3% 35% 41% -10% -9% -46% 4% 1% 1% -3% -14%																																						
grphre3	-4% -1% -2% -3% -2% -6% -7% -6% 16% 15% 8% 6% 0% -3% -5% -8% 3% -3% -8% -14% -13% 20% 2% 6% 2% -2% -2% 6% -16% -38%																																						
grphre4	0% -1% -1% 0% 0% 2% 4% 3% -2% -5% -5% -3% -2% 0% 1% 2% 4% -1% 3% 3% -27% -19% -6% 3% 36% -2% 1% 0% -2% -15% -36% -40%																																						
grphre5	2% -3% -3% -1% 2% 2% 1% 3% 0% -2% -3% -1% -1% 0% 1% 1% 2% -1% 4% 1% 12% 8% -2% 0% 0% 0% -6% -16% -17% -17%																																						
grphre6	2% -4% -4% 2% 6% 3% -1% 1% -1% -2% 0% 1% 0% 0% 0% 0% 0% 0% 0% 0% 18% 4% -4% -2% 1% 1% -5% -13% -15% -14% -6%																																						
dist	6% -13% -11% 0% 6% 4% 6% 10% 0% -7% -8% -9% 0% 3% 5% 7% 2% 21% 6% 20% -6% -12% -4% 0% -7% 11% -2% -14% 20% -1% -11% 4% -3% 0%																																						
ratio_tc_ref08	0% 6% 2% -2% -2% -1% -1% -2% 0% 1% 2% 2% 0% 0% -1% -2% -1% -5% -4% -3% 1% 2% -1% -1% 3% -2% 0% -1% -12%																																						
ratio_vel_ref08	5% -11% -9% -1% 3% 3% 5% 8% 1% -4% -5% -7% -6% 0% 2% 3% 5% 1% 18% 2% 16% -6% -6% -3% -1% -6% 9% -1% -12% 14% -2% -7% 3% -2% 0% -1% -12%																																						
ratio_mar_ref08	3% -7% -6% 0% 2% 2% 3% 5% 1% -2% -3% -5% -4% 0% 2% 2% 3% 1% 11% 1% 9% -4% -3% -1% -1% -3% 6% -1% -8% 8% -1% -4% 2% -1% 0% 40% 10% 57%																																						

Figure 5-11 : Facteurs de corrélation dépassant le seuil

## 5.2.2 Résultats du modèle

Les modèles logit ont été calculés avec le logiciel RStudio, plus particulièrement les packages nnet pour le calage des données et pscl pour le calcul des indicateurs statistiques. Les résultats tirés de ce modèle sont principalement les coefficients de risques relatifs rapprochés (RRR), aussi connus comme le rapport de cotes ou « Odds Ratios » (OR). Pour ce premier modèle, 255 969 observations ont été utilisées. Le Tableau 5-3 présente les coefficients RRR pour chaque facteur du modèle et pour chacune des alternatives modales. Les niveaux de significativité statistique sont indiqués par la quantité d'étoiles présente suite au coefficient, les seuils étant les suivants : \*p<0.1; \*\*p<0.05; \*\*\*p<0.01. Le terme en parenthèse est l'erreur standard associé au paramètre. Les facteurs suivants n'étant pas significatifs pour l'ensemble des modes, ils ont été retirés du modèle : motif école, motorisation nulle, départ en pré-pointe am et départ le soir.

Les rapports de cotes, ou Odds Ratios (OR), sont obtenus à partir de l'exponentielle des coefficients du modèle logit tel que définie à la formule (5.1) où  $P_M$  est la probabilité de choisir un mode autre que le mode de référence choisi, soit l'auto conducteur et  $P_{AuC}$  est la probabilité de choisir l'automobile conducteur.

$$(5.1) \quad RRR = \frac{P_M/(1 - P_M)}{P_{AuC}/(1 - P_{AuC})}$$

Une autre méthode est disponible pour calculer les RRR, l'exponentielle des coefficients du modèle. Ainsi, les RRR peuvent également être définis par la formule (5.2) où les  $\beta_i$  sont les divers coefficients issus du modèle logit multinomial.

$$(5.2) \quad RRR = e^{\beta_i}$$

La robustesse du modèle est évaluée par le pseudo- $R^2$  de McFadden. Ce ratio est défini à l'équation (5.3) où  $L_M$  est la probabilité prédite par le modèle et  $L_0$  est la probabilité prédite quand l'influence de tous les facteurs est réduite à 0.

$$(5.3) \quad R^2_{McFadden} = 1 - \frac{\log L_M}{\log L_0}$$

Avec l'utilisation du module pscl de RStudio, le pseudo $R^2$  de McFadden, aussi appelé  $\rho^2$ , est de 0,501. Bien que reconnue comme étant inférieure au  $R^2$  traditionnel, l'interprétation de cette valeur

est assez subjective et une valeur entre 0,2 et 0,4 est jugée comme un excellent calage de modèle (McFadden 1979). De plus, plus cette valeur est élevée, meilleur est le modèle. Une valeur de 0,501 est donc plus que suffisante pour analyser en profondeur ce modèle.

Tableau 5-3 : Rapports de cotes

	<i>Dependent variable:</i>				
	auto_p	Autre	marche	tc	velo
	(1)	(2)	(3)	(4)	(5)
homme	0.375*** (0.016)	0.543*** (0.027)	0.668*** (0.019)	0.620*** (0.018)	1.578*** (0.037)
age5_9	3.661*** (0.372)	1.585*** (0.372)	1.185*** (0.373)	0.595*** (0.380)	0.988*** (0.384)
age15_19	0.221*** (0.063)	0.268*** (0.065)	0.405*** (0.067)	0.471*** (0.065)	0.497*** (0.094)
age20_24	0.090*** (0.066)	0.019*** (0.108)	0.176*** (0.075)	0.244*** (0.067)	0.307*** (0.106)
age25_34	0.085*** (0.061)	0.008*** (0.113)	0.163*** (0.065)	0.220*** (0.064)	0.269*** (0.093)
age35_44	0.062*** (0.061)	0.005*** (0.117)	0.105*** (0.064)	0.166*** (0.063)	0.212*** (0.092)
age45_54	0.067*** (0.060)	0.006*** (0.109)	0.105*** (0.064)	0.167*** (0.063)	0.182*** (0.092)
age55_64	0.115*** (0.060)	0.009*** (0.107)	0.118*** (0.064)	0.213*** (0.064)	0.125*** (0.099)
age65_74	0.167*** (0.062)	0.012*** (0.111)	0.096*** (0.067)	0.223*** (0.068)	0.075*** (0.120)
age75plus	0.192*** (0.066)	0.018*** (0.105)	0.064*** (0.072)	0.168*** (0.076)	0.016*** (0.230)
rev20kmoins	0.912*** (0.045)	0.942*** (0.074)	1.017*** (0.049)	1.179*** (0.048)	1.759*** (0.100)
rev20k_39k	0.967*** (0.035)	0.951*** (0.058)	1.086*** (0.041)	1.087*** (0.038)	1.800*** (0.089)
rev40k_59k	0.993*** (0.034)	0.911*** (0.057)	1.007*** (0.041)	0.989*** (0.037)	2.034*** (0.089)
rev60k_79k	1.056*** (0.035)	0.997*** (0.059)	1.096*** (0.043)	1.023*** (0.039)	2.068*** (0.092)
rev80k_99k	1.112*** (0.037)	0.970*** (0.061)	0.993*** (0.046)	0.899*** (0.042)	2.047** (0.098)
rev100kplus	1.165*** (0.034)	0.824*** (0.057)	1.074*** (0.043)	0.960*** (0.038)	2.011*** (0.093)
rev_autre	1.103***	1.068***	1.088***	1.000***	1.453***

Tableau 5 3 : Rapport de cotes (suite)

	(0.034)	(0.056)	(0.041)	(0.037)	(0.094)
percond	0.003*** (0.103)	0.001*** (0.111)	0.005*** (0.105)	0.004*** (0.104)	0.006*** (0.112)
passeTC	2.897*** (0.023)	1.139*** (0.043)	2.831*** (0.030)	44.037*** (0.021)	1.462*** (0.058)
motif_travail	0.935*** (0.028)	0.330 (0.084)	1.019*** (0.034)	0.915*** (0.029)	1.257 (0.055)
motif_ecole	1.341 (0.027)	0.095 (0.148)	0.679 (0.031)	0.386 (0.038)	0.517 (0.075)
motif_loisir	1.638*** (0.028)	0.167* (0.099)	0.994*** (0.036)	0.561*** (0.037)	1.070*** (0.066)
motif_autre	0.912*** (0.027)	0.162*** (0.093)	0.449*** (0.032)	0.263*** (0.037)	0.280*** (0.082)
moto_part	2.457*** (0.017)	2.143*** (0.029)	2.593*** (0.021)	2.917*** (0.021)	2.475*** (0.046)
moto_none	23.709 (0.057)	59.200 (0.074)	126.124 (0.058)	151.060 (0.056)	171.834 (0.073)
grphre1	1.167 (0.042)	0.799 (0.079)	0.425 (0.075)	0.625 (0.044)	0.617 (0.103)
grphre2	1.022*** (0.025)	1.068*** (0.035)	0.785*** (0.027)	0.864*** (0.027)	0.866*** (0.052)
grphre4	1.291*** (0.021)	0.803*** (0.036)	0.763*** (0.024)	1.041*** (0.024)	0.946*** (0.048)
grphre5	1.822*** (0.031)	0.069* (0.154)	0.628*** (0.039)	0.420*** (0.040)	0.602*** (0.079)
grphre6	2.046* (0.034)	0.077 (0.164)	0.492 (0.051)	0.393 (0.043)	0.497 (0.092)
dist	1.013*** (0.001)	1.022** (0.002)	0.238 (0.012)	1.005*** (0.001)	0.851 (0.006)
ratio_tc_ref08	0.996*** (0.001)	0.933*** (0.003)	1.001*** (0.0003)	0.888*** (0.003)	1.000*** (0.001)
Constant	247.175*** (0.100)	662.264*** (0.110)	1,894.711*** (0.103)	86.498*** (0.104)	8.459*** (0.138)
Akaike Inf. Crit.	352 779	352 779	352 779	352 779	352 779

Note:

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

À la lumière de ce modèle, on peut identifier le niveau de motorisation partielle comme étant la variable catégorielle qui a le plus grand impact sur le choix modal et ce pour l'ensemble des modes. Provenant d'un modèle qui prédit les probabilités de choisir un mode autre que l'automobile conducteur, ce facteur semble très pertinent. Toutefois, une forte motorisation peut expliquer l'utilisation de l'automobile alors que le besoin de motorisation vient peut-être d'une offre TC déficiente. Ainsi, suite à une modification à l'offre TC, un ménage pourrait choisir de délaisser ses véhicules. L'utilisation de ce facteur ne semble donc pas appropriée dans l'élaboration des modèles.

Le second facteur commun à plusieurs modes est la possession d'un titre mensuel de transport collectif. Ce facteur est de loin le plus important pour le choix du TC. Néanmoins, tout comme le niveau de motorisation, il n'est pas le plus utile dans la compréhension du choix du transport en commun puisqu'il sera généralement acquis par les individus qui prennent régulièrement le transport en commun ; il est donc simultanément cause et conséquence. De plus, son utilisation nécessiterait le développement d'un modèle prédictif de sa possession à l'horizon 2031. Parmi les autres facteurs qui se démarquent, le revenu et le fait d'être un homme favorisent l'utilisation du vélo. Finalement, les enfants sont plus propices à se déplacer en tant que passager.

Afin de remédier au problème relié à l'utilisation de la possession du titre de transport mensuel et du niveau de motorisation comme variables explicatives, un second modèle sera calculé et présenté en section 5.2.3.

Il est important d'analyser les résidus, c'est-à-dire les erreurs faites par le modèle sur tous les déplacements. La Figure 5-12 illustre ces résidus par mode sous forme de graphiques moustaches. Ce type de graphique est utile pour représenter les caractéristiques statistiques d'un jeu de données, puisque les différentes caractéristiques des boîtes illustrent directement ces statistiques. Les limites de la boîte représentent les 25<sup>e</sup> et 75<sup>e</sup> percentiles des données, la longueur verticale de celle-ci représente l'intervalle interquartile, la division à l'intérieur de cette boîte la médiane et les extrémités des moustaches représentent une distance une fois et demi plus grande que l'étendue de la boîte. Ainsi, plus les boîtes sont petites, plus les résidus sont concentrés et meilleures sont les prédictions pour ce mode.

Pour parvenir à ce graphique, les résidus ont été normalisés et regroupés par mode prédit. L'échelle verticale représente donc la variance totale des résidus par mode. On peut y remarquer une grande répartition des erreurs pour l'ensemble des modes et de plus petites boîtes pour les modes auto conducteur, marche et TC. Ce serait donc les trois modes pour lesquelles le modèle arrive à faire les meilleures prédictions. De manière contraire, la qualité des prédictions pour le vélo et pour les autres modes est plus incertaine vu la plus grande dispersion de leurs résidus.

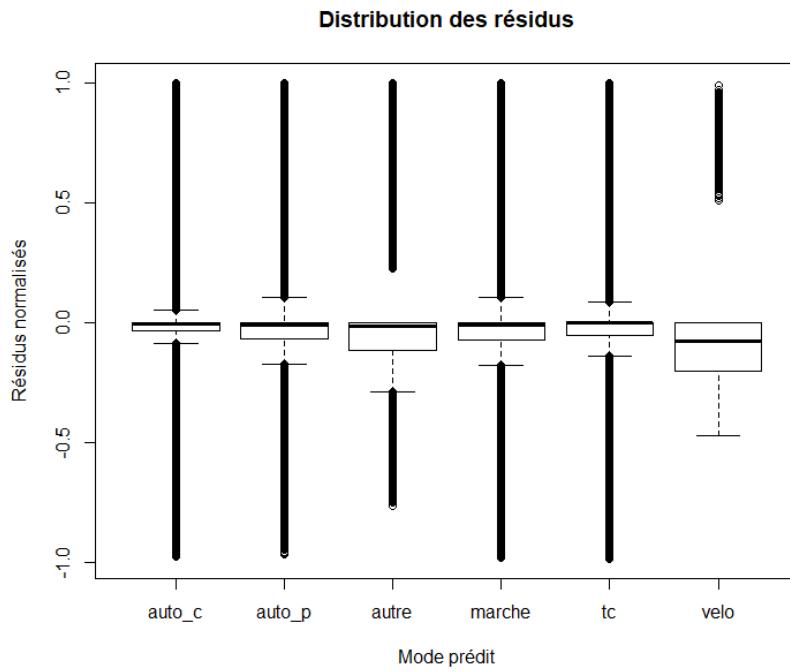


Figure 5-12 : Distribution des résidus du modèle logit

Afin de vérifier la validité des modèles, il est important de mesurer sa performance. Avec une méthode d'échantillonnage aléatoire de la base de données en deux jeux de données, le premier contenant 75% des données servant au calage du modèle, il est possible de mesurer la qualité des prédictions effectuées à partir du modèle pour le 25% des données restantes. Les figures suivantes présentent la qualité des probabilités calculées par le modèle en comparaison des modes déclarés. On remarque sur les Figure 5-13 à Figure 5-18 que le modèle semble être en mesure de bien expliquer les modes auto conducteur, auto passager, TC et marche mais semble éprouver de la difficulté avec le vélo, probablement dû à la faible quantité de déplacements fait à vélo dans l'enquête OD ainsi qu'à l'indisponibilité des variables pouvant expliquer son usage (sécurité, météo, présence d'infrastructure dédiée, etc.).

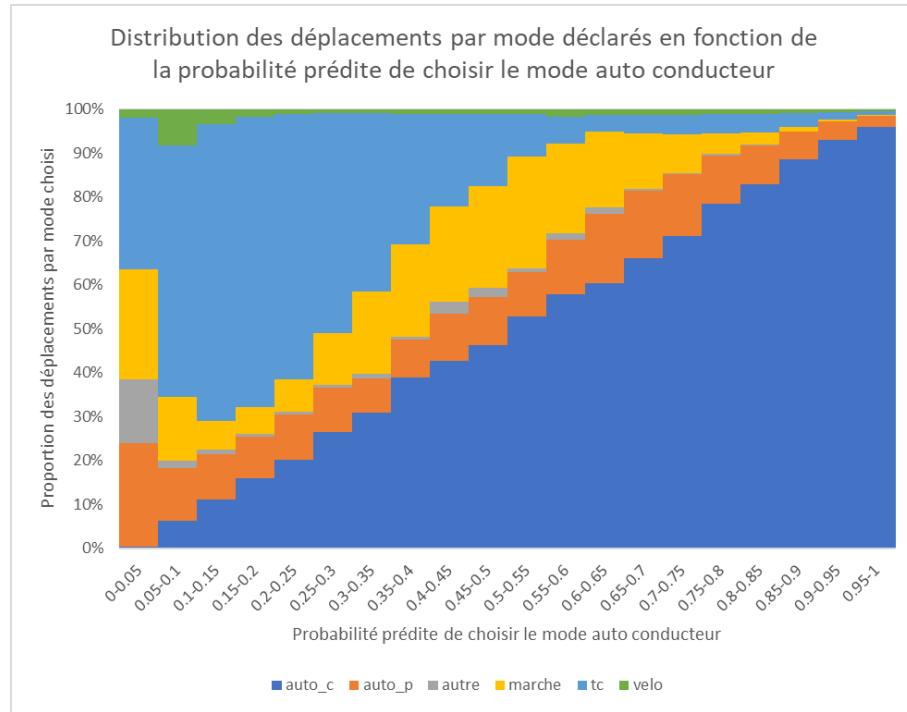


Figure 5-13 : Distribution des parts modales en fonction des probabilités de choisir le mode auto conducteur

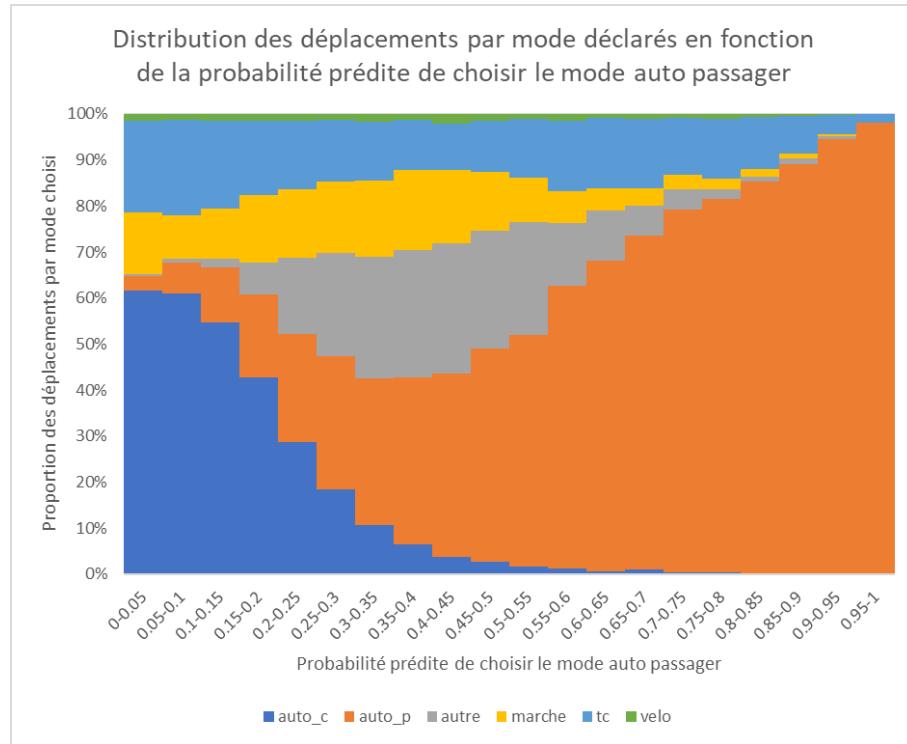


Figure 5-14 : Distribution des parts modales en fonction des probabilités de choisir le mode auto passager

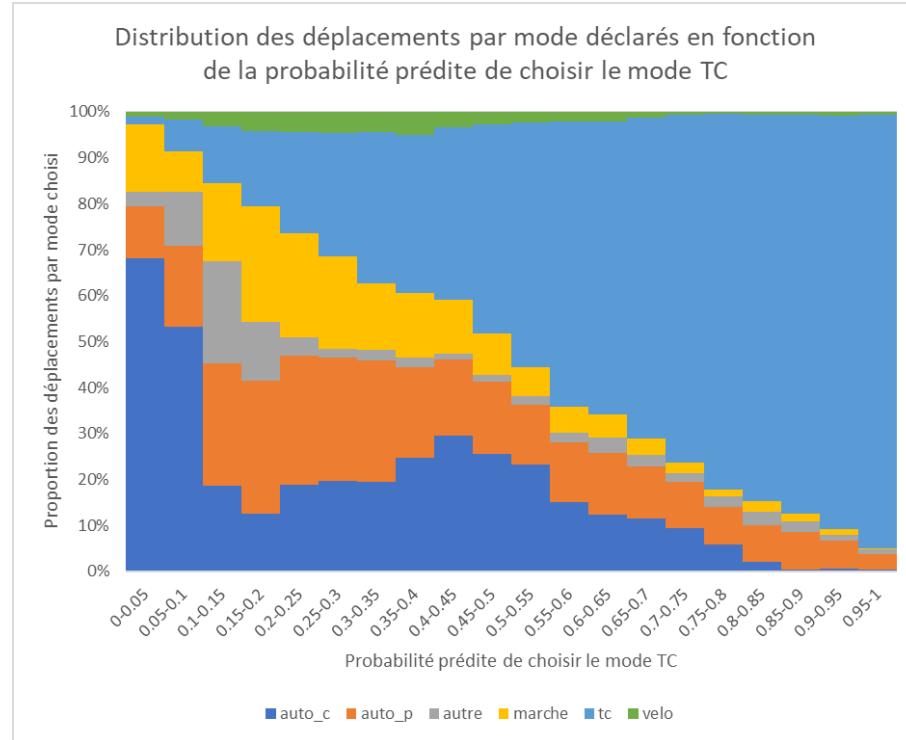


Figure 5-15 : Distribution des parts modales en fonction des probabilités de choisir le mode TC

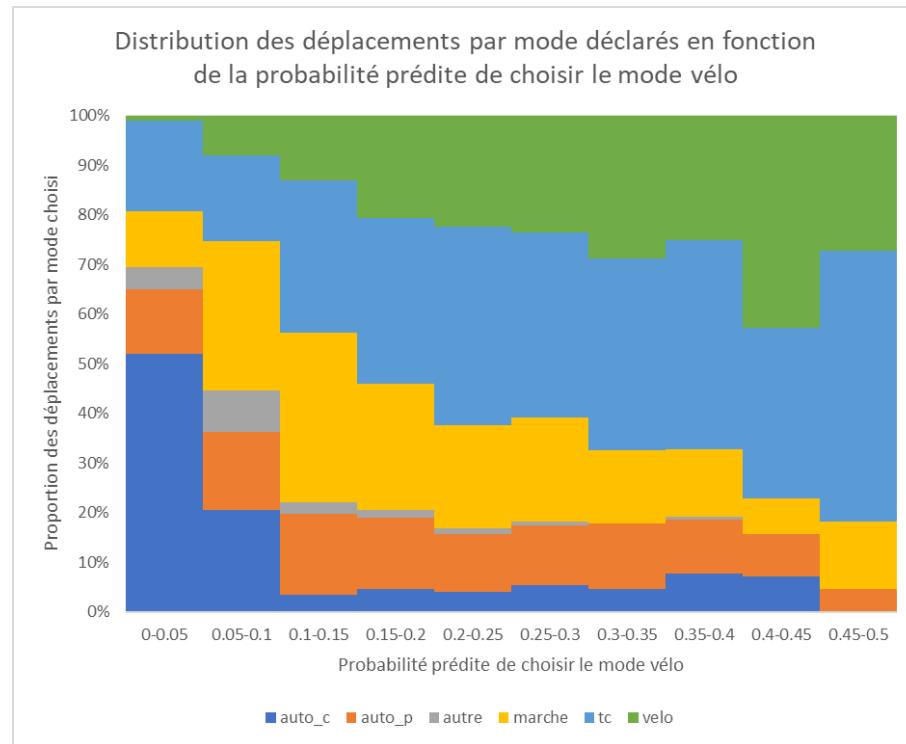


Figure 5-16 : Distribution des parts modales en fonction des probabilités de choisir le mode vélo

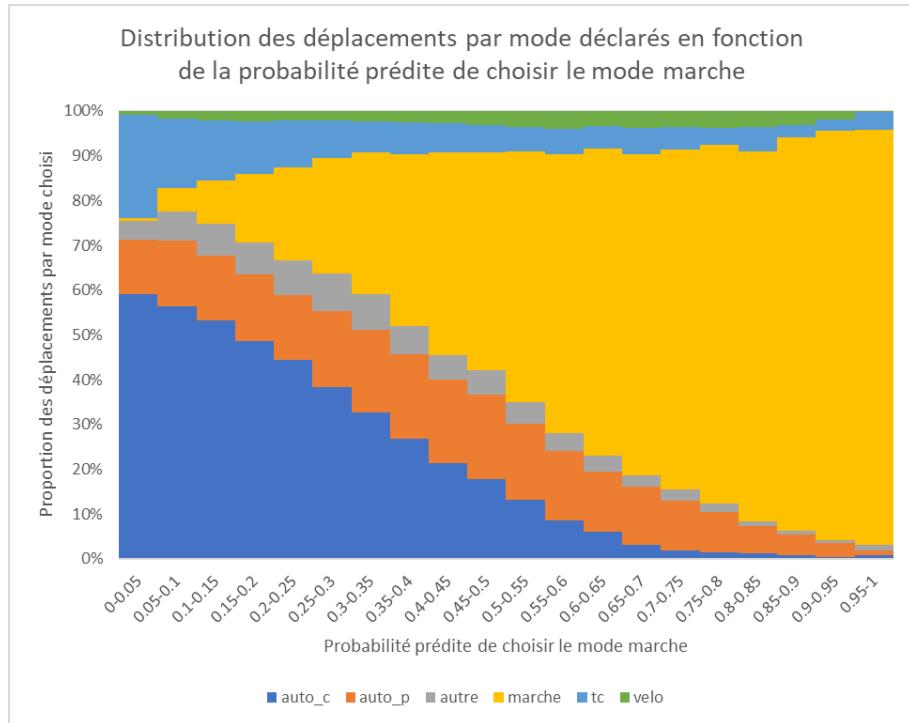


Figure 5-17 : Distribution des parts modales en fonction des probabilités de choisir le mode marche

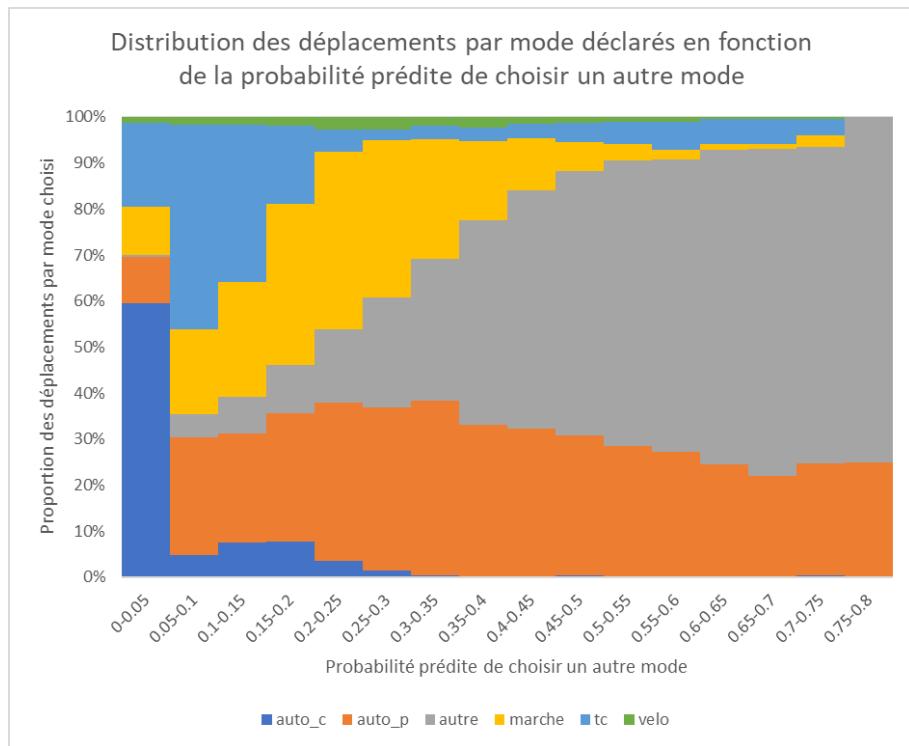


Figure 5-18 : Distribution des parts modales en fonction des probabilités de choisir un autre mode

### 5.2.3 Modèle excluant la possession du permis de conduire et du titre TC

Cette section présente brièvement un modèle logit évalué sans les facteurs de possession d'un permis de conduire ou d'un titre TC. Ainsi, de manière similaire au modèle précédent, le Tableau 5-4 présente les variables explicatives utilisées et celles retirées en rouge. En plus des différentes variables retirées pour des raisons de corrélation, la possession du permis du conduire et du titre mensuel de TC ont été retirés. Le niveau de motorisation a également été retiré par soucis de capacité de projection.

Tableau 5-4 : Variables explicatives utilisées

Variable indépendante à prédire		Variables explicatives catégorielles							Variable explicatives continues
Choix modal	Genre	Âge	Revenu	Motif	Possession permis de conduire	Possession de titre mensuel TC	Niveau de motorisation	Groupe d'heure de départ	
Auto Conducteur	Homme	5-9 ans	moins de 20k	Travail	Oui	Oui	Complète	Pré-pointe AM	Distance
Auto Passager	Femme	10-14 ans	20k à 39k	Magasinage	Non	Non	Partielle	Pointe AM	Ratio TC*
TC		15-19 ans	40k à 59k	École				Nulle	Jour
Vélo		20-24 ans	60k à 79k	Loisir				Pointe PM	Ratio Vélo
Marche		25-34 ans	80k à 99k	Retour				Post-pointe PM	Ratio Marche
Autre		35-44 ans	100k+	Autre				Soir	
		45-54 ans							
		55-64 ans							
		65-74 ans							
		75 ans+							

Le **Erreur ! Source du renvoi introuvable.** présente les coefficients RRR des variables explicatives trouvées significatives. Les coefficients non-significatifs ont été identifiés en gris. Dans ce modèle, le motif du déplacement n'est plus significatif du tout et le revenu ne l'est plus pour les revenus supérieurs à 80 000\$ par année. De manière cohérente avec le premier modèle, le fait d'être un homme augmente la propension à utiliser le vélo et les enfants ont beaucoup plus de chance d'être transportés en auto que de prendre n'importe quel autre mode. Finalement, ce modèle met en lumière que les ménages qui ont un revenu inférieur à 20 000\$ par année ont plus de chance de ne pas utiliser l'automobile.

Tableau 5-5 : Rapports de cotes

Dependent variable:						
		auto_p	autre	marche	tc	velo
		-1	-2	-3	-4	-5
homme		0.365***	0.556***	0.667***	0.605***	1.701***
age5_9		3.124***	1.478***	1.034***	0.213***	0.908***
age15_19		0.004***	0.003***	0.008***	0.015***	0.010***
age20_24		0.001***	0.0001***	0.002***	0.005***	0.005***
age25_34		0.0004***	0.00003***	0.001***	0.002***	0.003***
age35_44		0.0003***	0.00002***	0.001***	0.001***	0.002***
age45_54		0.0003***	0.00002***	0.001***	0.001***	0.002***
age55_64		0.001***	0.00003***	0.001***	0.001***	0.001***
age65_74		0.001***	0.00005***	0.001***	0.001***	0.001***
age75plus		0.001***	0.0001***	0.001***	0.001***	0.0003***
rev20kmoins		1.297***	1.814***	2.074***	2.818***	3.891***
rev20k_39k		0.957***	1.069***	1.101***	1.261***	2.001***
rev40k_59k		0.825***	0.816***	0.723***	0.720***	1.531**
rev60k_79k		0.750***	0.746***	0.647***	0.589***	1.281***
dist		0.999	0.995	0.245	0.979*	0.833
ratio_tc_ref08		.993**	0.934	1.000***	0.765	0.997
Constant		499.831***	1,013.421***	4,335.272***	1,357.669***	18.502***

Note:

\* p&lt;0.1; \*\* p&lt;0.05; \*\*\* p&lt;0.01

Ce modèle a été calculé à partir de 255 969 observations et a un pseudo R<sup>2</sup> de McFadden de 0,305. Très simplifié, ce modèle a l'avantage d'être indépendant de l'état de possession d'un permis de conduire ou d'un titre mensuel de TC mais comparé au pseudo R<sup>2</sup> de McFadden du modèle précédent à 0,501, celui-ci est sensiblement moins précis, mais tout de même dans la fourchette de valeur que McFadden qualifie d'excellente.

## CHAPITRE 6 CONCLUSION ET RECOMMANDATIONS

L'objectif était de développer une méthode permettant d'analyser des scénarios d'offre de transport et de répartition de la population. La méthode devait permettre de quantifier l'utilisation de chacun des modes afin d'être utilisée dans des bilans de santé publique. Cette méthode devait aussi pouvoir servir à identifier des quartiers négligés par l'offre de transport en commun. Finalement, l'utilisation des ratios de compétitivité donnait l'espoir de pouvoir intégrer les modes actifs aux modèles.

En plus de faire la synthèse des travaux réalisés durant le développement de cette méthode, ce dernier chapitre fera un rappel des contributions apportées par ce mémoire ainsi que des limitations et des perspectives d'utilisation de cette méthode.

### 6.1 Synthèse des travaux

Une revue de l'état de l'art a permis de parfaire notre compréhension des variables explicatives du choix modal et de choisir celles qui étaient pertinentes et disponibles pour l'élaboration des modèles. Une liste de douze variables a ainsi été identifiée, quatre étant numériques (ratio TC, ratio vélo, ratio marche, distance) et les huit autres étant catégorielles (âge, sexe, revenu, possession du permis de conduire, possession d'un titre mensuel de TC, heure de départ, motorisation du ménage, motif).

En un second temps, une méthode d'analyse a été développée autour du concept de ratio de compétitivité, concept défini comme un outil de caractérisation de la performance modale ayant le potentiel d'être un facteur explicatif du choix modal. Bien qu'intéressant dans les analyses de performances modales, la corrélation entre la distance, le ratio du vélo et de la marche limite leur utilisation simultanée dans les modèles.

La méthode développée débute par la mise en commun des données provenant des diverses simulations et par le nettoyage de ces données par une série de filtres. Basés sur les données fournies dans l'enquête OD 2008, les filtres visaient à retirer les observations qui étaient non mobiles, des enfants de 4 ans et moins, effectués avant 5h ou passé 24h et celles dont l'origine ou la destination était hors du territoire d'enquête. Un dernier filtre retirait tous les déplacements n'ayant pas trouvé d'itinéraire en TC dû aux différents critères tel que le temps de marche maximal d'accès à l'origine et à la destination.

Après avoir comparé les données provenant de la Chaire Mobilité avec celles provenant du MOTREM08, les données de l'enquête OD 2008 et les simulations de temps de parcours de la Chaire Mobilité ont été utilisées pour définir la situation de référence qui allait être utilisée pour le calage de différents modèles explicatifs du choix modal. Les résultats de la comparaison montraient que les temps automobiles simulés par le MO semblaient être surestimés et que les temps TC étaient quant-à-eux probablement sous-estimés dû au modèle de simulation en régime permanent, menant à des ratios de compétitivité utiles en termes de comparaison de la performance relative.

Afin de s'affranchir du caractère arbitraire de la division horaire de la journée, une division basée sur le test de similarité statistique de Kolmogorov-Smirnov a été proposée.

À partir du scénario de référence, les déplacements ont été regroupés par vecteur de ratios de compétitivité afin de tenter d'identifier des typologies de vecteurs. Pour parvenir à regrouper les 255 969 observations présentes dans la base de données, une méthode de clustering à deux étapes est nécessaire, c'est-à-dire une méthode qui débute par la création de sous-ensembles par la méthode des k-moyennes avant de procéder au regroupement hiérarchique de ces sous-ensembles.

Finalement, deux types de modèles explicatifs du choix modal ont été développés, un arbre de décision et un modèle probabiliste logit multinomial. Ces modèles mettent en lumière l'influence prédominante qu'ont les facteurs associés à la possession d'un permis de conduire et d'un titre de transport mensuel. Des modèles excluant ces facteurs ont également été présentés et mettent en lumière le rôle joué par le ratio TC dépendant du motif du déplacement. On y voit entre autres que les modèles basés sur l'ensemble des données ont de mauvaises performances en classification et que des modèles ciblés sont plus pertinents. Dans ces modèles, on y voit également que les travailleurs sont plus exigeants en termes de performance du TC que les étudiants dans leurs choix modaux.

## 6.2 Rappel des contributions

Ce projet a apporté plusieurs contributions dans le développement de la méthode d'analyse des impacts des scénarios de transport et de répartition de la population au sein du collectif de recherche. Ces contributions sont principalement d'ordre méthodologique et d'analyse.

Au niveau méthodologique, la principale contribution de ce mémoire est la base d'une méthode d'analyse de scénarios d'offre TC mettant en lumière l'impact sur le choix modal qu'ont les temps

de parcours, représentés sous forme de ratios de compétitivité. Également, plusieurs limites ont été identifiées et des recommandations ont été formulées afin de servir de guide pour la suite des recherches.

Au niveau analytique, une méthode d'analyse spatiale des ratios de compétitivité a été proposée afin d'identifier des secteurs privilégiés ou démunis de service TC. Basé sur une segmentation du territoire en zones uniformes hexagonales d'un kilomètre carré, l'analyse de la répartition spatiale des valeurs d'autocorrélation spatiale (LISA) des ratios de compétitivité peut mettre en évidence les secteurs privilégiés et ceux négligés.

Une distance seuil pour la compétitivité du TC a également été proposée. Bien qu' étant sujet à discussion, des limitations, tel que l'omission de la pénibilité associée à la marche, ont été identifiées et des recommandations ont été formulées afin de raffiner le concept.

L'ensemble des variables explicatives disponibles à partir de l'enquête OD 2008 ont été étudiées et intégrées dans des modèles explicatifs du choix modal. Il a été recommandé d'exclure les variables associées à la possession du permis de conduire, à la possession d'un titre mensuel de transport collectif et au niveau de motorisation des ménages afin de permettre l'usage du modèle pour des fins de prévision.

Finalement, l'élaboration des ratios de compétitivité comme mesure d'attractivité des diverses options modales s'avère une piste intéressante dans le développement de modèles explicatifs du choix modal.

### 6.3 Limitations

De nombreuses simplifications ont été posées au cours de ce mémoire et doivent être rappelées. Tout d'abord, la disparité de définitions des heures de départs entre l'enquête OD et le GTFS servant aux simulations TC a causé le retrait d'environ 2000 déplacements nocturnes. Cette limitation peut être évitée en ajustant les heures de départ de l'enquête OD sur une période de 5h à 28h. Un autre facteur limitatif à cette méthode est la négligence des facteurs individuels tels que la préférence modale ou la perception des temps et des coûts. Ces facteurs pourront être ajoutés dans la poursuite des recherches.

Aussi, une faiblesse dans la méthode a été remarquée tardivement qui induit une limitation potentielle au niveau de la qualité des résultats des modèles. Tel que décrit au filtre #5, les

déplacements n'ayant pas trouvé d'itinéraires TC ont été exclus des modèles. Toutefois, avec l'évolution des réseaux TC tel que proposée dans les scénarios, certains de ces déplacements pourraient trouver un itinéraire. Ainsi, une méthode alternative, tel que l'utilisation d'un temps infini, aurait pu être définie en cas d'absence d'itinéraire afin de conserver ces déplacements et d'analyser leur évolution.

Finalement, la dernière limitation de cette méthode est liée à la qualité et la disponibilité des données utilisées. La qualité des ratios de compétitivité est directement liée à la qualité des temps de parcours utilisés. Les données provenant de diverses sources, il est possible que les ratios tels que présentés dans ce mémoire ne soient pas représentatifs de la compétitivité modale réelle. La disponibilité des données des scénarios est conditionnelle à un complexe travail de codification de réseaux de transport. Les lignes principales étant bien définies, le risque se situe davantage dans la codification des autres lignes à proximité et des lignes de rabattement qui doivent être ajustées en fonction des nouveaux pôles intermodaux. De plus, plusieurs simplifications peuvent avoir un impact sur les temps de parcours, notamment l'omission du temps de recherche d'un espace de stationnement pour les parcours effectués en voiture.

## 6.4 Perspectives

Les bases de la méthode d'évaluation des impacts, sur le choix modal, de scénarios de localisation résidentielle de transport ont été posées dans ce mémoire et plusieurs pistes d'approfondissement ont également été identifiées.

La première piste qui pourrait être poursuivie lors de recherches futures concerne l'analyse des scénarios 2031 PMAD, TOD 100% et CENTRAL. Les données n'étant pas disponibles à l'heure de déposer ce mémoire, cette section de l'analyse reste à compléter.

Ensuite, la définition du seuil minimal pour la distance TC pourrait être intéressante. Telle que discutée, la méthode proposée ici ne semble pas tout à fait adéquate ou est trop simplificatrice. Entre autres, la prise en compte de la pénibilité de la marche, du vélo et des différentes sections du parcours en transport en commun sont des facteurs qui ont été négligés qui seraient pertinent d'intégrer au calcul d'un seuil de distance minimal pour le TC.

La revue de littérature a mis en lumière le fait que les chaînes de déplacements sont rarement étudiées malgré leur significativité. L'intégration de ce facteur au modèle pourrait lui être bénéfique.

Afin de mieux saisir la variation des performances modales au cours de la journée, les paires origine-destination de chaque déplacement pourraient être simulés à chaque période horaire. Bien que nécessitant une grande quantité de calculs, cette méthode permettrait de mesurer l'évolution des ratios à travers les différentes plages horaires de la journée. Finalement, la principale difficulté dans la méthode proposée dans ce mémoire est la manipulation des données. La rédaction d'un script automatisant la manipulation de données est vitale à la mise en opération de cette méthode. Majoritairement développée avec l'aide de RStudio, l'automatisation du traitement des données géoréférencées demande l'utilisation de logiciel tel que PostGreSQL.

## BIBLIOGRAPHIE

- ANSELIN, L. (2015, Avril). Local Indicators of Spatial Association - LISA. *Geographical Analysis*, Vol 27, No. 2, pp. 93-115.
- BONNEL, P. (2002). Prévision de la demande de transport. *Rapport présenté en vue de l'option de diplôme d'habilitation à diriger les recherches*.
- BRINCKERHOFF, P., & FARRADYNE, P. (2002). *Rapport d'étape 1 - Réalisation des études d'achalandages et revenus pour les projets autoroutiers en partenariat public-privé dans la région de Montréal*. Montréal: PB Consult Inc.
- CERVERO, R. (2002). Built environments and mode choice: toward a normative framework. *Transportation Research Part D*, 7 (2002) pp. 265-284.
- CMM. (2011). *Un Grand Montréal attractif, compétitif et durable - Plan métropolitain d'aménagement et de développement*. Communauté métropolitaine de Montréal.
- COLLET, R. (2008). Dynamique du niveau de motorisation automobile des ménages français. ED EPS <hal-00318695>.
- CRANE, R. (2007). Is There a Quiet Revolution in Women's Travel? - Revisiting the Gender Gap in Commuting. Dans *Research on Women's Issues in Transportation, Volume 1: Conference Overview and Plenary Papers* (pp. 73(3) pp. 1-36). ISBN 978-0-309-09956-1 | DOI 10.17226/23274.
- DECAESTERCKER, C. S., SAERENS, M. (2017, Septembre). *Les arbres de décisions (decision trees)*. Récupéré sur Université Paris Descartes: <http://www.math-info.univ-paris5.fr/~lomn/Cours/DM/Material/ComplementsCours/decisiontree.pdf>
- DÉSILETS, S. (2012). Tendances liées à la composition, la motorisation et la localisation spatiale des ménages. *Mémoire de maîtrise*. École polytechnique de Montréal, Montréal, QC.
- DeWITTE, A., Hollevoet, J., HUBERT, M., DOBRUSZKES, F., & MACHARIS, C. (2013). Linking modal choice to motility: a comprehensive review. *Transportation Research. Part A: General*, no. 49, pp.329-341.
- DUBÉ, A., TÉTRAULT, L., & Al. (2016). Impact de projets de transport sur la santé publique : Développement de scénarios pour 2008-2031. Montréal: 20es Journées annuelles de santé publique.
- EWING, R., & CERVERO, R. (2010). Travel and the Built Environment. *Journal of the American Planning Association*, Vol 76, No.3, pp. 1-30.

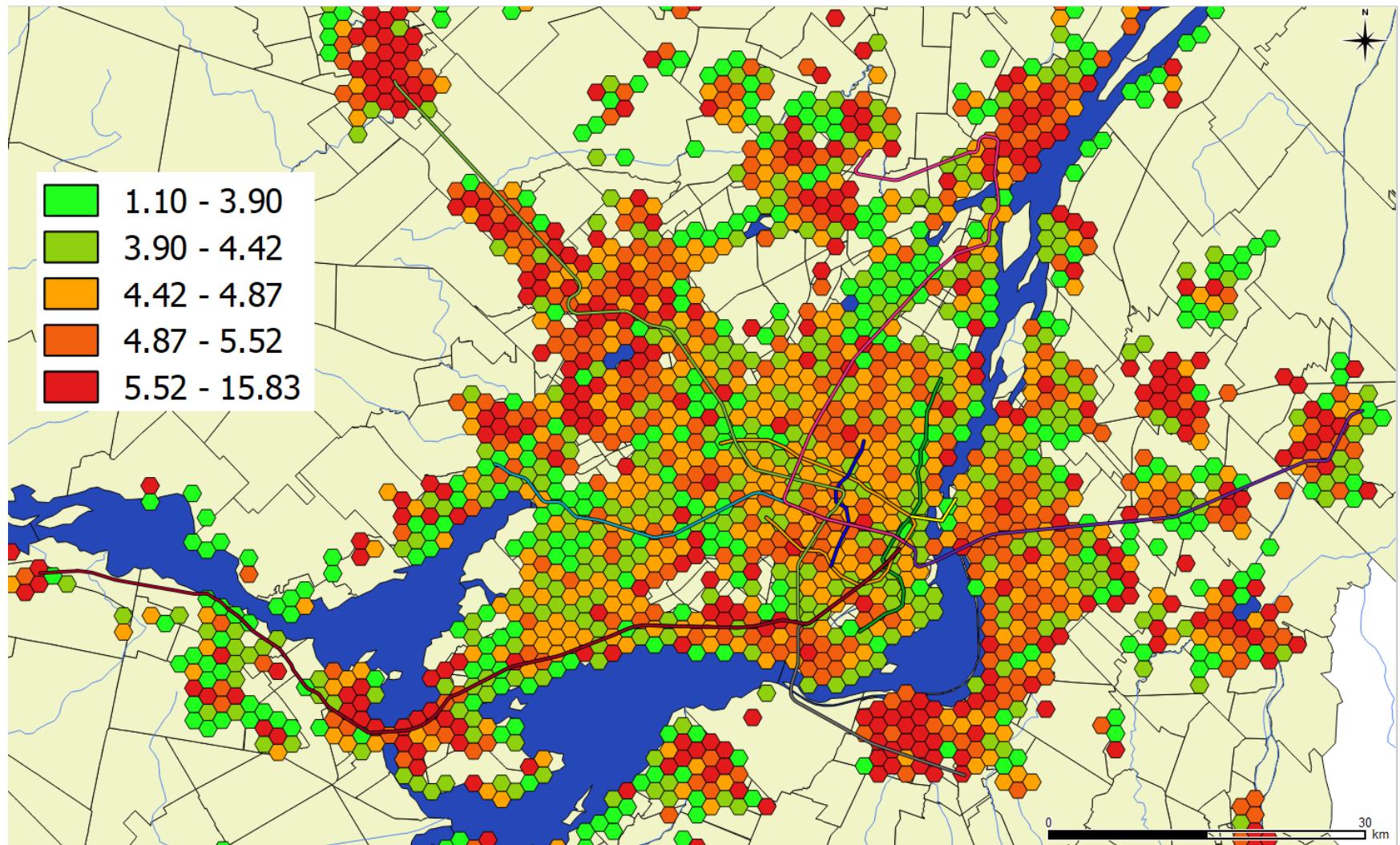
- GAGNON, L. (2006, Septembre). *Émissions de gaz à effet de serre des options de transport des personnes et des marchandises.* Récupéré sur Hydro-Québec: <http://www.hydroquebec.com/developpementdurable>
- HANDY, S. (2005). Planning for Accessibility: in Theory and in Practice. Dans *Access to Destinations* (pp. 131-147).
- HARTIGAN, J. A., & WONG, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, Vol. 28, No. 1, pp. 100-108.
- HUNT, D. (1990). A Logit Model of Public Transport Route Choice. *ITE Journal*, pp. 26-30.
- INSTITUTE, M. T. (2014). *Public Bikesharing in North America During a Period of Rapid Expansion: Understanding Business Models, Industry Trends, and User Impacts*. San José, CA: Library of Congress Catalog Card Number : 2014949233.
- JAFFE, E. (2015, Octobre 2015). *The Poor Bike, the Rich Bike-Share*. Récupéré sur CITYLAB: <https://www.citylab.com/transportation/2015/10/the-poor-bike-the-rich-bike-share/413119/>
- KAJITA, Y., TOI, S., CHISHAKI, T., & MATSUOKA, A. (2004). Structural Mechanism of Modal Choice Based on the Linked Structure of Trip Purpose and Transportation Choice. *Memoirs of the Faculty of Engineering, Kyushu University*, Vol. 64, No. 1, pp. 17-33.
- KAUFMANN, V., & JEMELIN, C. (2004). *La motilité, une forme de capital permettant d'éviter les irréversibilités socio-spatiales?* Lausanne: École Polytechnique Fédérale de Lausanne (EPFL).
- LITMAN, T. (2017). *Evaluating Accessibility for Transport Planning*. Victoria: Victoria Transport Policy Institute.
- LITMAN, T. (2017). *Evaluation Public Transit Benefits and Costs - Best Practices Guidebook*. Victoria: Victoria Transport Policy Institute.
- McFADDEN, D. (1974). Conditional logit analysis of qualitative choice behavior. Dans *Analysis of qualitative choice behavior* (pp. 105-141).
- MOBILITÉ, C. (2011). *Chaire de recherche sur l'évaluation et la mise en oeuvre de la durabilité en transport - Rapport d'activités 2010-2011*. Montréal.
- MORAN, P. (1950). Notes on Continuous Stochastic Phenomena. Dans *Biometrika* (pp. Vol. 37, No. 1/2, pp. 17-23).
- MORENCY, C. T.-A.-S. (2017). *Projet de recherche sur le développement d'un outil de suivi et d'analyse des conditions de circulation sur le réseau routier du MTMDET de la région de Montréal - Rapport Final*. Montréal: MTMDET.

- MORENCY, C., TRÉPANIER, M., & DEMERS, M. (2011). Walking to transit: An unexpected source of physical activity. *Transport Policy*, Vol. 18, pp. 800-806.
- MTMDET. (2017). Guide de l'analyse Avantages-Coûts des projets publics en transport routier. Bibliothèque et Archives nationales du Québec , ISBN 978-2-550-77782-3 (PDF).
- MTQ, & ADEC. (2014). *Évaluation des coûts de la congestion routière dans la région de Montréal pour les conditions de référence de 2008 - Rapport final préliminaire*. Montréal: Les Conseillers ADEC inc.
- NOROUSIS, M. (2009). Chapter 16 : Cluster Analysis. Dans *SPSS 17.0 Statistical Procedures Companion* (pp. 361-391).
- PATTERSON, Z., DAMANT-SIROIS, G., & EL-GENEIDY, A. (2014). *Nouveau modèle urbain: vers la mise en place d'un cadre intégré de modélisation du développement urbain et des transports dans la région de Montréal*. Montréal.
- POLIQUIN, E., MORENCY, C., & DEMERS, M. (2013). Estimation du marché actuel et potentiel de la marche à Montréal. Dans *La ville sous nos pieds : Connaissances et pratiques favorables aux mobilités piétonnes* (pp. 51-64). Montréal: INRS.
- RAKOTOMALALA, R. (2005). Arbres de Décisions. *Revue MODULAD*, No. 33, pp. 163-187.
- SAWADA, M. (2009). *Global Spatial Autocorrelation Indices - Moran's I, Geary's C and the General Cross-Product Statistic*. Récupéré sur University of Ottawa: <http://www.lpc.uottawa.ca/publications/moransi/moran.htm>
- TESSIER, M.-A. (2014, Mai 13). *Indicateurs d'analyse et de suivi de la congestion routière*. Récupéré sur Chaire Mobilité - Polytechnique Montréal: [http://www.polymtl.ca/mobilite/doc/Affiche\\_MA\\_Tessier\\_reduit.pdf](http://www.polymtl.ca/mobilite/doc/Affiche_MA_Tessier_reduit.pdf)
- TESSIER, M.-A. (2015). Développement d'indicateurs d'analyse et de suivi de la congestion routière. *Mémoire de maîtrise*. École polytechnique de Montréal, Montréal, Qc.
- TORRES-REYNA, O. (2014, Décembre). *Logit, Probit and Multinomial Logit models in R*. Récupéré sur Princeton University: <http://dss.princeton.edu/training>
- TREMBLAY, P. (2002). Survol technique du modèle de transport de la région de Montréal. Montréal: Ministère des Transport du Québec.
- VITON, P. (2014). *Mode Choice - The IIA Problem*. Récupéré sur Ohio State University: <http://facweb.knowlton.ohio-state.edu/pviton/courses2/crp5700/mode-choice-iiia.pdf>
- WEGENER, M., & FÜRST, F. (1999). *Land-Use Transport Interaction: State of the Art*. Dortmund: Institut für Raumplanung, Universität Dortmund.

## ANNEXE A – MATRICE COMPLÈTE – TEST KS – REF08

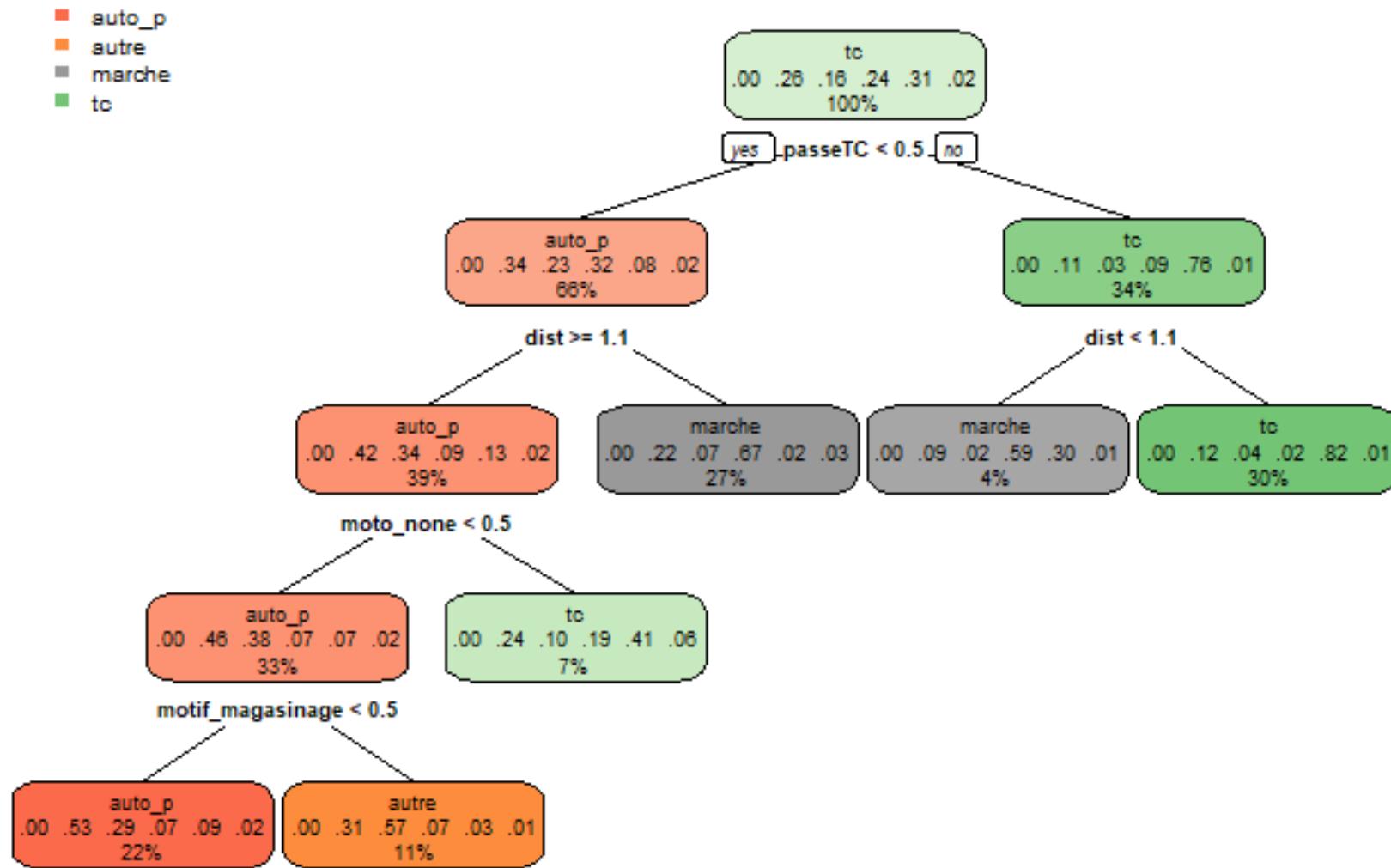
	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
5		2.1%	11.4%	17.0%	16.4%	20.5%	26.3%	24.3%	23.9%	22.7%	22.3%	13.8%	11.4%	12.3%	18.6%	19.9%	14.9%	12.6%	7.6%	5.5%	12.4%	7.3%	14.9%	14.3%
6	2.1%		11.8%	17.5%	17.1%	21.0%	27.1%	25.0%	24.6%	23.6%	23.3%	14.8%	12.0%	12.9%	19.4%	20.4%	15.6%	13.1%	8.7%	6.5%	13.5%	7.6%	16.5%	13.8%
7	11.4%	11.8%		5.9%	6.1%	10.3%	15.9%	13.7%	13.3%	13.2%	12.5%	5.0%	1.1%	2.0%	8.3%	9.2%	4.5%	2.4%	6.8%	8.9%	3.4%	10.6%	9.2%	9.3%
8	17.0%	17.5%	5.9%		1.9%	4.9%	10.1%	7.9%	7.5%	7.8%	7.4%	4.7%	5.8%	5.4%	3.6%	4.0%	4.6%	6.4%	11.3%	14.0%	8.1%	15.1%	14.4%	14.2%
9	16.4%	17.1%	6.1%	1.9%		4.4%	10.2%	8.1%	7.7%	7.2%	6.9%	3.4%	5.4%	4.5%	2.5%	3.6%	3.1%	4.9%	10.1%	13.2%	6.4%	15.5%	13.4%	14.6%
10	20.5%	21.0%	10.3%	4.9%	4.4%		6.4%	4.7%	3.9%	3.4%	2.8%	6.9%	9.5%	8.6%	2.5%	1.7%	6.6%	8.7%	14.1%	17.0%	10.4%	19.4%	16.7%	18.8%
11	26.3%	27.1%	15.9%	10.1%	10.2%	6.4%		2.4%	3.0%	4.7%	4.4%	12.8%	15.2%	14.4%	8.1%	7.0%	11.8%	14.7%	20.1%	22.8%	15.9%	24.1%	22.7%	23.2%
12	24.3%	25.0%	13.7%	7.9%	8.1%	4.7%	2.4%		1.9%	3.5%	2.5%	10.8%	13.1%	12.3%	6.2%	6.0%	9.8%	12.4%	18.0%	21.1%	13.9%	22.3%	21.2%	21.5%
13	23.9%	24.6%	13.3%	7.5%	7.7%	3.9%	3.0%	1.9%		2.0%	2.0%	10.3%	12.9%	11.9%	5.7%	4.5%	9.3%	12.1%	17.7%	20.5%	13.4%	22.1%	20.2%	21.1%
14	22.7%	23.6%	13.2%	7.8%	7.2%	3.4%	4.7%	3.5%	2.0%		1.9%	9.0%	12.4%	11.5%	5.4%	4.3%	9.6%	11.2%	16.3%	18.8%	13.4%	22.4%	18.5%	21.6%
15	22.3%	23.3%	12.5%	7.4%	6.9%	2.8%	4.4%	2.5%	2.0%	1.9%		8.8%	11.7%	10.8%	4.7%	4.3%	9.1%	11.0%	15.9%	18.6%	12.9%	21.9%	18.8%	21.3%
16	13.8%	14.8%	5.0%	4.7%	3.4%	6.9%	12.8%	10.8%	10.3%	9.0%	8.8%		4.3%	3.5%	5.2%	6.2%	3.0%	4.4%	8.1%	10.6%	6.0%	15.4%	10.5%	14.1%
17	11.4%	12.0%	1.1%	5.8%	5.4%	9.5%	15.2%	13.1%	12.9%	12.4%	11.7%	4.3%		1.1%	7.6%	8.6%	3.8%	1.6%	6.3%	8.3%	3.6%	11.3%	8.7%	10.1%
18	12.3%	12.9%	2.0%	5.4%	4.5%	8.6%	14.4%	12.3%	11.9%	11.5%	10.8%	3.5%	1.1%		6.8%	7.8%	2.8%	1.7%	6.3%	9.3%	3.0%	12.0%	9.8%	11.1%
19	18.6%	19.4%	8.3%	3.6%	2.5%	2.5%	8.1%	6.2%	5.7%	5.4%	4.7%	5.2%	7.6%	6.8%		1.5%	4.6%	7.0%	12.2%	15.3%	8.3%	17.6%	15.4%	16.7%
20	19.9%	20.4%	9.2%	4.0%	3.6%	1.7%	7.0%	6.0%	4.5%	4.3%	4.3%	6.2%	8.6%	7.8%	1.5%		5.7%	7.9%	13.6%	16.5%	9.4%	18.4%	16.5%	17.7%
21	14.9%	15.6%	4.5%	4.6%	3.1%	6.6%	11.8%	9.8%	9.3%	9.6%	9.1%	3.0%	3.8%	2.8%	4.6%	5.7%		3.3%	8.6%	11.6%	4.3%	13.2%	11.7%	12.4%
22	12.6%	13.1%	2.4%	6.4%	4.9%	8.7%	14.7%	12.4%	12.1%	11.2%	11.0%	4.4%	1.6%	1.7%	7.0%	7.9%	3.3%		6.3%	9.3%	4.4%	12.0%	9.9%	10.8%
23	7.6%	8.7%	6.8%	11.3%	10.1%	14.1%	20.1%	18.0%	17.7%	16.3%	15.9%	8.1%	6.3%	6.3%	12.2%	13.6%	8.6%	6.3%		3.9%	7.5%	8.4%	11.2%	8.9%
24	5.5%	6.5%	8.9%	14.0%	13.2%	17.0%	22.8%	21.1%	20.5%	18.8%	18.6%	10.6%	8.3%	9.3%	15.3%	16.5%	11.6%	9.3%	3.9%		9.2%	8.0%	12.0%	11.9%
25	12.4%	13.5%	3.4%	8.1%	6.4%	10.4%	15.9%	13.9%	13.4%	13.4%	12.9%	6.0%	3.6%	3.0%	8.3%	9.4%	4.3%	4.4%	7.5%	9.2%		11.0%	8.7%	9.6%
26	7.3%	7.6%	10.6%	15.1%	15.5%	19.4%	24.1%	22.3%	22.1%	22.4%	21.9%	15.4%	11.3%	12.0%	17.6%	18.4%	13.2%	12.0%	8.4%	8.0%	11.0%		14.6%	14.3%
27	14.9%	16.5%	9.2%	14.4%	13.4%	16.7%	22.7%	21.2%	20.2%	18.5%	18.8%	10.5%	8.7%	9.8%	15.4%	16.5%	11.7%	9.9%	11.2%	12.0%	8.7%	14.6%		11.5%
28	14.3%	13.8%	9.3%	14.2%	14.6%	18.8%	23.2%	21.5%	21.1%	21.6%	21.3%	14.1%	10.1%	11.1%	16.7%	17.7%	12.4%	10.8%	8.9%	11.9%	9.6%	14.3%	11.5%	

## ANNEXE B – CARTE DES RATIOS TC\*, 7H À 9H59



## ANNEXE C – ARBRE DE DÉCISION, NON-DÉTENTEURS DE PERMIS DE CONDUIRE

### Arbre de décision du choix modal pour les non-détenteurs de permis de conduire



## ANNEXE D – ARBRE DE DÉCISION, DÉTENTEURS DE PERMIS DE CONDUIRE

### Arbre de décision du choix modal pour les détenteurs de permis de conduire

